Genetic Analysis of Leaf and Sprout Traits of Cabbage and Brussels Sprout

(MSc Thesis Plant Breeding)

By: **A M Sajedur Rahman**

WAGENINGEN
UNIVERSITY & RESEARCH

March 2019

MSc Thesis Plant Breeding (PBR-80439)

Genetic Analysis of Leaf and Sprout Traits of Cabbage and Brussels Sprout

By

A M Sajedur Rahman

900713679060

Supervisor:	Dr. ir. Guusje Bonnema &
	Johan Bucher
Examiner:	Dr. ir. Guusje Bonnema &
	Dr. Marieke Jeuken
University:	Department of Plant Breeding
	Wageningen University and Research
	6708 PB Wageningen
	The Netherlands

March 2019

Abstract

Brassica oleracea is an important vegetable species with diverse morphotypes. For this thesis heading cabbages and Brussels sprouts were considered. Different traits of heading cabbages and Brussels sprout had been phenotyped in the field and in Unifarm. The aim of the thesis was to assess variation in phenotypic traits at harvest stage of heading cabbage and Brussels sprout and study their relatedness and genetic regulation. In addition, a cultivation practice, decapitation was tested to observe its effect on the overall yield of Brussels sprout. The correlation between different traits was observed and interesting correlation was found between different traits. Population structure was corrected with PCO and it was found that it can potentially reduce the no. of false positive SNPs. To accomplish the aim of the thesis, Genome Wide Association Study (GWAS) was conducted. The phenotypic data, PCO and genotypic data were combined for conducting of GWAS in TASSEL software. After analysing, TASSEL generated 110 significant SNPs for the studied traits of heading cabbages and Brussels sprout. For the significant SNPs, gene within 50 Kb were searched in the *Brassica* genome browser and their orthologues in *Arabidopsis* genome browser. Several important genes were also found which can be considered as potential candidate genes as they are reported to be involved in growth and development. It was also observed that, the decapitation practice significantly influenced the yield of Brussels sprout. Moreover, it was observed that genetic regulation of yield differs between treatment as different significant marker trait association was found in different treatments. No previous literature was found for genetic regulation of sprout development in Brussels sprout and axillary shoot development of cabbages, so, this thesis can help the researcher to conduct a successful GWAS approach on several yield traits of these two morphotype of *B. oleracea*.

Keywords: GWAS, Heading cabbages, Brussels sprout, Decapitation, yield trait

Acknowledgement

I would like to thank dr. ir. Guusje Bonnema for her advice, supervision, caring and guidance during this thesis period. Thanks, Johan Bucher, for inspiring me to use the word “challenge” instead of “problems”, which helped me believe in me and made me confident to accept challenges. Marieke Jeuken also helped me to get the thesis which gave me the opportunity to work with a very dynamic group. I learned a lot from my supervisors and from fellow students. Specially Chunmei Zou, Pim van de Mortel and Zakaria Alam explained a lot of things to me, which helped me to understand the thesis easily. Moreover, it was my parents and fiancée who constantly support me in this journey, without whom it was not possible to continue the thesis. Last but not least, I would also like to thank the Dutch government and Netherland Fellowship Program (NFP) for giving me the scholarship to study in Wageningen University and conducting this research.

Contents

1 INTRODUCTION

1.1 BRASSICACEAE AND BRASSICA

The *Brassicaceae* is one of the largest and important family of the angiosperms, popularly known as crucifers or mustard family. It comprises 370 genera and more than 4000 species with 10 poorly defined tribes. Several species of *Brassicaceae* has great scientific, agronomic and economic importance, as it includes the famous model plant *Arabidopsis thaliana* (*L.*), and several vegetable, fodder and oil producing *Brassica oleracea* (cabbage, cauliflower, broccoli, sprouts etc.), *B. rapa* (chinese cabbage, pak choi, etc.) and *B. napus* (oilseed rape, canola, etc.) crops (Sun, 2018; Zou, 2019).

Brassica is a large and diverse genus of *Brassicaceae* family, which encompasses several important vegetable, oil and condiment crops (Cartea et al., 2011). The *Brassica* genus is a monophyletic group and evolutionary closely related to the model plant *A. thaliana*; which diverged from a common ancestor about 14.5-24 million years ago (MYA) (Browers et al., 2003). Some other researcher also reported from the fossil evidence that, the *Brassica* and *A. thaliana* had evolved from the same ancestor ~43 MYA (Beilstein et al., 2010). 13-17 MYA a whole genome triplication (WGT) event occurred which yielded different species in *Brassica* (Town et al., 2006). However, along with WGT some other human interventions, e.g. domestication upon different traits and breeding contributed to the evolvement of different extreme morphotypes within different species of *Brassica* (Bonnema et al., 2011).

The *Brassica* includes six economically important species which are interrelated and can be described by the "triangle of U" (Nagaharu, 1935) (Figure 1). Among the six species, *Brassica rapa* (AA, 2*n* = 20), *Brassica oleracea* (CC, 2*n* = 18) and *Brassica nigra* (BB, 2*n* = 16) are diploid and hybridized to form three allotetraploid species *Brassica juncea* (AABB, 2*n* = 36), *Brassica carinata* (BBCC, 2*n* = 34), and *Brassica napus* (AACC, 2*n* = 38) (Zhao et al., 2013). Due to domestication and selection, different morphotypes are also present within a species.

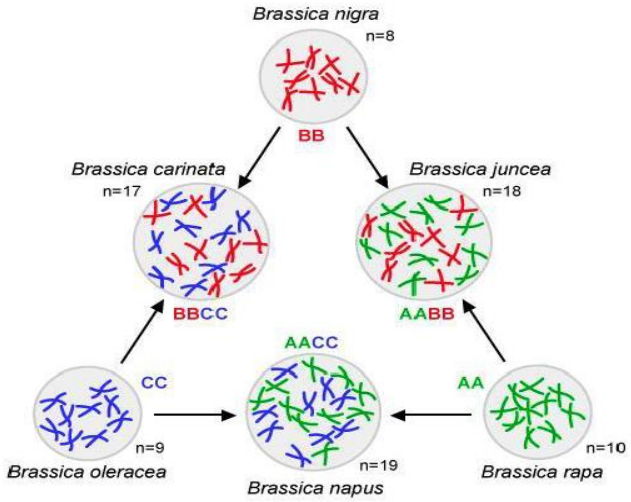

Figure 1. "Triangle of U" illustrating the genetic relationships among six economically important species of brassica genus Nagaharu, 1935.

1.2 BRASSICA OLERACEA

B. oleracea is a species with one of the most diversified and extreme morphotypes among the cultivated species of *Brassica* genus. *B. oleracea* represent different extreme morphotypes like heading cabbages (red, white, savoy and pointed), cauliflower, brussels sprout, kohlrabi, broccoli, tronchuda, collard green, chinese kale and different ornamentals. It is assumed that the *B. oleracea* is evolved from a wild cabbage (Smyth, 1995). North Atlantic and Mediterranean region is considered to be the centre of origin, whereas domestication was taken place in Europe (Cartea et al., 2011, Maggioni et al., 2010). During domestication, selection took place at different plant parts of the wild species and resulted different extreme morphotypes (Figure 2) (Kalloo and Bergh 1993; Stromberg, 2015).

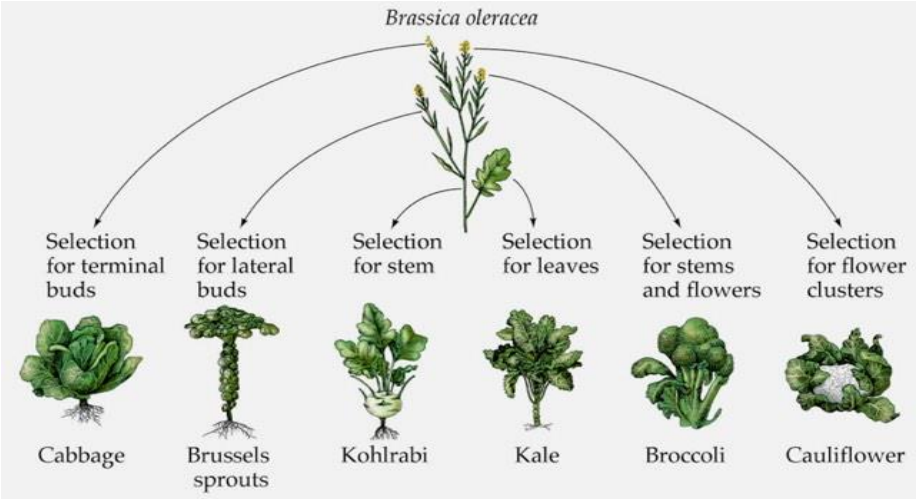

Figure 2. Domestication by selection of different traits from wild mustard, *brassica oleracea* (Kallo and Bergh 1993, Stromberg, 2015).

1.3 HEADING CABBAGE AND BRUSSELS SPROUT

There are different important morphotypes of *B. oleracea*, for this thesis the focus was on different late traits of heading cabbages (ssp. *capitata*) and Brussels sprout (ssp. *gemmifera*) as these were in their mature or harvest stage at the start of my thesis (Third September 2018). Heading cabbages (white, red, savoy and pointed) are leafy vegetables that are well known for their entire heads. Nishijima & Fukino, 2006 found that heads are developed due to self-shading and blanching of the leaves. Moreover, they also suggested, round shaped leaves with low length to width ratio and overlapping leaves leads to form leafy heads in cabbages. In the modern varieties of heading cabbages, it is desired to have single entire head at the top of the plant with no or relatively fewer side axillary shoots developing from the axillary buds. The axillary shoots can develop into small heads in the later stages of the plant growth of the heading cabbages. That means the photosynthates are also translocated to the secondary heads, which may affect the final yield of the main head. However, in comparison with heading cabbages, the Brussels sprouts generally do not form a compact head like structure at the top of the plant. Though in some accession a loose head is formed at the top of the plant at the end of the growing season. Brussels sprout is cultivated for its round or oval shaped miniature cabbage like structures. The miniature cabbages are developed from the axillary bud of the plant which are popularly known as sprouts.

Though both the axillary shoots of cabbages and the sprouts of Brussels sprout are developed from the axillary buds, they are not morphologically similar. In the cabbages, the axillary bud develops into a shoot like structure with expanded leaves whereas; the buds of Brussels sprout develop into a round or oval shaped miniature cabbages with almost no expanded leaves. The classical hypothesis of axillary bud outgrowth suggests that, there are three stages for axillary shoot development: dormancy, transition and sustained growth which are regulated by phyto-hormones. The master regulatory hormone, auxin, plays the vital role for controlling the fate of the dormant axillary buds (Yaish, et al., 2010). One hypothesis is that, the Brussels sprouts may have a prolonged transition stage, whereas the cabbages may switch to the sustained growth stage so quickly and form axillary shoots.

1.4 Association mapping, GWAS and Population structure

Genetic locus that are associated with different traits can be mapped to a genomic region and there are two different methods for doing so; quantitative trait locus (QTL) mapping and association mapping (AM). AM is also known as linkage disequilibrium (LD) mapping or Genome Wide Association Study (GWAS). Biparental segregating population is used in QTL mapping, whereas in the AM, natural genetic variation in the mapping population is used for linking genomic regions to variation in a phenotypic trait. AM considers more recombination than QTL mapping and gives better resolution to detect more alleles with higher speed (Zhu et al., 2008). But, GWAS has some limitations as well, as it led to detect some irrelevant alleles (false positive) due to population structure, as the genetic relatedness among the genotypes varies, in contrast to offspring of biparental population. However, the challenge may also be solved by correcting the population structure.

GWAS and candidate gene association mapping are the two ways for conducting AM. Genome wide markers are screened over the population in GWAS and in candidate gene approach the allelic variation in the candidate genes are profiled (Zhu et al., 2008). In this study both the GWAS and candidate gene approach has been followed. The genotypes within the sub-morphotypes are genetically more related than different morphotypes, so, false positive results may also occur. This issue had been solved by using DARwin software by correcting population structure.

The population of heading cabbages contained four different sub-morphotypes (white, red, pointed and savoy) and the total number of accessions were 180. It is considered that breeding and selection have taken place at geographically different locations and no intercrossing has taken place among the morphotypes. So, different sub-morphotypes of the heading cabbages may have different levels of relatedness. However, Brussels sprout had a relatively smaller group with 49 accessions, but their origin was diverse. So, it is assumed the accessions have different degree of relatedness. For being more precise to detect SNPs that may be associated to a trait and eliminating the false positive SNPs, population structure was corrected for the heading cabbages and Brussels sprouts group separately (Alam, 2018; Korte and Farlow, 2013; Mortel, 2018).

Population structure correction can be done by using STRUCTURE or PCO. Both the STRUCTURE and PCO has been used for population structure correction for *B. oleracea* and *B. rapa*. But, however, PCO gives a better correction than the STRUCTURE (Alam, 2018; Brouwer, 2018; Del Carpio et al. 2011; Earl & VonHoldt, 2012; Mortel, 2018; Pang et al., 2015). So, for this thesis the population structure was

corrected with PCO with the 1383 SNPs that are equally distributed over the genome and have low number of missing values (Alam, 20018; Del Carpio et al., 2011).

In the last four years, quite a lot of research has been conducted in a view to detect different genomic regions and candidate genes for leaf and heading traits of cabbage under the TKI project. Different leaf related traits were phenotyped in earlier stages of plant growth but there are some additional traits that may have an influence on the quality and yield of cabbage at later stages of plant growth. Yield related traits of Brussels sprouts were not also phenotyped in the past years. So, for this thesis different trait data were considered for cabbages and a treatment was considered for Brussels sprout at harvesting stages of plant growth.

1.5 AIM OF THE THESIS

Different traits of heading cabbages and Brussels sprout had been phenotyped in the field and in Unifarm. The aim of the thesis was to assess variation in phenotypic traits at harvest stage of heading cabbage and Brussels sprout and study their relatedness and their genetic regulation. In addition, a cultivation measure was tested to observe its effect on the overall yield of Brussels sprout. To achieve the aim of the thesis several goals were defined.

1.6 GOAL

1. a) Define, collect and analyse different traits relevant for cabbage head or sprout development like, leaf, axillary shoot (cabbage) and sprout traits (B. sprout)
   b) Analyse correlation among different traits
2. a) Analyse population structure of the studied population
   b) Use trait data as input for GWAS
   c) Analyze genomic regions on the *B. oleracea* genome for significant SNP trait relations
3. Test the hypothesis that decapitation influence the overall and marketable yield of Brussels sprout.

1.7 THESIS OUTLINE

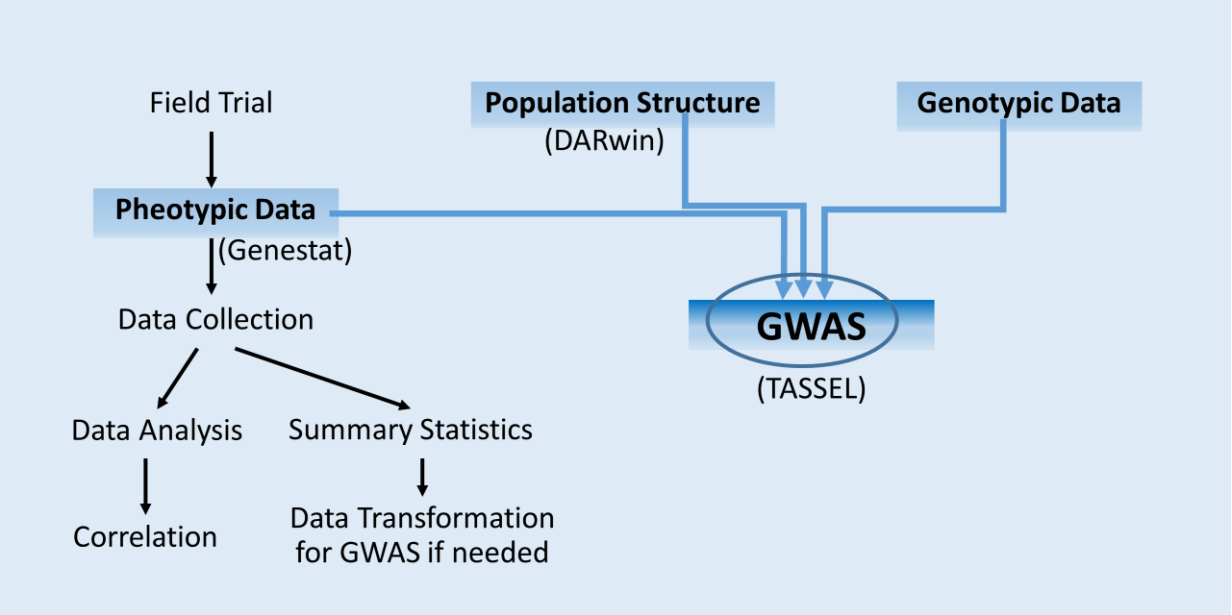

Figure 3. Schematic representation of the thesis overview; software packages are within the brackets that were used for different purposes.

2 MATERIALS AND METHODS

In this chapter, the experimental setup of the field experiment, phenotyping of the different traits, data collection, data analysis, population structure correction and GWAS is explained. The chapter is divided into several paragraphs which give an outline of the whole study.

2.1 FIELD TRIAL AND EXPERIMENTAL SETUP

404 accessions of 7 different morphotypes of *B. oleracea* were planted in the field at 31st May 2018. Randomized complete block design with two replications was followed during this experiment. 5 plants were planted per accessions per block. The morphotypes along with their numbers of accessions are represented in Table 1. A summary of the field lay out can be found in appendix 1. The heading cabbages have the highest number of accessions (180) whereas the collard green represents the least number of accessions (20).

Table 1. Different morphotypes with their associated accessions numbers

Morphotype	Number of hybrids	Number of accessions	Total number
1. Heading Cabbage			
a. White Cabbage	24	75	99
b. Red Cabbage	30	23	53
c. Savoy Cabbage	0	20	20
d. Pointed Cabbage	5	3	8
Total Heading Cabbage (a+b+c+d)	59	121	180
2. Cauliflower	24	36	60
3. Brussels Sprout	10	39	49
4. Kholrabi	17	31	48
5. Tronchuda	1	24	25
6. Ornamentals	10	12	22
7. Collard Green	0	20	20
Grand Total (1+2+3+4+5+6+7)	121	283	404

For this thesis, phenotyping was done at the late of the season so, leaf numbers were counted for the whole plant and within a head of heading cabbages. Number of axillary shoots and secondary heads that were developed from the axillary shoots were also recorded. For Brussels sprout, plant height, stalk height, total yield and marketable yield were also recorded. In the experiment field of 2018, there were 180 accessions of cabbage and 49 accessions of Brussels sprout. One of the objectives of this thesis was to collect and define different leaf and axillary shoot (cabbages)/ bud (Brussels sprout) related data and study the correlation among different traits and use those data as input of GWAS and later on tried to find the genomic regions for those traits. To accomplish this objective, different field data were collected and transformed if needed to study GWAS and correlation.

However, for another objective of this thesis, a treatment was given to the Brussels sprout accessions by removing the top portion of the plant (decapitation) and compared those after 3 weeks with the controls to see if there is any impact of the top leaves on the overall uniformity of the sprout formation (Figure 4). For doing so, decapitation was done with two plants per accession per block. During harvesting, the decapitated two plants were harvested separately from each accession. Then all the

harvested plants were brought to the Unifarm, WUR. Plant height and stalk height was recorded, and the sprouts were separated from the plants. After that, overall yield was recorded, then the sprouts were graded into three categories upon their size and the weight of all three categories were recorded.

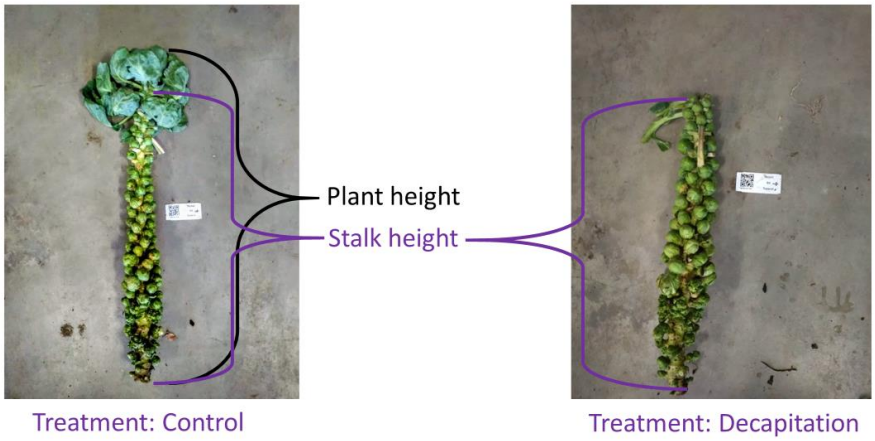

Figure 4. Decapitation of Brussels sprout to see the impact of top leaves on overall uniformity of sprouts

2.2 PHENOTYPING

In 2018, the other students of the growth and development group had phenotyped several leaf related and heading data of cabbages (Alam, 2018; Mortel, 2018; Zou, 2019) (Table 2). But, no axillary shoot related data and the number of head leaves were recorded. So, for cabbages, different axillary shoot related traits were recorded (Figure 5). The no. of dead and alive leaves was also recorded. The developed axillary shoots were graded into three categories upon their length and then the no. of axillary shoots was recorded. Furthermore, it was observed that some of the sprouts developed into secondary heads. All the heads were categorized and recorded again into three different groups upon their sizes (Table 3). For measuring length of a shoots or a secondary head, base to the tip of the shoots or head were considered. 5cm x 5cm, and 10cm x 10cm squares were drawn to the measuring table to fasten the speed of the work. The squares were used occasionally for classifying the confusing axillary shoots and secondary head in an appropriate group.

Table 2. Different traits that had been phenotyped for heading cabbages in 2018

Phenotyped Traits	Unit	Researcher(s)
Number of scars	#	Mortel, 2018
Number of leaf	#	Mortel, 2018
Total # non-heading leaves	#	Mortel, 2018, Zou, 2019
Leaf length	mm	Alam, 2018, Mortel, 2018, Zou, 2019
Leaf width	mm	Alam, 2018, Mortel, 2018, Zou, 2019
Leaf area	mm^2	Alam, 2018, Mortel, 2018, Zou, 2019
Leaf ratio	-	Mortel, 2018, Zou, 2019
Petiole length	mm	Mortel, 2018
Petiole width	mm	Mortel, 2018
Head weight	gram	Mortel, 2018, Zou, 2019
Head diameter	mm	Mortel, 2018, Zou, 2019

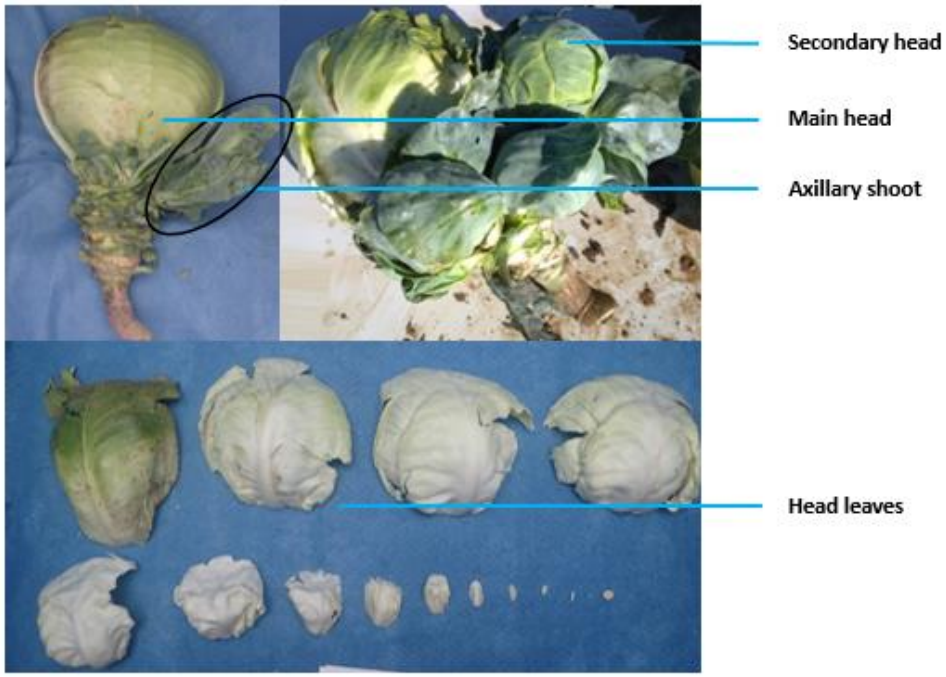

Figure 5. Different traits that were recorded from cabbage accessions

For Brussels sprout focus was on yield. The harvested sprouts were categorized into three different groups upon their size. Along with yield, plant height and stalk height was also recorded. The sprouts were harvested from the plant stalks by some mechanical means. Then the harvested sprouts were sorted and graded into three categories. 15 mm, 25 mm, 35 mm meshes had been used to categorize the sprouts. The overall weight and individual weight of all the categories were recorded. Figure 6 illustrates the working procedure that were followed for phenotyping of b. sprouts.

Figure 6. Harvesting, grading and weighing of sprouts

The complete list of all the traits that had been phenotyped for this thesis can be found in table 3, the table also gives an impression of the traits.

Table 3. Different traits that had been phenotyped for heading cabbages and b. sprouts

Trait name	Trait Description	Unit	Time of measurement
Heading cabbage subset			
Scars	No. of dead leaves before heading	#	19 Sep-19 Oct
Leaves	No. of alive leaves before heading	#	19 Sep-19 Oct
Total leaves	Total no. of alive leaves + scars before heading	#	19 Sep-19 Oct
Head leaves	Total no. of leaves within the head	#	19 Sep-19 Oct
Axillary shoot gr1	<5 cm sized axillary shoots developed from the axillary bud	#	19 Sep-19 Oct
Axillary shoot gr2	5.1-10 cm sized axillary shoots developed from the axillary bud	#	19 Sep-19 Oct
Axillary shoot gr3	>10 cm sized axillary shoots developed from the axillary bud	#	19 Sep-19 Oct
Total no. of axillary shoot	Total no. of axillary shoots developed from the axillary bud + all the secondary heads	#	19 Sep-19 Oct
Secondary head gr1	<5 cm sized secondary heads that developed from the axillary shoots	#	19 Sep-19 Oct
Secondary head gr2	5.1-10 cm sized secondary heads that developed from the axillary shoots	#	19 Sep-19 Oct
Secondary head gr3	>10 cm sized secondary heads that developed from the axillary shoots	#	19 Sep-19 Oct
Total no. of secondary head	Total no. of secondary heads that developed from the axillary shoots	#	19 Sep-19 Oct
Brussels sprout subset			
Plant height	Above ground height of the plant	cm	03 Dec
Stalk height	Height of the plant without the top crown	cm	03 Dec
Total yield	Total plant yield that includes, total marketable yield + <15 mm sprout + burst/cracked sprouts	gram	04 Dec- 11 Dec
Yield gr1	Weight of 15-25 mm sized un-opened round or oval shaped sprout	gram	04 Dec- 11 Dec
Yield gr2	Weight of 25.1-35 mm sized un-opened round or oval shaped sprout	gram	04 Dec- 11 Dec
Yield gr3	Weight of >35 mm sized un-opened round or oval shaped sprout	gram	04 Dec- 11 Dec
Total market yield	Total weight of yield group 1, 2 and 3	gram	04 Dec- 11 Dec
Yield loss	Difference between total yield and total marketable yield	gram	04 Dec- 11 Dec
%Yield loss	Difference between total yield and total marketable yield in percentage	-	04 Dec- 11 Dec

2.3 STATISTICAL ANALYSIS OF PHENOTYPIC DATA

The collected phenotypic data were analysed by Genstat software (VSN International, 2015). For cabbages one-way ANOVA was conducted to see if there is any effect of the block. If there is no effect of block, then the average value of each accessions was considered. Mixed model (REML) was used to see the effect of the decapitation treatment on Brussels sprout. Correlation between the traits were also checked by conducting Pearson's correlation. Then, the normal distribution of the residuals was checked by Q-Q plots. In some of the cases skewness was found, data transformation was applied to generate normal distribution data. However, in case of extreme skewness it was not possible to get the normal distribution even after transformation. For GWAS, extremely skewed data series and traits were not considered; as they tend to create false positive QTLs.

2.4 GENOTYPING

The growth and development group of plant breeding department has been working along with some other breeding companies to understand the genetic relationship of a wide range of germplasm collection of *B. oleracea* modern hybrids, land races and wild accessions under the "TKI 1000 genome project". This germplasm collection represents all the morphotypes and related species of *B. oleracea*. Appendix 2 represents the complete list of different morphotypes that had been considered for genotyping.

Hybrids were homogenous but the other accessions were heterogeneous, so, during DNA isolation two different strategies were followed. Cotyledons and hypocotyls of 50 to 100 seedlings were harvested for each modern hybrid, whereas, in case of other accessions of the germplasm collection one representative plant from each accession was harvested for genotyping. Theo Borm extracted the genotypic information from the Keygene generated sequence-based genotyping (SBG) data. At first, Theo found more than 200,000 Single Nucleotide Polymorphisms (SNPs) in 936 accessions, but however most of the SNPs were not present in most of the accessions. Then the number of SNPs were reduced to 18.580 SNPs on the basis of their occurrence in at least 80% of all the studied accession and having a minor allele frequency of more than 2.5% (Alam, 2018; Brouwer, 2018; Mortel, 2018; Slob, 2016; Zou, 2019). For population structure analysis of this study, 1376 SNPs were selected with a distance of ≥250 Kb, so that they evenly distributed over the genome (Alam, 2018; Groot, 2017; Zou, 2019).

2.5 POPULATION STRUCTURE

Three different PCOs had been constructed for three different data sets of the heading cabbages. DARwin software package was used for constructing those PCOs. During PCO construction 106 accessions, 133 accessions and 140 accessions were considered for head leaves, other leaf related traits and axillary shoot related traits respectively. The highest possible number of axes were also considered in every case for the construction of those PCOs (Table 4).

Table 4. Different traits with their no. of accessions and the no. of axes that were considered during construction of PCOs for heading cabbages

Traits	No. of accession	PCO axes
Head leaves	106	106 axes
Scars	133	133 axes
Leaves	133	133 axes
Total leaves	133	133 axes
Axillary shoot gr1	140	140 axes
Axillary shoot gr2	140	140 axes
Axillary shoot gr3	140	140 axes
Total no. of axillary shoot	140	140 axes
Secondary head gr1	140	140 axes
Secondary head gr2	140	140 axes
Secondary head gr3	140	140 axes
Total no. of secondary head	140	140 axes

Brussels sprout had a comparatively smaller group with 48 accessions. Though the group is smaller but there may have some degree of relatedness within the accessions as they belong to different geographic location. So, a PCO was also constructed for Brussels sprout by using DARwin. All the 48 accession and 48 axes were considered for this PCO.

2.6 Genome Wide Association Study (GWAS)

TASSEL (Trait Analysis by aSSociation, Evolution and Linkage) v5.2.33 software was used to establish association between genetic marker and phenotype (Alam, 2018; Bradbury et al. 2007, Islam, 2017; Mortel, 2018). To observe the marker-trait association, General Linear Model (GLM) was used. Genotypic data, phenotypic data and corrected population structure were used as input of GLM. GLM was performed with 999 combinations to reduce the error (Anderson and Braak, 2003). False Discovery Rate (FDR) was considered for marker-trait association. Then the significant SNPs were selected from the Manhattan plots for each trait. The selection was done upon some assumptions like; markers should have high LOD score, flanking markers should have to be in the same region with high LOD score, LOD score should be higher than the FDR threshold (Alam, 2018). A complete guideline for conducting GWAS in TASSEL has added to the appendix 3.

LOD is the value of the –log10 *p* value. In this thesis the threshold level was set to LOD score 4 as a rule of thumb, which is equivalent to $p=10^{-4}$ (Zou, 2019). A better correction with PCO and comparatively higher LOD score yielded fewer significant SNPs.

For selecting the candidate genes for a trait, a region of 50kb was considered to either side of the marker (Cheng et al., 2016). Then the genes were also entered in the brassica genome browser (BRAD, 2019) and *Arabidopsis* genome browser (TAIR, 2019) to find their actual function (Alam, 2018; Mortel, 2018; Yu et al., 2013). Furthermore, the genes were also searched in different literature and in the previous theses.

3 RESULT

Phenotypic results for heading cabbages and Brussels sprout are presented at first, then the population structure for these two collections is presented and after that the GWAS results. Physically linked SNPs and the SNPs that are associated with different traits are also presented here. At the end of the chapter the candidate genes are shown. As different traits of two morphotypes were considered for this thesis, so each of sub-chapter is divided into two parts. The first part represents the result of heading cabbages and the second part represents Brussels sprout results.

3.1 PHENOTYPING DATA

3.1.1 Heading cabbages

All the phenotyping data for heading cabbages were recorded between mid-September to mid-October (133-163 DAS) except the head weight and head diameter. The head weight and diameter data were recorded by another thesis student between end of August till second week of September (111-124 DAS). There were 180 accessions for heading cabbages in the field, but it was not possible to collect data from all the accessions. Some of the accession started flowering and some did not produce a head. Moreover, there was severe pest damages in some of the accessions, so it was not possible to collect data from those accessions. So, at the end of the data collection process axillary shoot data were collected from 140 accessions, head weight data from 137 accessions, leaf data from 133 accessions, head leaves data from 106 accessions and secondary head data from 90 accessions.

One-way ANOVA test was performed to see the possible block effects. Block effect was found for some of the traits. This effect may have been occurred due to the variation of data collection period. Data collection started with block A and it took 17 days to collect all the data then the data from block B was recorded which took 14 more days. So, eventually the cabbages of Block B had at least 18 more days of growth which may have an influence on block effect. Then the rank of each block was observed for total # leaf and 6 out of the top 10 accession matches (Appendix 4). The normality of the data set was checked by using QQ plots with 95 % confidence interval. Data for some of the traits were not normally distributed, so some data sets were transformed with log and square root transformation. No transformation was needed for total # leaves before heading, head leaves and total # axillary shoots. Log transformation and square root transformation was used for leaf data and axillary shoot gr1 data respectively. Specific transformation mode that were used for achieving normal distribution of data are listed in table 5. Summary statistics of all the traits can also be found in appendix 5.

Table 5. Mode of data transformation for achieving normal distribution of data

Trait	Successful mode of data transformation for normality	Comments
Heading cabbages		
Scars	-	Not possible to achieve normality even after data transformation
Leaves	Log10	-
Total leaves	-	No transformation required
Head leaves	-	No transformation required
Axillary shoot gr1	Square root	-
Axillary shoot gr2	-	Not possible to achieve normality even after data transformation
Axillary shoot gr3	-	Not possible to achieve normality even after data transformation
Total no. of axillary shoot	-	No transformation required
Secondary head gr1	-	Not possible to achieve normality even after data transformation
Secondary head gr2	-	Not possible to achieve normality even after data transformation
Secondary head gr3	-	Not possible to achieve normality even after data transformation
Total no. of secondary head	-	Not possible to achieve normality even after data transformation
Brussels Sprout		
Plant height	-	No transformation required
Stalk height	-	No transformation required
Total yield	Square root	-
Yield gr1	Square root	-
Yield gr2	Square root	-
Yield gr3	-	Not possible to achieve normality even after data transformation
Total market yield	Square root	
Yield loss	Log10	-
%Yield loss	Log10	-

Twelve trait data were collected from the field and two trait data (head weight (gram) and diameter (mm)) were added from another master thesis of Pim van de Mortel (Mortel, 2018). To observe the correlation between the traits, Pearson's correlation test was performed and the correlation between all the 14 traits are presented in table 8. Total # leaves and # leaf, axillary shoot gr1 and total # axillary shoot, secondary head gr1 and total # secondary heads are highly positively correlated (0.88, 0.89 and 0.93 respectively). Moderately positive correlation was found between #head leaves and head weight (0.30) and between diameter and weight (0.35). Some moderately negative correlations were also observed as well. Head weight is negatively correlated with # leaf and total # leaves (-0.34 and -0.30 respectively).

3.1.2 Brussels sprout

All the phenotyping data for Brussels sprouts were collected from 03 December to 11 December 2018 (208-216 DAS). It was planned to harvest the Brussels sprout plants after 3-4 week of decapitation treatment. But, due to some unavoidable factors like weather, limited cold storage facility etc. it was not possible to harvest at planned time. The harvesting date was 10 days later than the planned day, which was 39 days after the decapitation treatment. There were 49 Brussels sprouts accessions in the field, but one accession did not produce any sprout. So, at the end of the growing season it was possible to collect all the data form 48 accessions.

Like heading cabbages the normality of the data set was checked by using QQ plots with 95 % confidence interval. Data was transformed with log and square root transformation for traits that were not normally distributed. No transformation was needed for plant height and stalk height traits as they were normally distributed. Square root transformation was used for yield gr1, yield gr2, total yield and total market yield. Log transformation was used for yield loss and % yield loss. However, it was not possible to achieve normal distribution even after transformation for the yield gr 3 (Table 5). Some of the accession did not produce larger than 35 mm sprouts, so in yield gr 3 there were some zero values. That was one of the reasons for not achieving normal distribution even after transformation. Specific transformation mode that were used for achieving normal distribution of data are listed in table 5.

Mixed models REML was used to test the hypothesis that the removal of top leaves (decapitation) has an influence on overall yield and marketable yield. It was confirmed that, decapitation has a significant effect on the overall yield of sprout (Table 6, 7). Boxplot of total yield per plant and total marketable yield also show that there is an increase of yield after decapitation (Figure 7). A general overview of decapitation treatment is presented in figure 8. Contribution of different yield groups were also calculated. It was also observed that 6.73% total yield and 7.21% marketable yield is increased after the decapitation treatment (Appendix 6).

Table 6. Total yield per plant

Source of variation	**d.f.**	**s.s.**	**m.s.**	**v.r.**	**F pr.**
Treatment	1	334197	334197	3.95	0.048
Genotype	47	19668239	418473	4.94	<.001
Treatment.Genotype	47	4229726	89994	1.06	0.371
Residual	277	23441546	84627		
Total	372	46552171			

Table 7. Total Marketable yield

Source of variation	**d.f.**	**s.s.**	**m.s.**	**v.r.**	**F pr.**
Treatment	1	283947	283947	4.12	0.043
Genotype	47	16164320	343922	4.99	<.001
Treatment.Genotype	47	3682550	78352	1.14	0.264
Residual	277	19101878	68960		
Total	372	38301595			

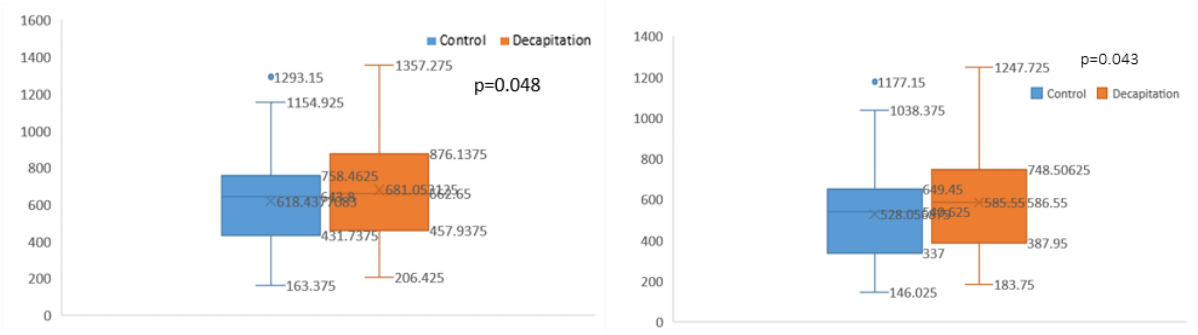

Figure 7. Boxplot of total yield per plant (left), and total marketable yield (right)

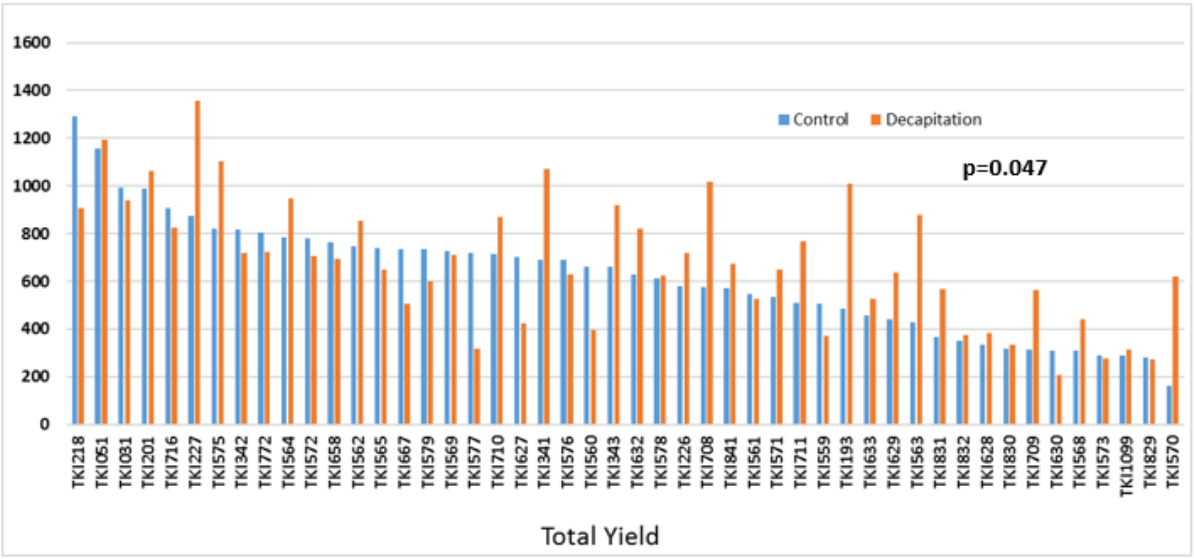

Figure 8. Total yield per accession in response to control and decapitation treatment

Several trait data were collected from the field and the correlation between traits was calculated with Pearson's correlation test (Table 9). A very high positive correlation was observed between stalk height and plant height (0.98), total yield and total marketable yield (0.97). A considerable positive high correlation was also found between yield gr2 and total yield per plant and between yield gr2 and total market yield (0.81 and 0.86 respectively). Similar trend was also found between yield gr3 and total yield per plant and between yield gr3 and total market yield (0.74 and 0.70 respectively). Negative correlations were also observed. Yield gr1 and yield gr3 are negatively correlated (-0.37). An interesting negative correlation was also observed between %yield loss and yield gr2 (-0.37).

Table 8. Correlation between all traits measured for the heading cabbages

Trait	#	Correlation													
Leaf	1	-													
Scars	2	0.08	-												
Total Leaves	3	0.88	0.53	-											
Axillary Shoot gr1	4	0.08	0.31	0.21	-										
Axillary Shoot gr2	5	-0.19	-0.08	-0.20	0.16	-									
Axillary Shoot gr3	6	-0.12	-0.08	-0.13	-0.10	0.44	-								
Total # axillary shoot	7	-0.01	0.23	0.10	0.89	0.54	0.29	-							
Secondary head gr1	8	-0.22	-0.15	-0.25	0.11	0.61	0.48	0.39	-						
Secondary head gr2	9	-0.17	-0.04	-0.17	0.00	0.54	0.49	0.27	0.55	-					
Secondary head gr3	10	-0.05	0.00	-0.04	-0.03	0.27	0.36	0.14	0.18	0.37	-				
Total # secondary head	11	-0.22	-0.12	-0.25	0.08	0.66	0.56	0.39	0.93	0.80	0.37	-			
Head Leaves	12	-0.08	-0.10	-0.11	0.08	0.14	-0.05	0.09	0.10	0.06	0.07	0.10	-		
Diameter	13	-0.09	0.08	-0.03	0.12	0.02	-0.07	0.09	0.00	0.01	0.02	0.01	0.11	-	
Weight	14	-0.34	-0.01	-0.30	0.19	0.07	-0.05	0.17	0.15	0.10	0.08	0.15	0.30	0.35	-
Trait number		1	2	3	4	5	6	7	8	9	10	11	12	13	14

Table 9. Correlation between all traits measured for the Brussels sprout (combined data

Trait	#	Correlation								
Plant_height	1	-								
Stalk_height	2	0.98	-							
Yield gr1	3	0.20	0.21	-						
Yield gr2	4	0.37	0.38	-0.02	-					
Yield gr3	5	0.31	0.26	-0.37	0.29	-				
Total yield per plant	6	0.47	0.45	-0.05	0.81	0.74	-			
Total market yield	7	0.48	0.46	-0.01	0.86	0.70	0.97	-		
Yield loss	8	0.16	0.16	-0.18	0.16	0.43	0.53	0.32	-	
%Yield loss	9	-0.24	-0.21	-0.24	-0.37	-0.03	-0.14	-0.32	0.62	-
Trait number		1	2	3	4	5	6	7	8	9

3.2 POPULATION STRUCTURE

3.2.1 Heading Cabbage

It was not possible to collect all the trait data from all the heading cabbages due to some unavoidable conditions like asynchronous maturity, insect-pest infestation etc. So, Different population structures were calculated for different traits with their relevant number of accessions and with the possible highest number of axes. For axillary shoot, leaf and head leaf related traits 140, 133 and 106 axes were considered respectively for construction of three different PCOs. The first PCO with 140 axes was for axillary shoot related traits and it was able to explain 99.32% of the variation of the studied population (Figure 9). The second PCO was for leaf related traits and with the 133 axes it explains 99.44% variation. The third PCO can explain 99.86% of the variation, which is constructed for head leaves trait and made with 106 axes(appendix 7). The percentage of variation that is explained by different axes can be found in Appendix 8, 9 and 10.

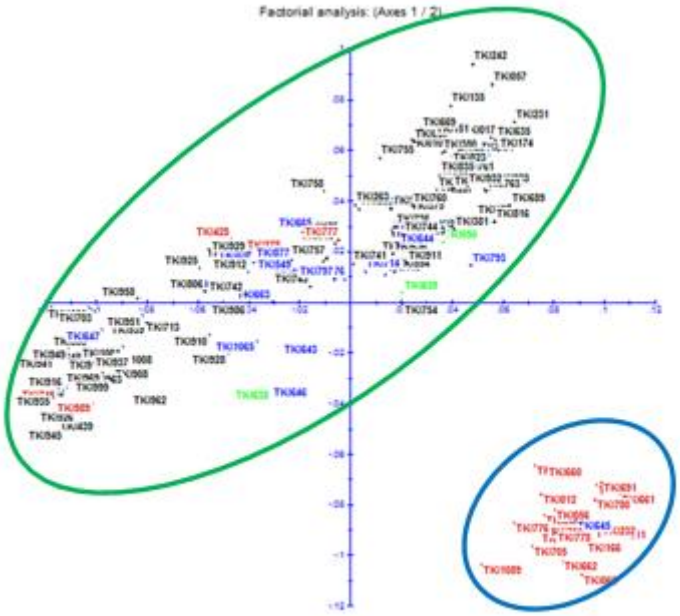

Figure 9. PCO of axillary shoot with 140 accessions. First two axes (x=1, y=2) altogether explain more than 19% of the total variation (12.13% and 6.93%). Different colour labels were used for different sub-morphotypes. Red colours were used for red cabbages, which form a separate group than the other cabbages. The green (gr1) and blue circles (gr2) indicate different groups.

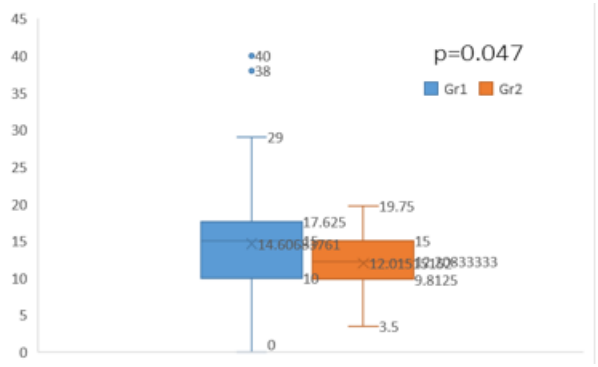

Fig 10. Boxplot of phenotypic variation in total # axillary shoot between gr1 and gr2.

It can be seen from all the PCOs that the red cabbages are clearly different than the other sub-morphotypes of the heading cabbages (Figure 9, Appendix 7). The other three morphotypes tends to form a group with a very few accessions of red cabbages.

3.2.2 Brussels sprout

There were considerably a small group for Brussels sprout, with 48 accessions. A PCO was calculated with 48 axes. The PCO can explain the 99.63% of the total variation of the Brussels sprout population. The percentage of variation that is explained by different axes can be found in Appendix 11. Two different groups were formed in the PCO (Figure 10). The origin was checked but the two groups were not surely for different geographic location. Two sample t-test (a variate with one grouping factor) was performed for all the traits but no significant difference was found between the two groups. We had no information about the maturity time of the sprouts. The two groups may be formed due to the variation of the maturity time.

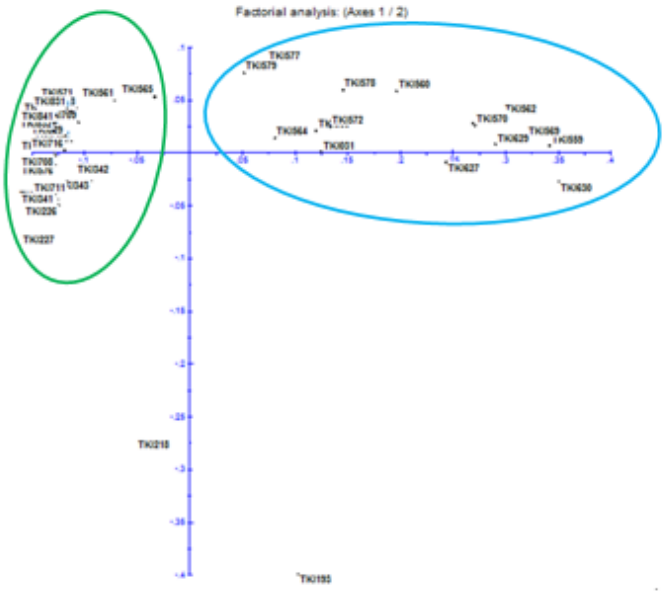

Figure 10. PCO of Brussels sprout with 48 accessions. First two axes (x=1, y=2) altogether explain more than 51% of the total variation (41.31% and 9.75%). The green (gr1) and blue circles (gr2) indicate different groups.

3.3 GWAS

GWAS was performed to study the genetic regulation of the phenotypic traits by using TASSEL. Phenotypic data, genotypic data and different PCOs were used for conducting GWAS. For cabbages three different PCOs were used for GWAS for different data sets. PCO with 140 axes were used for axillary shoot related traits, 133 axes were used for different leaf related traits and 106 axes were used for head leaf trait. Different PCO were considered as there were some missing values for different traits and the PCOs explained more than 99% variation of the studied population. However, for Brussels sprouts only 2 axes were used for GWAS.

3.3.1 Heading Cabbages

Figure 11 represent the calculated QQ plot for leaf and total # leaf traits. It can be seen from the figure that only a few markers have higher LOD scores than the expected values, and most of the markers followed the trend line of the expected values. For every trait different QQ plots were generated which can be found in the appendix 12.

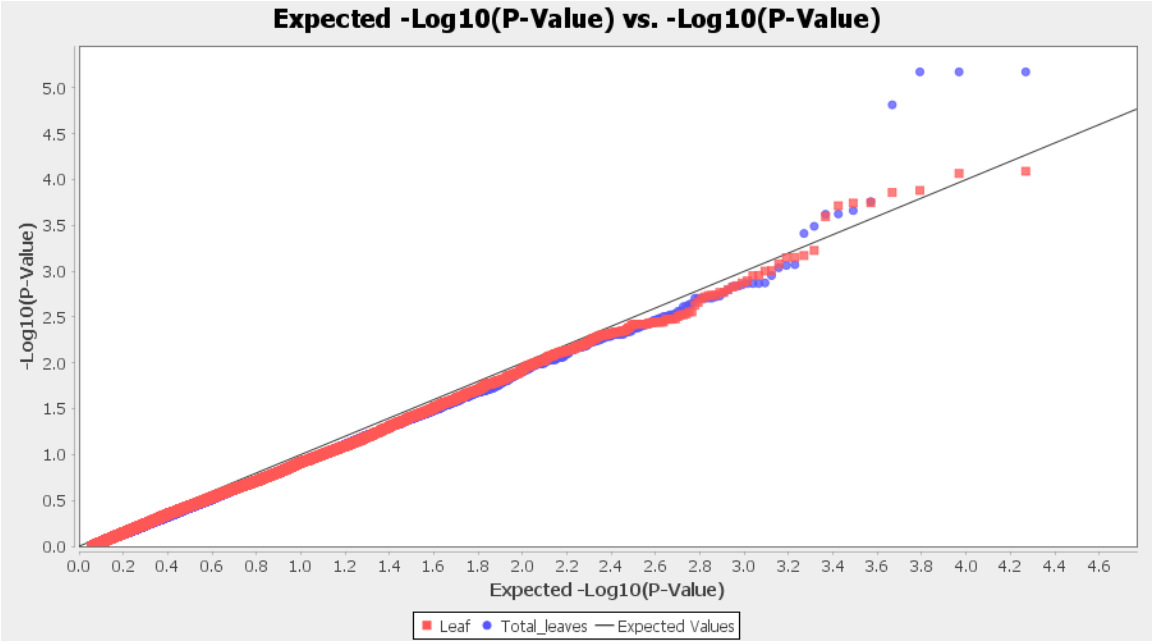

Figure 11. QQ plot of expected vs calculated –Log10(P-value) for leaf and total # leaf with a PCO of 133 axes. The red colour represents leaf and the blue colour represents total# leaf traits.

Manhattan plots were also generated for every trait to visualize the marker trait association. The Manhattan plot for leaf# is presented in figure 12. For this thesis, the threshold level was selected as 4.0. For leaf# a significant SNP in chromosome 2 was observed. TASSEL generated all the Manhattan plots for different traits of cabbages can be found in appendix 13. Several significant SNPs were found and in some cases several physically linked SNPs were also observed.

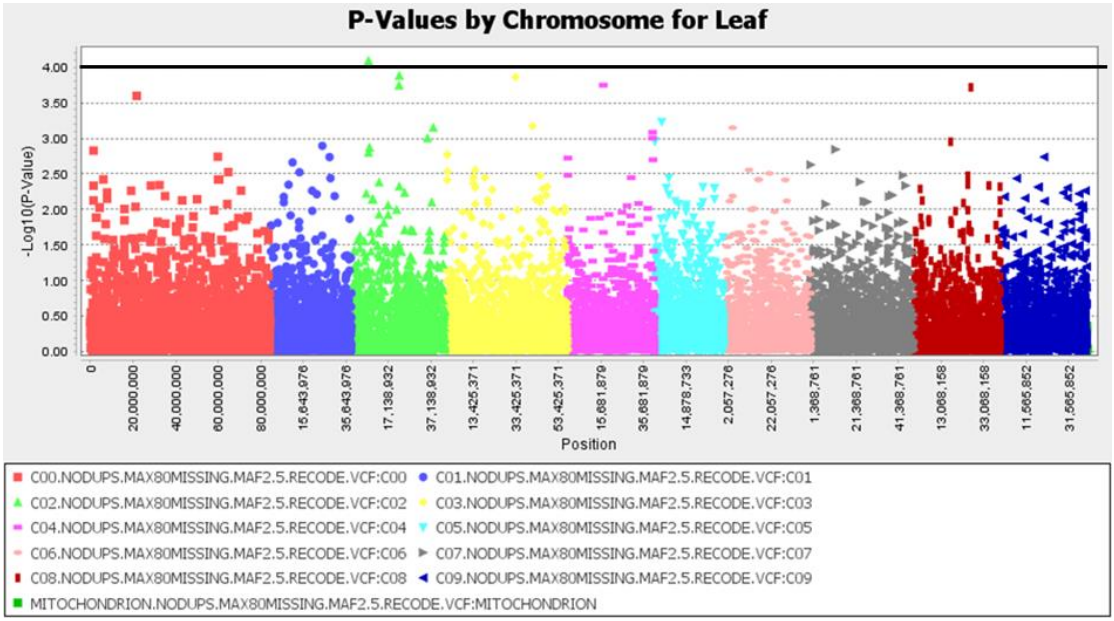

Figure 12. Manhattan plot for leaf#. The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNP above threshold level (LOD= 4.0) can be found only in chromosome 2.

3.3.2 Brussels sprout

The decapitation treatment had significant impact on the overall yield per plant and on total marketable yield. So, GWAS was performed separately for control and decapitated plant data. The phenotypic data, genotypic data and population structure were used for GWAS. The fairly small group with 48 accessions still showed evidence of population structure so a correction was considered for GWAS. QQ plots for different traits that were generated by using TASSEL can be found in the figure 13 and figure 14. A correction with 2 axes was chosen as it likely avoids false positive SNPs while not over correcting, and as such introducing false negative.

Figure 15, represent the calculated QQ plot for different traits of control treatment. Whereas, figure 16 represent the QQ plot for different traits of decapitation treatment. It can be seen from both the figure that a lot of markers have higher LOD scores than the expected values, which shows that the PCO may not correct for all false positives, but it is better than without PCO. As the correction may not be perfect, so, it may lead to yield some false positive SNPs after GWAS in TASSEL.

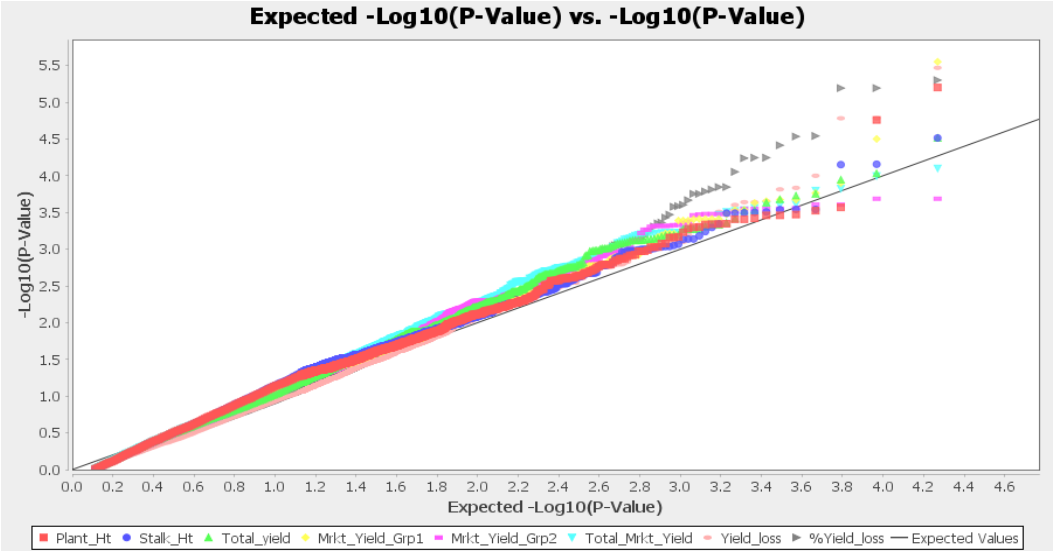

Figure 13. QQ plot of expected vs calculated –Log10(P-value) for different traits (**treatment: control**) of b. sprout with a PCO correction of 2 axes (51.06% explained variation). Different colours represent different traits, see legend.

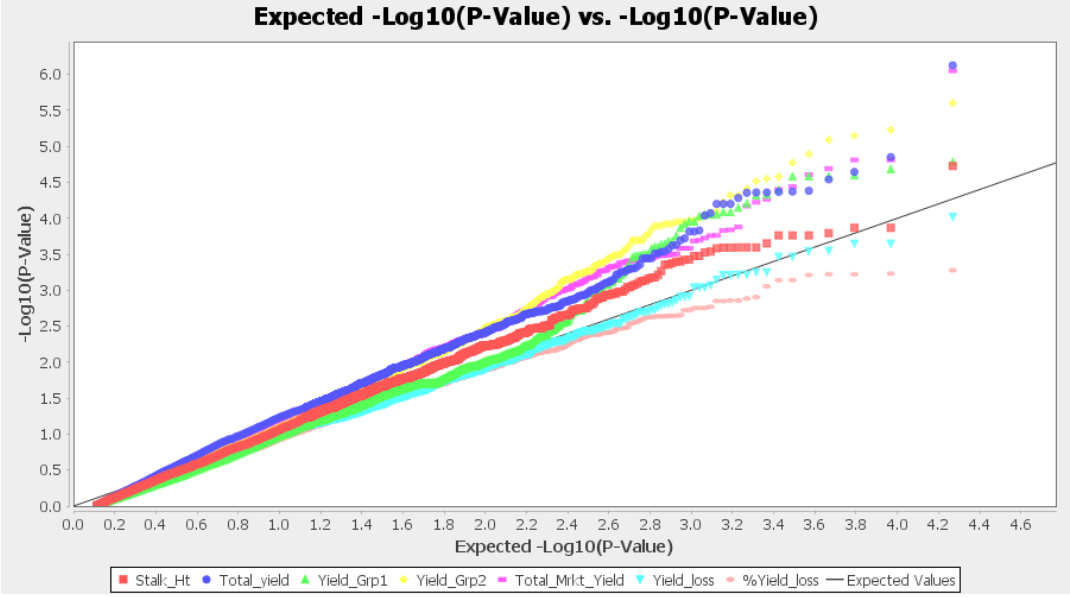

Figure 14. QQ plot of expected vs calculated –Log10(P-value) for different traits (**treatment: decapitation**) of b. sprout with a PCO correction of 2 axes (51.06% explained variation). Different colours represent different traits, see legend.

The Manhattan plot for total yield is presented in figure 15 (control treatment) and figure 16 (decapitation Treatment). For this thesis, the threshold level was selected as 4.0. For Total yield significant SNPs were observed in chromosome 3, 4 and 9 for control treatment, however, for decapitation treatment significant SNPs were found in chromosome 0, 1, 3, 6, 7, 8 and 9. Which proves that the genetic regulation of yield differs between treatments. For all the other Brussels sprout traits Manhattans plots were generated by using TASSEL, which can be found in appendix 14 (control) and appendix 15 (decapitation).

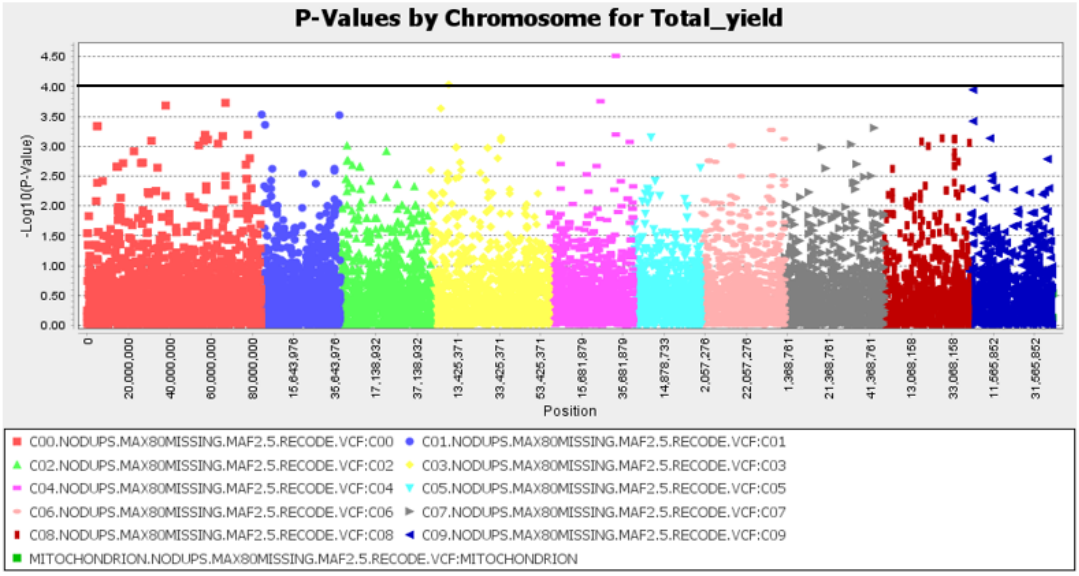

Figure 15. Manhattan plot for Total Yield (**Treatment: Control**). The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNPs above threshold level (LOD= 4.0) can be found in chromosome 3, 4 and 9.

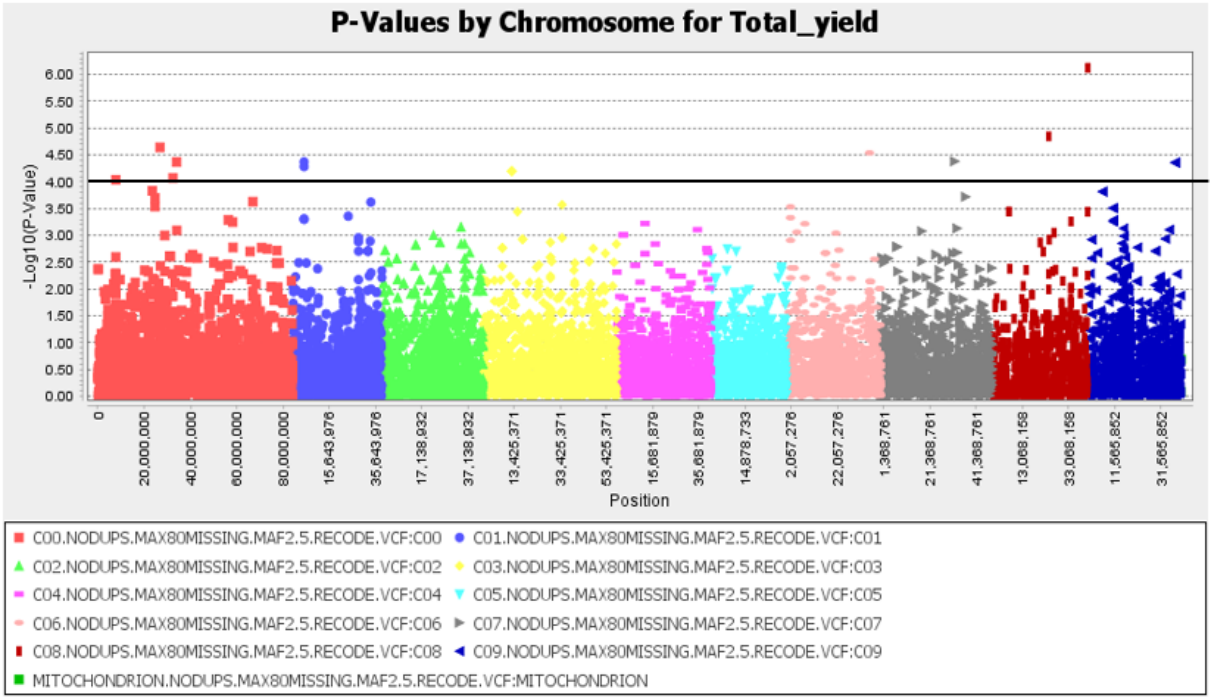

Figure 16. Manhattan plot for Total Yield (Treatment: Decapitation). The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNPs above threshold level (LOD= 4.0) can be found in chromosome 0, 1, 3, 6, 7, 8 and 9.

No significant SNP was found for % yield loss trait in decapitation treatment. But for other traits at least one SNP was found. 11 significant SNPs were found for heading cabbages, whereas for Brussels sprouts 41 and 62 SNPs were found for control and decapitation treatment respectively for different traits. The significant SNPs for Brussels sprouts are listed in Table 10.

Table10. Significant SNPs for different traits for b. sprout

Trait		#SNP	On chromosome
Plant height	Control	5	C01 and C08
Stalk height	Control	7	C01, C02, C04, C08 and 9
	Decapitation	1	C05
Total yield	Control	5	C00, C03, C04 and C09
	Decapitation	16	C00, C01, C03, C06, C07, C08 and C09
Yield gr1	Control	3	C01 and C09
	Decapitation	17	C00, C03, C04, C06 and C09
Yield gr2	Control	6	C03, C07 and C08
	Decapitation	17	C00, C02, C04, C06, C07, C08 and C09
Total market yield	Control	5	C00, C03, C04 and C09
	Decapitation	10	C00, C04, C07, C08 and C09
Yield loss	Control	4	C04 and C09
	Decapitation	1	C07
% Yield loss	Control	6	C01, C03 and C05
	Decapitation	-	-

3.4 PHYSICALLY LINKED SNPS AND SNPS ASSOCIATED WITH DIFFERENT TRAITS

Physically linked SNPs that are associated with same traits or with different traits were found for heading cabbages and for Brussels sprouts. In heading cabbages, physically linked SNPs were found for Leaf # and for total # leaf. In addition, a single SNP was associated with axillary shoot gr1 and total # axillary shoot (Table 11). For Brussels sprout, C09 seems to be important as it hosts many significant SNPs for different traits. Higher # of physically linked SNPs were found in Brussels sprout and some SNPs were also found that are associated with different traits as well. Table 12 and 13 represents the physically linked SNPs and SNPs that are associated with different traits for control and decapitation treatments of Brussels sprout.

Table 11. Significant SNPs associated with different traits, with their position on chromosome, LOD score and indication whether they are physically linked or associated with multiple traits for heading cabbages

Trait	Position	LOD	Chromosome										Remarks
			Unknown	1	2	3	4	5	6	7	8	9	
Head leaves	64590817	4.12	Unknown										
	17826699	4.04							6				
Leaf#	7830797	4.09			2								Physically linked
	7830770	4.07			2								
	9840886	5.18			2								Physically linked
	9840922	5.18			2								
	9840939	5.18			2								
Total # Leaf	22916038	4.82						5					
Axillary shoot gr1	**4440955**	5.72							6				Physically linked and associated with Axillary shoot gr1 and total # axillary shoot
Total # Axillary shoot	**4440955**	4.20							6				
	6228345	4.05		1									

Table 12. Significant SNPs associated with different traits, phenotyped in the collection of 48 Brussels sprouts, with their position on chromosome, LOD score and indication whether they are physically linked or associated with multiple traits for Brussels sprout (treatment: control)

Trait	Position	LOD	Chromosome										Remarks
			Unknown	1	2	3	4	5	6	7	8	9	
Plant Ht	29540525	4.10									8		Physically linked and same SNPs for plant ht, stalk ht and % yield loss (Position: 4808802)
	4808802	6.07		1									
	4808793	5.59		1									
	4808859	4.17		1									
	4808895	4.17		1									
Stalk Ht	**4808802**	5.34		1									
	4808793	4.97		1									
	24431048	4.55			2								
	33857294	4.28									8		
	6749478	4.20					4						Physically linked
	6749503	4.20					4						
	1070840	4.12										9	
Total yield	32336474	4.77					4						
	38063125	4.50	Unknown										
	67104184	4.43	Unknown										
	1594987	4.14										9	
	5027460	4.06				3							
Yield Grp1	29749078	5.32										9	Physically linked
	29749173	4.41										9	
	34175792	4.11		1									

Yield Grp2	5027460	4.29				3							
	46568493	4.07								7			
	37884795	4.05									8		Physically linked
	37884813	4.05									8		
	37884858	4.05									8		
	37884870	4.05									8		
Total Mrkt Yield	1594987	4.25										9	
	5027460	4.25				3							
	38063125	4.14	Unknown										
	24841089	4.10					4						
	9849598	4.09										9	
Yield loss	32336474	5.45					4						
	28726005	4.78										9	Physically linked
	28726031	4.78										9	
	28726101	4.00										9	
%Yield loss	3789194	5.30		1									
	723301	5.20				3							
	4808859	4.65		1									Physically linked and same SNP for plant ht, stalk ht and %yield loss (Position: 4808802)
	4808895	4.65		1									
	4808802	4.41		1									
	31762477	4.01						5					

Table 13. Significant SNPs associated with different traits, phenotyped in the collection of 48 Brussels sprouts, with their position on chromosome, LOD score and indication whether they are physically linked or associated with multiple traits for Brussels sprout (treatment: decapitation)

Trait	**Position**	**LOD**	**Chromosome**										**Remarks**
			Unknown	1	2	3	4	5	6	7	8	9	
Stalk Ht	6926291	4.73						5					
Total yield	41437911	6.12									8		
	24460367	4.85									8		
	26834050	4.64	Unknown										
	35902328	4.54							6				
	32140232	4.38								7			
	4808793	4.37		1									
	34059446	4.37	Unknown										
	37828582	4.36										9	Physically linked and same SNPs for Total Marketable yield
	37828596	4.36										9	
	37828587	4.36										9	
	4808802	4.28		1									
	12463416	4.20				3							Physically linked
	12463479	4.20				3							
	12463485	4.20				3							
	32432008	4.07	Unknown										
	7664591	4.04	Unknown										

Yield Gr1	20136867	4.78										9	
	9650062	4.68	Unknown										
	11032147	4.60				3							
	15378003	4.58										9	Physically linked
	15378032	4.58										9	
	15378033	4.58										9	
	1296469	4.40										9	
	68582022	4.33	Unknown										
	27707480	4.32							6				
	40972846	4.21				3							
	36081985	4.15					4						
	21597735	4.09										9	Physically linked
	21597791	4.09										9	
	24639032	4.05										9	Physically linked
	24639039	4.05										9	
	24639041	4.05										9	
	27707508	4.03							6				
Yield Gr2	35188372	5.60										9	
	28166275	5.23					4						
	36884892	5.15								7			
	23544547	5.09	Unknown										
	27544961	4.89									8		
	1730972	4.77							6				Physically linked
	1731037	4.41							6				
	12384325	4.58					4						
	34093019	4.55	Unknown										
	31308769	4.52								7			
	25124351	4.32			2								Physically linked
	25124352	4.32			2								
	22357634	4.22			2								
	62885802	4.10	Unknown										
	42996147	4.07								7			
	65879164	4.04	Unknown										
	32140232	4.01								7			
Total Mrkt Yield	41437911	6.05									8		
	37828582	4.81										9	Physically linked
	37828596	4.81										9	and same SNPs
	37828587	4.69										9	for total yield
	32140232	4.61								7			
	24460367	4.44									8		
	23544547	4.40	Unknown										
	12384325	4.27					4						
	34059446	4.22	Unknown										
	36884892	4.17								7			
Yield loss	25335605	4.02								7			

3.5 CANDIDATE GENES

Significant SNPs that had high LOD scores for every trait were entered into the *brassica* genome browser (BRAD, 2019) to find putative genes. Based on the function of *A. thaliana* orthologous genes, that are involved in growth and development, hormonal pathway, signalling pathway and in stress response were selected. The candidate genes that are found in this study is listed in table 14.

Table 14. Candidate genes for different traits of heading cabbages and Brussels sprout (source: BRAD, 2019 and TAIR, 2019)

Trait	Chr.	Position	LOD	Bol ID and best hit	Function
Heading cabbages					
Head leaves	C06	17826699	4.04	1. Bol019422 (AT5G62880 2. Bol019423 (AT5G25350) 3. Bol019425 (AT4G36140)	1. negative regulation of abscisic acid-activated signaling pathway, plant-type cell wall organization 2. negative regulation of ethylene-activated signaling pathway 3. defense response, signal transduction
Leaf#	C02	7830797	4.09	Bol015474 (AT5G59550)	protein auto ubiquitination, protein ubiquitination, response to abscisic acid,response to chitin, response to water deprivation
	C02	7830770	4.07		
Total # Leaf	C02	9840886	5.18	Bol014258 (AT5G56970)	It encodes a protein whose sequence is similar to cytokinin oxidase/dehydrogenase, which catalyzes the degradation of cytokinins.
	C02	9840922	5.18		
	C02	9840939	5.18		
Axillary shoot gr1	C06	**4440955**	5.72	Bol016157 (AT1G71090)	auxin homeostasis, auxin polar transport, auxin-activated signaling pathway, lateral root formation, regulation of growth rate, response to auxin, transmembrane transport
Total # Axillary shoot	C06	**4440955**	4.20		
	C01	6228345	4.05	1. Bol019713 (AT4G17870) 2. Bol019715 (AT4G17880)	1. abscisic acid-activated signaling pathway 2. It functions as a JAZ-interacting transcription factor that acts together with MYC2 and MYC3 to activate JA-responses. It also functions in blue light mediated secondary cell wall biogenesis via regulation of NST1 expression. MYC4 directly binds to NST1 promoter and activates its expression.
Brussels sprout (control)					
Plant height	C01	**4808802**	6.07	1. Bol020933 (AT4G29720) 2. Bol020942 (AT4G29830)	1. regulation of timing of transition from vegetative to reproductive phase 2. negative regulation of flower development
	C01	4808793	5.59		
	C01	4808859	4.17		
	C01	4808895	4.17		
Stalk height	C01	**4808802**	5.34		
	C01	4808793	4.97		
	C08	33857294	4.28	Bol045740 AT2G25680	Encodes a high-affinity molybdate transporter. Mutant has reduced concentrations of molybdate in roots and shoots, and reduced shoot and root length when growing on Mo-limited medium.
Total yield	C04	32336474	4.77	Bol033271 (AT2G29125)	Shoot system development

	C09	1594987	4.14	1. Bol032157 (AT4G03400) 2. Bol032164 (AT4G03440)	1. response to auxin, response to light stimulus 2.Salysilic acid defense Shoot system development
	C03	5027460	4.06	Bol025885 (AT5G60340)	developmental vegetative growth, nucleotide phosphorylation, regulation of growth, unidimensional cell growth
Yield Grp1	C09	29749078	5.32	1. Bol035782 (AT5G20900) 2. Bol035775 (AT5G20950)	1.defense response, regulation of defense response, regulation of jasmonic acid mediated signaling pathway, response to wounding 2. Carbohydrate metabolic process
	C09	29749173	4.41		
	C07	46568493	4.07	1. Bol033825 (AT5G51470) 2. Bol033806 (AT4G33210) 3. Bol033816 (AT4G33430)	1. brassinosteroid mediated signaling pathway, cell death, defense response,defense response to bacterium, defense response to fungus, defense response to oomycetes, protein phosphorylation 2. Encodes SLOMO (SLOW MOTION), a F-box protein required for auxin homeostasis and normal timing of lateral organ initiation at the shoot meristem 3 Auxin responsive gene
	C08	37884795	4.05	Bol031493 (AT1G14350)	Guard cell differentiation Auxin transport Stress
	C08	37884813	4.05		
	C08	37884858	4.05		
	C08	37884870	4.05		
Total Mrkt Yield	C09	1594987	4.25	Bol032157 (AT4G03400)	Encodes a GH3-related gene involved in red light-specific hypocotyl elongation.
	C03	5027460	4.25	Bol025880, Bol025881, Bol025885 (AT5G60340)	Encodes a nuclear adenylate kinase that interacts with a putative homolog of Rps14, AtRPS14-1 and affects the elongation of cells in the stem. regulation of growth, unidimensional cell growth
Yield loss	C04	32336474	5.45	---	---
	C09	28726005	4.78	---	---
	C09	28726031	4.78	---	---
	C09	28726101	4.00	---	---
%Yield loss	C01	3789194	5.30	1. Bol017973 (AT4G31610) 2. Bol017977 AT1G32330 3. Bol017978 (AT4G31610) 4. Bol017979 (AT4G31620) 5. Bol017980 and Bol0017981 (AT4G31610 6. Bol017985 AT4G31550	1&3,5 Expressed specifically in reproductive meristems, member of a moderately sized gene family distantly related to known plant DNA binding proteins, flower development, regulation of transcription, DNA-templated 2. Member of Heat Stress Transcription Factor (Hsf) family. Negatively regulated by HSP90.2. 4. Reproductive meristem formation 6. member of WRKY Transcription Factor; Group II-d; negative regulator of basal resistance to Pseudomonas syringae defense response to bacterium, induced systemic resistance, regulation of jasmonic acid mediated signaling pathway, regulation of transcription, DNA-templated, response to bacterium, response to chitin
	C03	723301	5.20	Bol015379 (AT5G04770)	amino acid transmembrane transport, response to nematode, transmembrane transport

Brussels sprout (decapitation)					
Trait	**Chr.**	**Position**	**LOD**	**Name**	**Function**
Total yield	C08	41437911	6.12	1. Bol018499 (AT1G02090) 2. Bol018500 (AT1G02100)	1. Photomorphogenic phenotype 2. negative regulation of brassinosteroid mediated signaling pathway
	C09	**37828582**	4.36	1. Bol043819 (AT5G08130) 2. Bol043820 (AT5G08120) 3. Bol043802 (AT5G08290) 4. Bol043802 (AT5G08280)	1. Shade avoidance 2. Water deprivation response & host response against virus 3. Leaf senescence 4.Defense response bacteria
Yield Gr1	C09	20136867	4.78	1. Bol030298 (AT5G53290) 2. Bol030297 (AT4G22910)	1. Cotyledon development, cytokinin-activated signaling pathway, ethylene-activated signaling pathway, leaf development, regulation of transcription, DNA-templated 2. DNA endoreduplication, cell division, multidimensional cell growth, positive regulation of ubiquitin protein ligase activity, protein ubiquitination, signal transduction, trichome branching
Yield Gr2	C09	35188372	5.60	1. Bol043397 (AT5G13910) 2. Bol043396 (AT5G13930)	1. Ethylene-activated signaling pathway, gibberellic acid mediated signaling pathway, positive regulation of seed germination, regulation of transcription, DNA-templated, response to gibberellin 2. Auxin polar transport, chalcone biosynthetic process, flavonoid biosynthetic process, regulation of anthocyanin biosynthetic process, response to UV-B, response to auxin, response to gravity, response to jasmonic acid, response to oxidative stress, response to wounding
	C04	28166275	5.23	1. Bol009985 (AT4G13980) 2. Bol009982 (AT1G05470)	1. Cellular response to heat, positive regulation of transcription from RNA polymerase II promoter in response to heat stress, regulation of transcription, DNA-templated 2. Abscisic acid-activated signaling pathway, cell differentiation, cotyledon vascular tissue pattern formation, inositol phosphate dephosphorylation, inositol phosphate-mediated signaling, inositol trisphosphate metabolic process, leaf vascular tissue pattern formation, phosphatidylinositol dephosphorylation, procambium histogenesis, response to abscisic acid, xylem and phloem pattern formation
	C07	36884892	5.15	1. Bol017077 (AT5G23080) 2. Bol017070 (AT5G23010)	1. Mutants display developmental defects, including reduced plant height, polycotyly, and reduced vascularization 2. response to insect, response to water deprivation

Total Mrkt Yield	C08	41437911	6.05	-----	-----
	C09	**37828582**	4.81	1. Bol043819 (AT5G08130) 2. Bol043820 (AT5G08120) 3. Bol043802 (AT5G08290) 4. Bol043802 (AT5G08280)	1. Shade avoidance 2. Water deprivation response & host response against virus 3. Leaf senescence 4.Defense response bacteria
	C09	37828596	4.81		
	C09	37828587	4.69		
	C07	32140232	4.61	-------	----------
Yield loss	C07	25335605	4.02	Bol017164 (AT5G53170)	PSII associated light-harvesting complex II catabolic process, proteolysis,response to heat

4 DISCUSSION AND RECOMMENDATIONS

Variation in traits of Brussels sprout and cabbages was evaluated at harvesting stage. The collected data were used to understand the variation in these traits and to use it as input of GWAS. The other students of growth and development group worked on mainly leaf morphology and heading traits from the same field experiment of 2018 (Table 2). They harvested the best three uniform cabbages from each accession earlier for collecting different trait data (Mortel, 2018). So, the worst two cabbages from each accession were in the field. There was a diverse collection of heading cabbages in terms of maturity, origin and in sub-morphotypes. Some of the accessions started early flowering, whereas some accessions didn't produce a head during data collection period. Differences in insect pest tolerance was also observed and some of the accessions were found completely damaged. Due to all these causes it was not possible to collect data from all the accessions and in some of the cases data were less reliable. Leaf, head leaf, axillary shoot and secondary head related trait data were collected from the field for heading cabbages. In some cases, it was not possible to collect data from both blocks for an accession. For Brussels sprout, plant height, stalk height and different yield traits were considered. Only one accession of Brussels sprout didn't produce any sprout, which was not considered for this study.

The measurement of the cabbage traits was time intensive, and it took 31 days to collect all the data. As a result, the cabbages of the last accessions got 31 more days to grow than the first harvested ones. There were two blocks and data collection were started from block A which took 17 days, then the data from block B were collected which took 14 days. So, the cabbages of block B got at least 18 more days for growth. To test whether this caused a block effects, ANOVA were run and interestingly most traits that were considered for GWAS was not affected by block (appendix 16). The only trait that showed a block effect was total # leaf. Then the mean value of each accession was considered for each block and ranked them separately from low to high and 6 accession were matched within the first ten accessions (Appendix 04). Block B is closer to the river than the block A, so there may have some difference in water gradient which may have an influence on the total # leaf. But, saying so, would be an overestimation as the two blocks were adjacent to each other. However, for this new crop season I will suggest to make vertically align blocks to the river.

Pearson's correlation test was conducted to see the correlation between different traits. Several highly correlated traits were found. For heading cabbages, the head weight is a very important trait. Moderately negative correlation was found between leaf # and head weight (-0.34) and total # leaf and head weight (-0.30). Almost similar negative correlations were also observed (-0.3 and -0.2) in a last year thesis (Mortel, 2018). Head leaves and diameter contributes to the weight of the cabbages, so highly positive correlation was expected. But moderately positive correlations were observed between head leaves and weight (0.30), and diameter and weight (0.35). All the head leaves were recorded, however, the smaller leaves near to shoot apical meristem (top of the core) were tiny and had fairly low weight in comparison to the outer large leaves, which likely resulted in moderate positive correlations between head leaves and weight (0.30). But, it can still be improved by stop counting head leaf after reaching at a certain size (e.g. smaller than 3 cm) for all the accessions.

Generally, it was observed in the field that the larger heads have greater weight. But, some of accessions were found with less weight and higher diameter. For this reason, moderately positive correlations were observed between diameter and weight (0.35). A new trait, density, can be considered in the next year, which would give a better impression of the correlation between #head leaf, diameter and weight of cabbage.

For Brussels sprout three different sets of correlations were conducted. The first one with the combined data, second one with the data of control treatment and the third one with the decapitation treatment data (Table 9 and Appendix 17). Almost similar values were found between different traits in the all three-correlation test. So, the correlation between different traits, with the combined data is discussed here. Some highly positive correlations were observed between different traits. Generally, supermarkets sell 21-32 mm size sprouts, which is comparable to the yield group 2 (25.1-35 mm). So, yield group 2 was considered as an important trait and correlations with all other traits were observed carefully. It was found that yield gr2 is highly positively correlated with total yield per plant and total marketable yield (0.81 and 0.86 respectively). Moderately positive correlations were observed between plant height and yield gr2 (0.37) and between stalk height and yield gr2 (0.38). So, it can be assumed that larger plants facilitate higher yield of Brussels sprouts by containing more sprouts throughout the plant.

For this thesis, population structure was corrected with PCO as it is more efficient than STRUCTUTRE in terms of time requirement and in visualizing different groups and outliers. A PCO is more powerful when it is tested with the highest number of axes as the higher no. of axes explain more variation in a population. Three different PCOs was calculated for three different population of cabbages with the highest possible no. of axes and each of the PCOs successfully explained more than 99% of the variation within the population. For Brussels sprout A PCO calculation was conducted with 48 accessions and it also explained more than 99% of the variation of the studied population.

Two distinct group was found after PCO correction for heading cabbage. The smaller group for cabbages is composed of red cabbages with one savoy cabbage accession. Significant differences were found for total # axillary shoot and for # head leaves, between the small and large groups. However, no significant differences were observed for #leaf, total # leaf and axillary shoot gr1. Two separate group was also observed for the Brussels sprout population. These are not separated according to hybrid-no hybrid, origin etc. May be the groups are formed due to the variation of maturity or harvesting time.

For GWAS; phenotypic, genotypic and population structure were considered for finding significant markers for different traits. As a strong correction for population structure was applied, we likely reduced the false positive SNPs. But for reducing false positive SNPs in TASSEL, 999 permutations were considered for all the dataset. PCO was strictly corrected and used with the optimum no. of axes for conducting GWAS. Moreover, a comparatively high FDR threshold level (LOD 4.0) was considered, which also helped to reduce the no. of false positive SNPs. But this strict correction may also lead to lose some important false negative SNPs, that may have been correlated to the population structure and may have associated to a trait variation.

For heading cabbages only ten different SNPs were found above the threshold level which can possibly be associated with the targeted traits. For Brussels sprout 40 SNPs were found in control treatment and 59 SNPs were found in decapitation treatment above FDR threshold. Some physically linked SNPs were found and interestingly, some SNPs were also found which were associated with more than one trait. Moreover, the traits share a common SNP are found to be highly correlated in some cases (e.g. plant height and stalk height (0.98), and total yield and total marketable yield (0.97)). The significant SNPs were searched in the previous thesis, but no common SNPs were found. It may be due to the traits that were considered for this thesis was different than the traits of the other thesis students, as they basically focused on earlier growth stages of cabbages.

At first, it was planned to search for nearby candidate genes for all the SNPs that are above the threshold level. But due to time constraint it was not possible to search for possible candidate genes for all the SNPs. So, at least few candidate genes for every trait were searched around the comparatively higher LOD score bearing SNPs in the *Brassica* genome browser (BRAD, 2019). Like some previous master student LD of 50 Kb was considered for this thesis (Alam, 2018, Islam, 2017, Topper, 2016, Zou, 2019). Cheng et al., 2016 calculated a LD of 36.8 Kb for a brassica collection and Mortel, 2018 estimated a LD of 50 Kb for heading cabbages from the same article. So, while searching for candidate genes a region of 100 Kb (50 Kb both way of a SNP) were searched in the browser.

In almost every time more than ten genes were found around the SNPs while searching in *Brassica* genome browser in a window of 100 Kb. Most of the genes had their orthologues, which were further searched in *Arabidopsis* genome browser (TAIR, 2019). Then the function of the gene, involvement in biological process, location of expression etc. were critically observed. After that, the genes that were found to be involved in different hormonal pathway, signalling pathway for biotic or abiotic stresses, leaf development and plant structure development were selected as potential candidate genes. Then the genes were further searched to see if the literature supports the function of the genes.

In case of decapitation treatment of Brussels sprout, C09_37828582, C09_37828587 and C09_37828596 SNPs were found to be physically linked, moreover, these three SNPs were associated to total yield and total marketable yield trait. Four interesting genes *AT5G08130*, *AT5G08120*, *AT5G08290* and *AT5G08280* were found that can possibly be associated with the traits. *AT5G08130* is reported to be involved in shade avoidance, *AT5G08290* for leaf senescence, *AT5G08120* for water deprivation response & host response against virus and *AT5G08280* for defense response against bacteria (TAIR, 2019). During shade, plant want to get more light and it becomes taller to capture more light, gibberellic acids (GA) plays important role for shoot elongation. Shade avoidance is an important mechanism in term of reducing the gibberellic acids (GA) concentration as the higher concentration stimulates conversion of dormant axillary buds into shoot and decreases the quality of sprout (Crocco et al., 2015). Leaf senescence is a common phenomenon for Brussels sprout, which may also help in yield as the older leaves can act as sink of the produced food rather than acting as source. The other two genes are also important yield traits as these genes can potentially help the plant to survive against harsh condition.

GWAS for Brussels sprouts without decapitation treatment (control) resulted in some marker trait associations with interesting potential candidate genes. For stalk height in C08_33857294 SNP, *AT2G25680* gene was found that have been reported as high affinity Molybdate transporter and mutant show reduced shoot length. For total yield, in C03_5027460, AT5G60340 were found which is involved in regulation and development of vegetative growth. C01_4808802, SNP was found for plant height, stalk height and % yield loss trait. *AT4G29720* and *AT4G29830* genes were found nearby of the SNP. *AT4G29720* gene was reported as regulator of transition from vegetative to reproductive phase and *AT4G29830* for negative regulator of flower development. So, these genes can also be considered as important candidate genes (TAIR, 2019).

For heading cabbages several genes were found and most of the genes are related with different hormonal pathway and with defence response. The hormonal pathways and their interplay should be critically judged to understand the mechanism of axillary shoot development. However, I think that the axillary shoots and the secondary heads are developed in the later stage of the plant growth of the heading cabbages. The heading cabbages may have deliberately produced the axillary shoot and sent the excess amount of assimilates to those shoots to maintain the entirety of the main head and wait until the proper environmental signal, so that it can grow flower and continue its lifecycle. So, the initiation time of the axillary shoots should have to be observed critically to understand the mechanism why heading cabbage invest energy to the axillary shoots and secondary heads.

The potential candidate genes that are found during this thesis need to be validate by molecular confirmation. Some previous thesis student developed Cleaved Amplified Polymorphic Sequence (CAPS) marker to confirm the genes (Baez, 2018; Pirzada, 2018). However, multi-location or multi-year data can also give a good insight about the associated markers. Several omics approaches (transcriptome, epigenome, metablome, proteome, etc.) can also be practiced to observe the association of a candidate gene to the target trait.

In conclusion, it was found that for different set of accessions of the main population it is better to conduct separate PCOs that can be used for population structure for GWAS as it can explain the appropriate amount of variation in the concerned population. The heading cabbages may have produced the axillary shoots and secondary heads for their own benefit. Decapitation of Brussels sprouts, several weeks prior to harvest, can significantly increase the yield and total marketable yield. Genetic regulation of yield differs between treatment as different significant marker trait association was found in different treatments. Moreover, GWAS led to find out several potential candidate genes which are associated with different traits.

4.1 RECOMMENDATIONS FOR FURTHER STUDY

1. Timing of axillary shoot initiation should be recorded for heading cabbages
2. Some small experiments can be conducted to see whether the heading cabbages delibarely produce axillary shoots or not, by removing the axillary shoots and observing the impact on entire head
3. If possible multi-location trial or trial for several years or trial with more replications with larger plots should be conducted to see the impact of decapitation treatment of Brussels sprout
4. Timing of decapitation should have to be optimized on the basis of maturity time
5. To observe a strong correlation between head leaves and yield of cabbages, some parameters should have to fixed (e.g. counting larger than three cm leaves within the head, density of the head should be considered etc.)
6. More traits that may contribute to the yield of Brussels sprouts should have to be considered (e.g. leaf number, overall distribution of the sprouts, density of sprout in the stalk, harvest index, time of leaf dropping, trend of leaf dropping, #of green leaves in the stalk during harvesting etc.)
7. The candidate genes that are found in this thesis need to be validated.

5 REFERENCES

Alam, Z. (2018). *Genetic regulation of variation of leaf traits of a Brassica oleracea collection: an association mapping approach*. Wageningen University Thesis.

Anderson, M., & Braak, C. T. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of statistical computation and simulation*, 73(2), 85-113.

Báez, J. A. (2018). *Approach in the development and evaluation of DNA molecular markers in Brassica species.* Wageningen University Thesis.

Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R., & Mathews, S. (2010). Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *107*(43), 18724-18728.

Bonnema, G., Carpio, D. P. D., Zhao, J. J., Kole, C., & Sadowski, J. (2011). Diversity analysis and molecular taxonomy of Brassica vegetable crops. *Genetics, Genomics and Breeding of Crop Plants*. Science Publishers, Enfield, NH, 81-124.

BRAD. (2019). Brassica Genome Browse. Retrieved March 1, 2019, from http://brassicadb.org/cgi-bin/gbrowse/Oleracea/

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome

evolution by phylogenetic analysis of chromosomal duplication events. *Nature, 422*(6930), 433.

Brouwer, H. (2018). *Genetic Analysis on Brassica oleracea breeding traits*. Wageningen University Thesis.

Cartea, M. E., Lema, M., Francisco, M., & Velasco, P. (2011). Basic information on vegetable Brassica crops. *Genetics, Genomics and Breeding of Vegetable Brassicas*, 1–33. https://doi.org/doi:10.1201/b10880-2\r10.1201/b10880-2

Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., ... & Liu, D. (2016). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in Brassica rapa and Brassica oleracea. *Nature genetics*, *48*(10), 1218.

Crocco, C. D., Locascio, A., Escudero, C. M., Alabadí, D., Blázquez, M. A., & Botto, J. F. (2015). The transcriptional regulator BBX24 impairs DELLA activity to promote shade avoidance in Arabidopsis thaliana. *Nature communications*, *6*, 6202.

Del Carpio, D. P., Basnet, R. K., De Vos, R. C. H., Maliepaard, C., Paulo, M. J., & Bonnema, G. (2011). Comparative Methods for Association Studies: A Case Study on Metabolite Variation in a Brassica rapa Core Collection. *PLoS ONE*, *6*(5), 1–10. https://doi.org/10.1371/journal.pone.0019624

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Groot, T. (2016). *Genetic analysis of heading cabbage traits*. Wageningen University Thesis.

Islam, M. (2017). *Association Study on Leaf Morphology in a Diverse Collection of Brassica Oleracea*. Wageningen University Thesis.

Kalloo, G., and B. O. Bergh. 1993. *Genetic Improvement of Vegetable Crops*. 1st ed. Oxford: Pergamon Press.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*(29), 1–9.

Maggioni, L., von Bothmer, R., Poulsen, G., & Branca, F. (2010). Origin and domestication of cole crops (Brassica oleracea L.): linguistic and literary considerations1. *Economic botany*, *64*(2), 109-123.

Mortel, P. (2018). *Genetic analysis of leaf and head morphology traits in heading cabbage (B. oleracea)*. Wageningen University Thesis.

Nagaharu, U. (1935). Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. *Jpn J Bot*, *7*(7), 389-452.

Nishijima, T., & Fukino, N. (2006). Autonomous Development of Erect Leaves Independent of Light Irradiation during the Early Stage of Head Formation in Chinese Cabbage (*Brassica rapa* L. var. pekinensis Rupr.). *Journal of the Japanese Society for Horticultural Science, 75*(1), 59–65. https://doi.org/10.2503/jjshs.75.59

Pang, W., Li, X., Choi, S. R., Dhandapani, V., Im, S., Park, M. Y., … Lim, Y. P. (2015). Development of a leafy Brassica rapa fixed line collection for genetic diversity and population structure analysis. *Molecular Breeding*, *35*(54), 1–15. https://doi.org/10.1007/s11032-015-0221-9

Pirzada, R. A. (2018). *Development of molecular markers for candidate gene profiling in Brassica oleracea for leaf development traits*. Wageningen University Thesis.

Smyth, D. R. (1995). Flower Development: Origin of the cauliflower. Current Biology, 5(4), 361-363. http://dx.doi.org/10.1016/S0960-9822(95)00072-8.

Stromberg, J. 2015. “Kale, Brussels sprouts, cauliflower, and cabbage are all varieties of a single magical plant species.” https://www.vox.com/xpress/2014/8/6/5974989/kale-cauliflower-cabbage-broccoli-same-plant/ (September 05, 2018).

Sun, X. X. (2018). *Morphological and genetic characterization of the leafy head of Chinese cabbage (Brassica rapa)*(Doctoral dissertation, Wageningen University).

Topper, F. (2016). *Association Mapping of Leaf Morphology Traits in Brassica oleracea*. Wageningen University Thesis.

Town, C. D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., ... & Vigouroux, M. (2006). Comparative genomics of Brassica oleracea and Arabidopsis thaliana reveal gene loss, fragmentation, and dispersal after polyploidy. *The Plant Cell, 18*(6), 1348-1359.

Yaish, M. W. F., Guevara, D. R., El-Kereamy, A., & Rothstein, S. J. (2010). Axillary shoot branching in plants. In *Plant Developmental Biology-Biotechnological Perspectives* (pp. 37-52). Springer, Berlin, Heidelberg.

Yu, J., Zhao, M., Wang, X., Tong, C., Huang, S., Tehrim, S., … Liu, S. (2013). Bolbase: A comprehensive genomics database for Brassica oleracea. *BMC Genomics*, *14*, 1–7. https://doi.org/10.1186/1471-2164-14-664

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., ... & Nordborg, M. (2007). An Arabidopsis example of association mapping in structured samples. PLoS genetics, 3(1), e4.

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome Journal*, *1*(1), 5–20. https://doi.org/10.3835/plantgenome2008.02.0089

Zou, C. (2019). Genetic analysis of field related phenotypic data of a very diverse *Brassica oleracea* collection. Wageningen University Thesis.

6 APPENDICES

Appendix 1. Field layout for 2018

Block B 3 & 4	Block B 1 & 2	Block A 3 & 4	Block A 1 & 2
Ornamental (22)	Kholrabi (48)	Cabbage white (102)	Brussels sprout (48)
Cabbage Savoy (20)			
Cabbage red (43)	Cauliflower (60)		Collard green (20)
			Cabbage pointed (4)
Cabbage pointed (8)			Ornamental (22)
Cabbage white (109)		Cabbage pointed (4)	
		Cabbage white (8)	Tronchuda (24)
	Brussels sprout (48)	Kholrabi (46)	
			Cabbage Savoy (20)
		Cauliflower (60)	Cabbage red (43)
	Collard green (22)		
	Tronchuda (23)		Brussels sprout (1)
			Cabbage white (42)
Cauliflower (26)	Cabbage white (30)	Kohlrabi (2)	
		Cauliflower (30)	

Figure 17. Overview of field layout with the number of accession per morphotype

Appendix 2. Different morphotypes and related species of *B. oleracea* with their numbers that were used for genotyping

Table 15. Morphotype and number of accessions that were used for genotyping

Morphotype	Number of hybrids	Number of accessions	Total number
Heading cabbage (total)	130	184	314
White	78	103	181
Red	21	23	44
Savoy	11	39	50
Pointed	5	4	9
Unknown	15	15	30
Cauliflower	137	93	230
Kohlrabi	17	34	51
Brussels Sprouts	10	39	49
Ornamentals	27	1	28
Tronchuda	1	25	26
Collard Green	0	22	22
Broccoli	54	39	93
Wild C9 species (not oleracea)	0	58	58
Kale	5	30	35
Wild B. oleracea	0	18	18
Chinese Kale	1	7	8
Off types	1	3	4
Total	383	553	936

Appnedix 3. Guideline for GWAS in TASSEL (adapted from Mortel, 2018)

The Genome Wide Association Study (GWAS) of this research was done in the Program Tassel. The guideline for conducting GWAS in TASSEL.

For the GWAS three types of data are needed:

- **Genotypic data**
- **PCO**
- **Phenotypic data**

These data files all need to be inserted in a certain way.

Genotypic data: This is the 'Genotypic data - 80% GT - 2.5% MAF - 60% missing values - 18.850 SNP.vcf' file. Nothing has to be adjusted to this file, this file just needs to be imported.

PCO: has to be a text file with tabs delimited which starts in the following way:

<covariate>

<Trait>	Axis1	Axis2	etc.->
TKI008	0.050394	0.065680	
TKI010	0.020356	0.059508	

Etc.

The rest of the information which is in this file (% of variation explained, etc.) all has to be deleted.

Phenotypic data: This also has to be a text file. The file needs to start in the following way:

<Trait>	<Scars>	<Leafs>	etc.->
TKI008	5.167	17.83	
TKI010	5.83	19.67	

Etc.

For this file the complete data with all numbers behind the point needs to be added.

After preparing these files they can all be added to Tassel.
File -> open -> select the 3 different files (Genotypic data, PCO & Phenotypic data)

Then the three different files need to be combined to one file.
Select all three files together, press present -> data ->
With missing data: press intersect join
Without missing data: press union join
A new file will appear where all the three files are combined.

Then the analyses can be run. Select the new file in which all three data files (Genotypic data, PCO & Phenotypic data) were combined. Press Analysis -> Association -> GLM
Select 'Save file to' for the genotypic and for the statistics data and give it a clear name to find back later.
Select the box: 'Run permutations', type 999 at number of permutations.
Then press OK and the analyses will run.

When the analyses is finished the genotypic and statistics data file can be opened and the results can be analysed.

At the genotypic file the significant SNP's can be found per trait. This data can also be copied and pasted into an excel file to only save the significant SNP's.

At the statistics file figures can be made as output. For instance a QQ-plot and Manhattan Plots can be made.
Select the statistics file -> go to Results -> select QQ Plot ->Select trait
-> Select Manhattan Plot ->Select trait

Appendix 4. Rank of accession (descending order) on the basis of # leaf before heading

Table 16. Accession with their associated leaf numbers (ranked for both block)

Rank	Block A	#leaf		Block B	#leaf
1	TKI421	25		TKI777	22
2	**TKI780**	25		**TKI428**	23
3	**TKI050**	26		**TKI780**	24.5
4	**TKI428**	26		TKI295	25
5	**TKI028**	**27**		**TKI638**	25
6	TKI928	27		**TKI050**	26
7	TKI691	28		**TKI056**	26
8	**TKI638**	30		TKI382	28
9	**TKI056**	30.5		TKI935	28
10	TKI925	31.5		**TKI028**	29

Appendix 5. Summary statistics of different traits that were considered for this thesis

Table 17. Descriptive Statistics for all the traits that were considered for this thesis

Trait	Mean	Minimum	Maximum	Standard deviation	CV
Heading cabbage subset					
Scars	11.77	6	38	5.175	43.96
Leaves	29.97	13	98	9.676	32.29
Total leaves	41.17	0	136	12.49	30.34
Head leaves	66.54	35	114	13.78	20.70
Axillary shoot gr1	10.79	0	37	6.324	58.59
Axillary shoot gr2	2.192	0	10	2.324	106.0
Axillary shoot gr3	1.119	0	14	1.998	178.5
Total no. of axillary shoot	13.79	0	40	7.361	53.38
Secondary head gr1	1.221	0	11	1.991	163.1
Secondary head gr2	0.373	0	7	1.062	284.6
Secondary head gr3	0.0728	0	5	0.414	568.5
Total no. of secondary head	1.667	0	18	2.877	172.6
Brussels sprout subset (Control treatment)					
Plant height	61.39	28	115	13.93	22.50
Stalk height	56.01	25	94	13.68	24.42
Total yield	609.4	14.2	1612	332.5	54.56
Yield gr1	108.0	0	407.4	69.08	63.96
Yield gr2	271.6	0	932.5	197.7	72.78
Yield gr3	141.2	0	971.4	180	127.5
Total market yield	519.4	3.7	1564	294.3	56.65
Yield loss	89.99	5.5	513.6	84.39	93.78
%Yield loss	15.47	0.756	73.94	11.57	74.80
Brussels sprout subset (Decapitation treatment)					
Stalk height	57.02	28	92	11.82	20.73
Total yield	665.9	45.8	2006	357.0	53.62
Yield gr1	108.8	3.8	394.1	69.30	63.69
Yield gr2	306.3	0	1035	207.6	67.77
Yield gr3	154.6	0	1194	204.7	132.4
Total market yield	569.7	28.8	1918	326.4	57.29
Yield loss	96.19	6.5	556	94.56	98.31
%Yield loss	15.58	0.797	57.44	12.26	78.69

Appendix 6. Contribution of different yield group on total yield, total marketable yield and increase of yield after treatment

Table 18. Contribution to total yield (%)

Treatment	Yield gr1	Yield gr2	Yield gr3
Control	17.21691	45.70324	22.56451
Decapitation	15.85302	47.40185	22.6692

Table 19. Contribution to total marketable yield (%)

Treatment	Yield gr1	Yield gr2	Yield gr3
Control	20.14035	53.46368	26.39598
Decapitation	18.45004	55.16713	26.38283

Table 20. Yield increase after treatment (%)

Total yield	Total Marketable yield	Yield gr1	Yield gr2	Yield gr3
6.73	7.21	-1.29	10.07	7.16

Appendix 7. Population structure correction for different set of accessions for heading cabbages

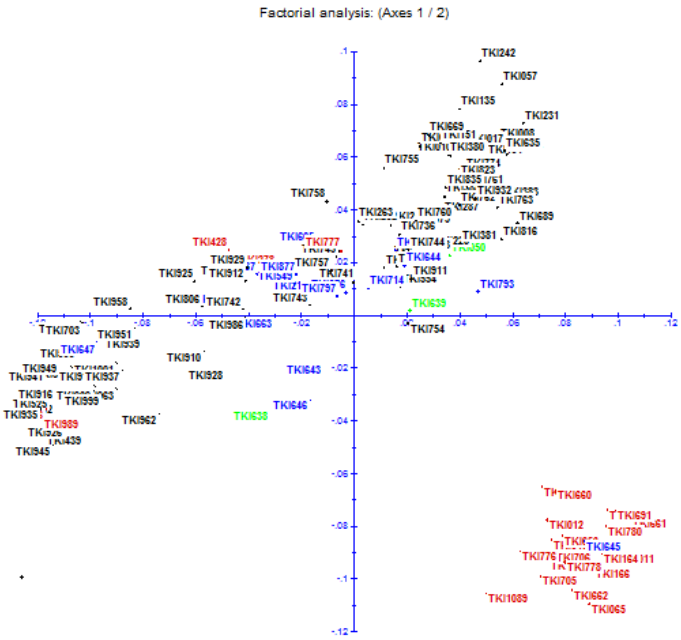

Figure 18. PCO of leaf with 133 accessions. First two axes (x=1, y=2) altogether explain 19.3% of the total variation (12.18% and 7.11%). Different colour labels were used for different sub-morphotypes. Red colours were used for red cabbages, which form a separate group than the other cabbages.

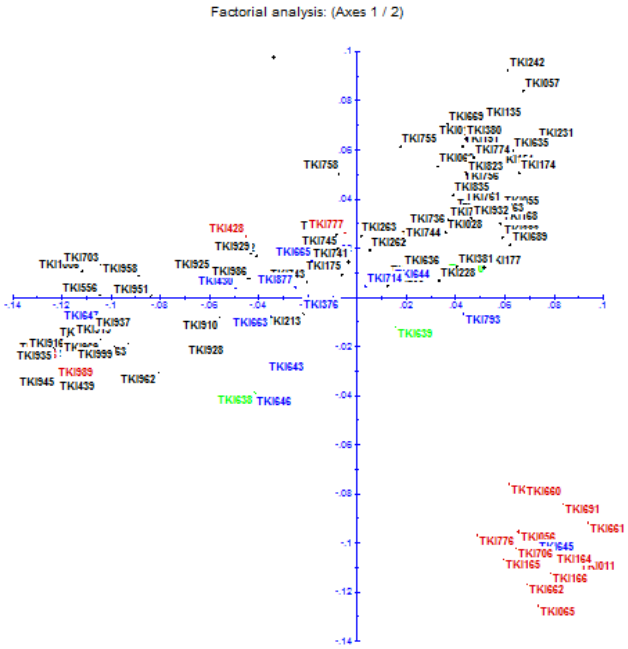

Figure 19. PCO of head leaf with 106 accessions. First two axes (x=1, y=2) altogether explain about 20% of the total variation (12.4% and 7.44%). Different colour labels were used for different sub-morphotypes. Red colours were used for red cabbages, which form a separate group than the other cabbages.

Appendix 8. PCO of 140 axes for axillary shoot related data of heading cabbage

Table 21. PCO of 106 heading cabbage accessions with 140 axes. The variation that can be explained by every axis is shown here

Axis	% Explained	Axis	% Explained	Axis	% Explained	Axis	% Explained
1	12.13	36	0.78	71	0.39	106	0.13
2	6.93	37	0.77	72	0.37	107	0.11
3	4.7	38	0.75	73	0.37	108	0.11
4	3.53	39	0.73	74	0.36	109	0.1
5	3.09	40	0.73	75	0.35	110	0.1
6	2.41	41	0.73	76	0.34	111	0.08
7	2.26	42	0.69	77	0.32	112	0.08
8	1.96	43	0.68	78	0.32	113	0.08
9	1.9	44	0.67	79	0.32	114	0.07
10	1.82	45	0.66	80	0.31	115	0.06
11	1.72	46	0.65	81	0.31	116	0.06
12	1.64	47	0.64	82	0.3	117	0.05
13	1.52	48	0.62	83	0.28	118	0.05
14	1.49	49	0.61	84	0.28	119	0.04
15	1.38	50	0.6	85	0.27	120	0.03
16	1.35	51	0.59	86	0.27	121	0.03
17	1.3	52	0.57	87	0.26	122	0.02
18	1.24	53	0.57	88	0.25	123	0.01
19	1.2	54	0.55	89	0.25	124	0
20	1.19	55	0.54	90	0.24	125	
21	1.13	56	0.53	91	0.23	126	
22	1.11	57	0.52	92	0.21	127	
23	1.07	58	0.52	93	0.21	128	
24	1.06	59	0.51	94	0.2	129	
25	1.01	60	0.5	95	0.19	130	
26	1.01	61	0.48	96	0.18	131	
27	0.97	62	0.47	97	0.18	132	
28	0.96	63	0.45	98	0.17	133	
29	0.94	64	0.45	99	0.17	134	
30	0.93	65	0.44	100	0.16	135	
31	0.91	66	0.43	101	0.16	136	
32	0.89	67	0.42	102	0.15	137	
33	0.85	68	0.42	103	0.14	138	
34	0.84	69	0.41	104	0.13	139	
35	0.82	70	0.4	105	0.13	140	

Appendix 9. PCO of 133 axes for leaf related data of heading cabbage

Table 22. PCO of 133 heading cabbage accessions with 133 axes. The variation that can be explained by every axis is shown here

Axis	% Explained	Axis	% Explained	Axis	% Explained	Axis	% Explained
1	12.18	35	0.82	69	0.4	103	0.13
2	7.11	36	0.8	70	0.39	104	0.11
3	4.76	37	0.78	71	0.37	105	0.1
4	3.6	38	0.77	72	0.37	106	0.09
5	3.11	39	0.75	73	0.36	107	0.09
6	2.45	40	0.73	74	0.34	108	0.08
7	2.2	41	0.72	75	0.33	109	0.08
8	2.03	42	0.7	76	0.33	110	0.07
9	1.93	43	0.69	77	0.32	111	0.07
10	1.87	44	0.67	78	0.32	112	0.05
11	1.73	45	0.65	79	0.31	113	0.05
12	1.69	46	0.64	80	0.3	114	0.04
13	1.57	47	0.63	81	0.29	115	0.04
14	1.53	48	0.62	82	0.28	116	0.03
15	1.41	49	0.61	83	0.28	117	0.02
16	1.38	50	0.6	84	0.27	118	0.01
17	1.31	51	0.58	85	0.26	119	0.01
18	1.25	52	0.57	86	0.25	120	0
19	1.24	53	0.57	87	0.24	121	
20	1.2	54	0.56	88	0.23	122	
21	1.16	55	0.55	89	0.23	123	
22	1.12	56	0.54	90	0.22	124	
23	1.11	57	0.53	91	0.21	125	
24	1.09	58	0.52	92	0.2	126	
25	1.04	59	0.51	93	0.19	127	
26	1.02	60	0.49	94	0.19	128	
27	1.01	61	0.48	95	0.18	129	
28	0.98	62	0.46	96	0.17	130	
29	0.95	63	0.46	97	0.16	131	
30	0.94	64	0.44	98	0.16	132	
31	0.92	65	0.44	99	0.15	133	
32	0.91	66	0.43	100	0.15		
33	0.89	67	0.41	101	0.14		
34	0.83	68	0.41	102	0.13		

Appendix 10. PCO of 106 axes for head leaves related data of heading cabbage

Table 23. PCO of 106 heading cabbage accessions with 106 axes. The variation that can be explained by every axis is shown here

Axis	% Explained	Axis	% Explained	Axis	% Explained	Axis	% Explained
1	12.4	31	0.98	61	0.44	91	0.11
2	7.44	32	0.93	62	0.44	92	0.07
3	4.92	33	0.93	63	0.43	93	0.07
4	3.85	34	0.89	64	0.4	94	0.07
5	3.35	35	0.88	65	0.4	95	0.06
6	2.8	36	0.87	66	0.38	96	0.04
7	2.51	37	0.84	67	0.36	97	0.03
8	2.19	38	0.81	68	0.35	98	0.02
9	2.06	39	0.79	69	0.34	99	0.01
10	1.96	40	0.77	70	0.33	100	
11	1.95	41	0.76	71	0.33	101	
12	1.88	42	0.75	72	0.31	102	
13	1.73	43	0.72	73	0.3	103	
14	1.58	44	0.7	74	0.3	104	
15	1.56	45	0.68	75	0.28	105	
16	1.47	46	0.68	76	0.27	106	
17	1.45	47	0.67	77	0.25		
18	1.38	48	0.65	78	0.24		
19	1.32	49	0.63	79	0.22		
20	1.31	50	0.61	80	0.22		
21	1.29	51	0.6	81	0.21		
22	1.23	52	0.57	82	0.2		
23	1.2	53	0.56	83	0.18		
24	1.18	54	0.53	84	0.17		
25	1.12	55	0.53	85	0.17		
26	1.11	56	0.5	86	0.15		
27	1.09	57	0.5	87	0.15		
28	1.06	58	0.49	88	0.13		
29	1.04	59	0.47	89	0.12		
30	1.01	60	0.46	90	0.12		

Appendix 11. PCO of 48 axes for 48 harvested Brussels sprout accessions

Table 24. PCO of 48 harvested Brussels sprout accessions with 48 axes. The variation that can be explained by every axis is shown here

Axis	% Explained	Axis	% Explained
1	41.31	25	0.74
2	9.75	26	0.7
3	4.57	27	0.66
4	3.78	28	0.64
5	3.24	29	0.58
6	2.96	30	0.52
7	2.66	31	0.49
8	2.31	32	0.46
9	2.23	33	0.44
10	2.01	34	0.39
11	1.88	35	0.36
12	1.77	36	0.32
13	1.53	37	0.3
14	1.49	38	0.26
15	1.43	39	0.24
16	1.32	40	0.19
17	1.15	41	0.17
18	1.08	42	0.14
19	1.03	43	0.11
20	1	44	0.08
21	0.89	45	0
22	0.85	46	
23	0.81	47	
24	0.79	48	

Appendix 12. QQ plot of expected vs calculated –Log10(P-value) for total #leaf,

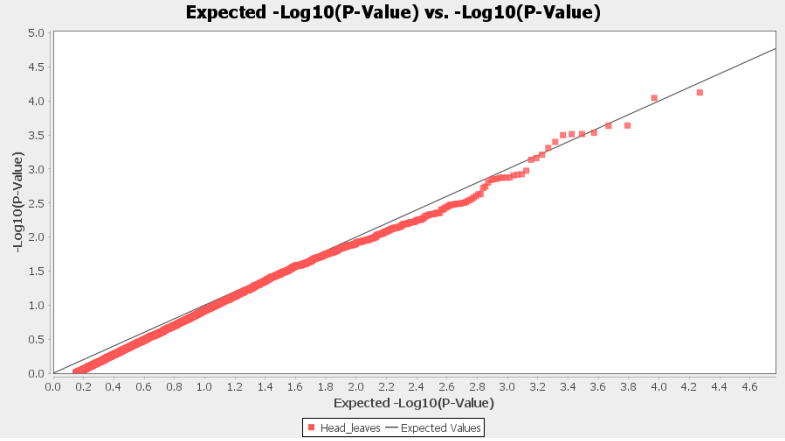

Figure 20. QQ plot of expected vs calculated –Log10(P-value) for leaf and total # head leaves with a PCO of 106 axes. The found LOD score runs parallel to the line of the expected LOD scores.

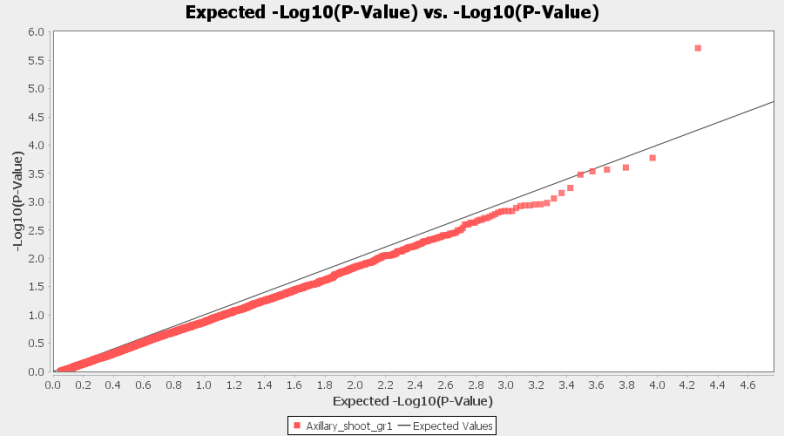

Figure 21. QQ plot of expected vs calculated –Log10(P-value) for axillary shoot gr1 with a PCO of 140 axes. The found LOD score runs parallel to the line of the expected LOD scores.

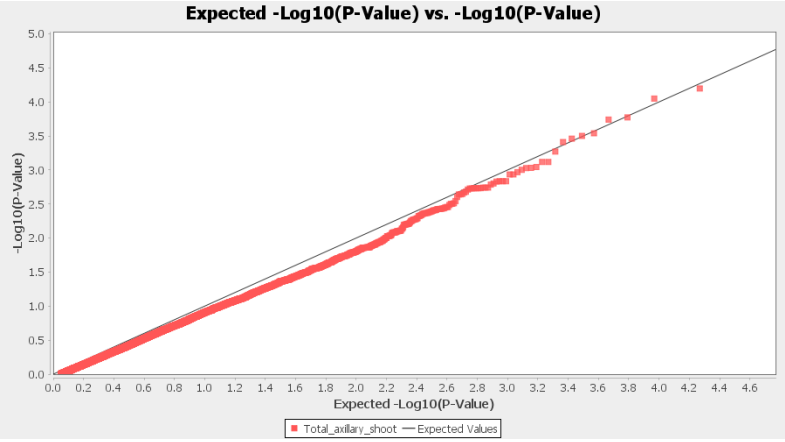

Figure 22. QQ plot of expected vs calculated –Log10(P-value) for leaf and total # axillary shoot with a PCO of 140 axes. The found LOD score runs parallel to the line of the expected LOD scores.

Appendix 13. Manhattan plot of GWAS for different accessions of heading cabbages with different axes and PCOs (treatment: control)

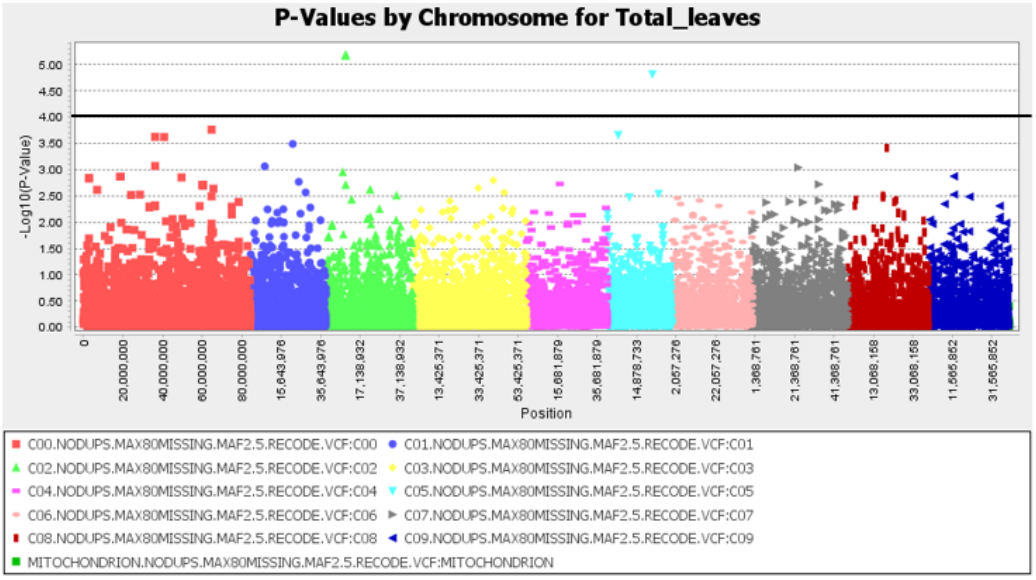

Figure 23. Manhattan plot for total # leaf with a population structure correction with 133 accession and 133 axes. The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNP above threshold level (LOD= 4.0) can be found only in chromosome 2.

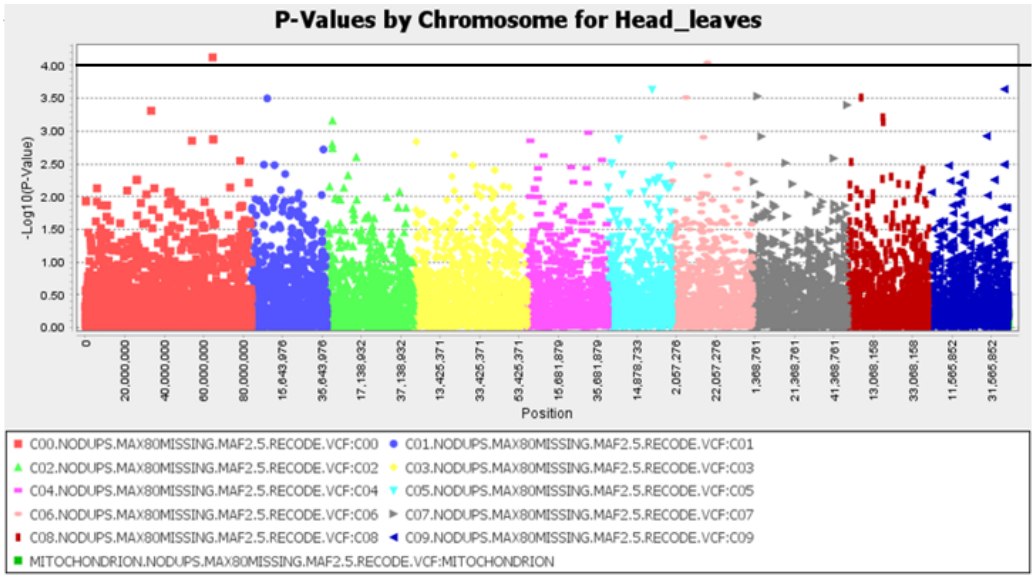

Figure 24. Manhattan plot for # head leaf with a population structure correction with 106 accession and 106 axes. The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNP above threshold level (LOD= 4.0) can be found only in chromosome 2.

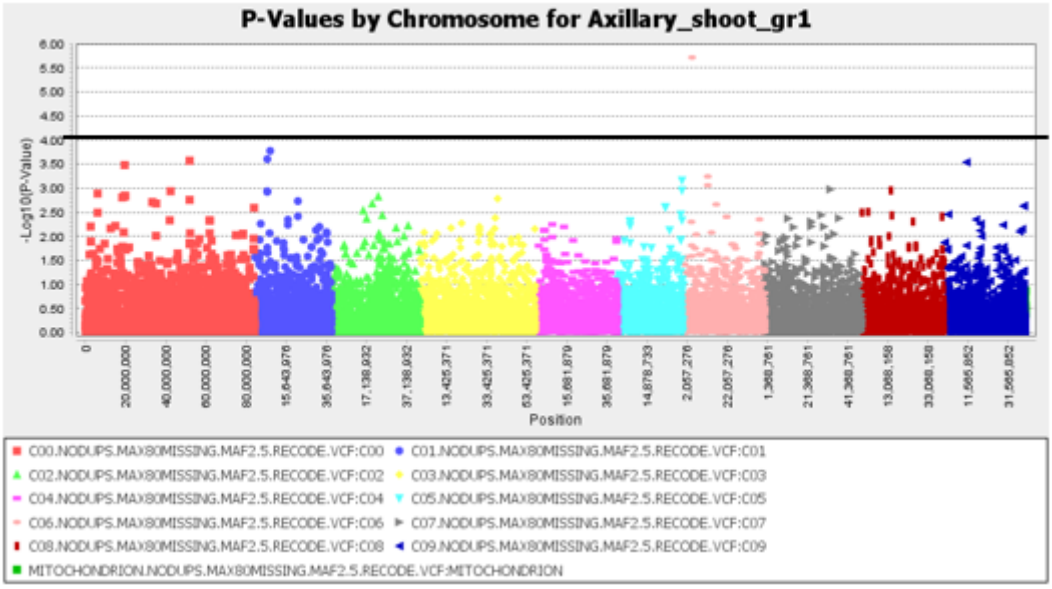

Figure 25. Manhattan plot for axillary shoot gr1 with a population structure correction with 140 accession and 140 axes. The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNP above threshold level (LOD= 4.0) can be found only in chromosome 2.

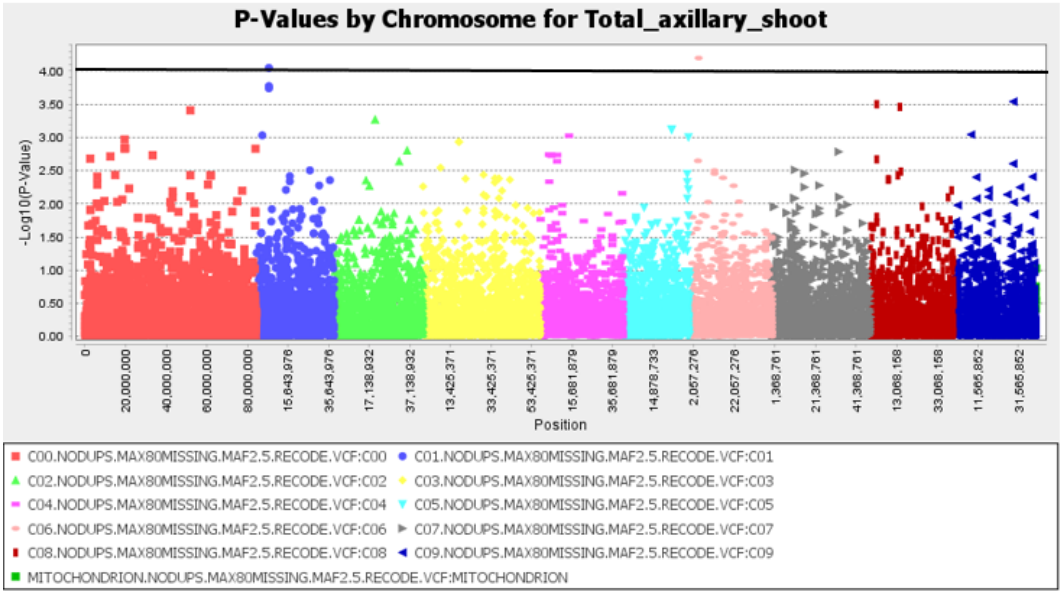

Figure 26. Manhattan plot for total# axillary shoot with a population structure correction with 140 accession and 140 axes. The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNP above threshold level (LOD= 4.0) can be found only in chromosome 2.

Appendix 14. Manhattan plot of GWAS for 48 accessions of Brussels sprout with a PCO of 2 axes (treatment: control)

Figure 27. Manhattan plot for different traits (**Treatment: Control**). The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNPs above threshold level (LOD= 4.0) can be found in chromosome 3, 4 and 9.

Appendix 15. Manhattan plot of GWAS for 48 accessions of Brussels sprout with a PCO of 2 axes (treatment: decapitation)

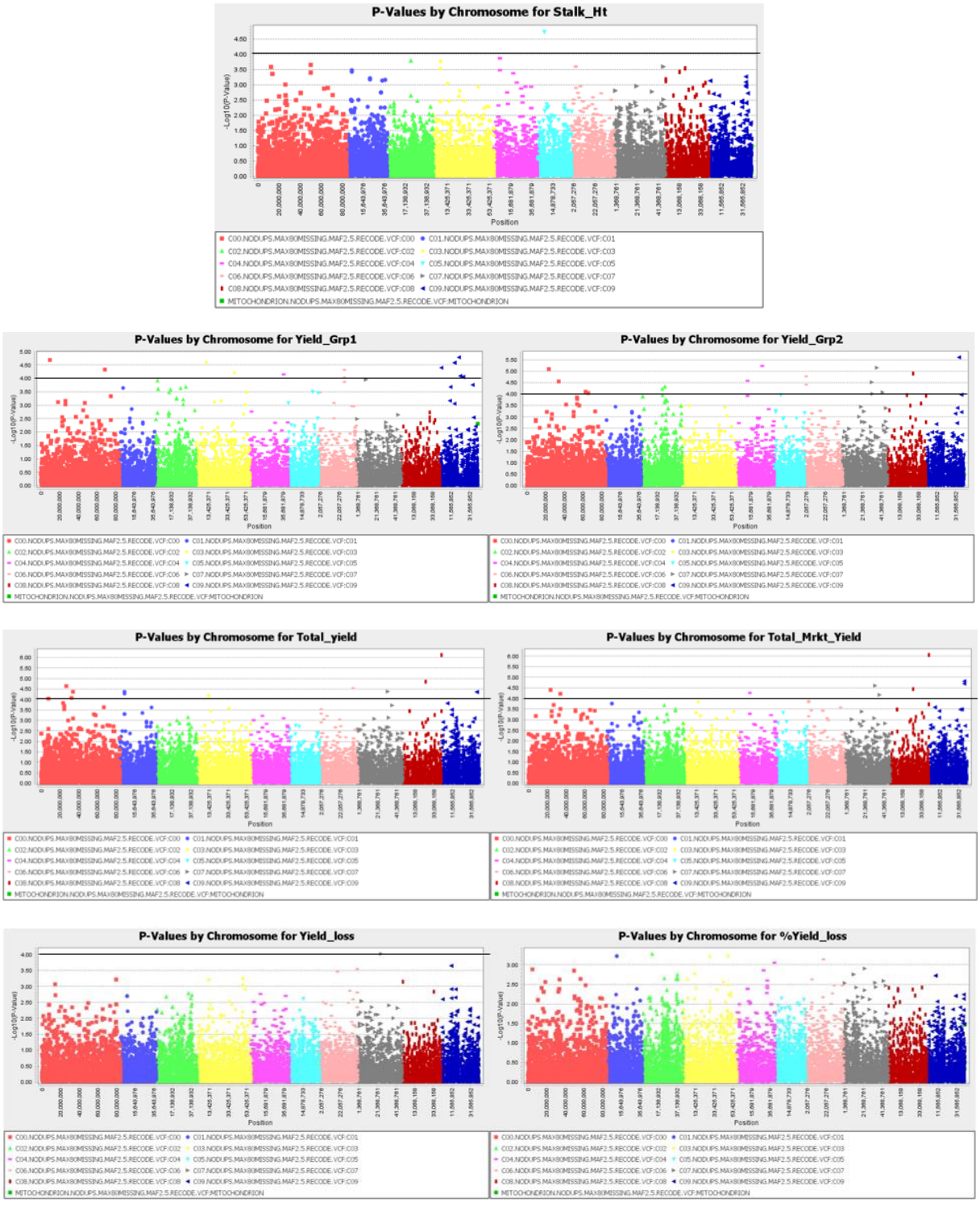

Figure 28. Manhattan plot for different traits (**Treatment: Decapitation**). The first block is for chromosome 0, representing scaffold with an unknown chromosomal location. The following blocks from left to right with different colours represent different chromosome from 1 to 9. Significant SNPs above threshold level (LOD= 4.0) can be found in chromosome 3, 4 and 9.

Appendix 16. Analysis of variance (ANOVA) for heading cabbages for all the traits that were considered for this thesis

Analysis of variance Scars

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	193.965	193.965	22.5	< 0.001
TKI	135	8059.155	59.697	6.93	< 0.001
Residual	254	2189.622	8.621		
Total	390	10442.74	26.776		

Analysis of variance Leaf

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	386.6	386.6	6.41	0.012
TKI	132	18857.65	142.86	2.37	< 0.001
Residual	204	12306.39	60.33		
Total	337	31550.64	93.62		

Analysis of variance Total # leaf

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	2.07	2.07	0.02	0.88
TKI	132	34296.48	259.82	2.84	< 0.001
Residual	210	19207	91.46		
Total	343	53505.56	155.99		

Analysis of variance Axillary Shoot Gr1

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	858.44	858.44	35.1	< 0.001
TKI	139	9167.24	65.95	2.7	< 0.001
Residual	285	6970.15	24.46		
Total	425	16995.82	39.99		

Analysis of variance Axillary Shoot Gr2

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	0.268	0.268	0.07	0.788
TKI	139	1174.876	8.452	2.3	< 0.001
Residual	244	898.633	3.683		
Total	384	2073.777	5.4		

Analysis of variance Axillary Shoot Gr3

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	0.638	0.638	0.19	0.665
TKI	139	705.715	5.077	1.5	0.003
Residual	245	830.165	3.388		
Total	385	1536.518	3.991		

Analysis of variance Total # Axillary Shoot

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	466.92	466.92	13.96	< 0.001
TKI	139	13028.94	93.73	2.8	< 0.001
Residual	285	9533.12	33.45		
Total	425	23028.99	54.19		

Analysis of variance Secondary Heads Gr1

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	0.084	0.084	0.03	0.867
TKI	139	831.779	5.984	2	< 0.001
Residual	285	853.396	2.994		
Total	425	1685.258	3.965		

Analysis of variance Secondary Heads Gr2

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	0.3657	0.3657	0.42	0.519
TKI	139	228.976	1.6473	1.88	< 0.001
Residual	285	250.3133	0.8783		
Total	425	479.6549	1.1286		

Analysis of variance Secondary Heads Gr3

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	0.1373	0.1373	0.89	0.345
TKI	139	28.7918	0.2071	1.35	0.019
Residual	285	43.815	0.1537		
Total	425	72.7441	0.1712		

Analysis of variance Total # Secondary Heads

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	1.598	1.598	0.27	0.602
TKI	139	1845.045	13.274	2.26	< 0.001
Residual	285	1672.023	5.867		
Total	425	3518.667	8.279		

Analysis of variance Weight_grams

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	12228477	12228477	47.63	< 0.001
TKI	126	2.46E+08	1950455	7.6	< 0.001
Residual	236	60584568	256714		
Total	363	3.19E+08	877604		

Analysis of variance Head_Leaves

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Block	1	372.13	372.13	5.4	0.022
TKI	105	36184.27	344.61	5.01	< 0.001
Residual	136	9363.89	68.85		
Total	242	45920.3	189.75		

Appendix 17. Correlation between all traits measured for Brussels sprout in control and decapitation treatment

Table 25. Correlation between all traits measured for the Brussels sprout in control treatment

Plant_height	1	-								
Stalk_height	2	0.98	-							
Yield gr1	3	0.20	0.22	-						
Yield gr2	4	0.37	0.38	-0.02	-					
Yield gr3	5	0.30	0.26	-0.36	0.29	-				
Total_Marketable_yield_g	6	0.48	0.46	-0.01	0.86	0.70	-			
Total_yield_per_plant	7	0.47	0.46	-0.05	0.81	0.74	0.97	-		
Yield_loss	8	0.17	0.17	-0.19	0.16	0.44	0.33	0.53	-	
%_yield_loss	9	-0.13	-0.11	-0.26	-0.36	-0.03	-0.32	-0.15	0.58	-
		1	2	3	4	5	6	7	8	9

Table 26. Correlation between all traits measured for the Brussels sprout in decapitation treatment

Stalk_height	1	-							
Yield gr1	2	0.10	-						
Yield gr2	3	0.38	-0.10	-					
Yield gr3	4	0.20	-0.36	0.30	-				
Total_Marketable_yield_g	5	0.40	-0.08	0.83	0.75	-			
Total_yield_per_plant	6	0.37	-0.10	0.77	0.76	0.96	-		
Yield_loss	7	0.04	-0.11	0.03	0.29	0.17	0.42	-	
%_yield_loss	8	-0.22	-0.17	-0.46	-0.11	-0.41	-0.19	0.70	-
		1	2	3	4	5	6	7	8