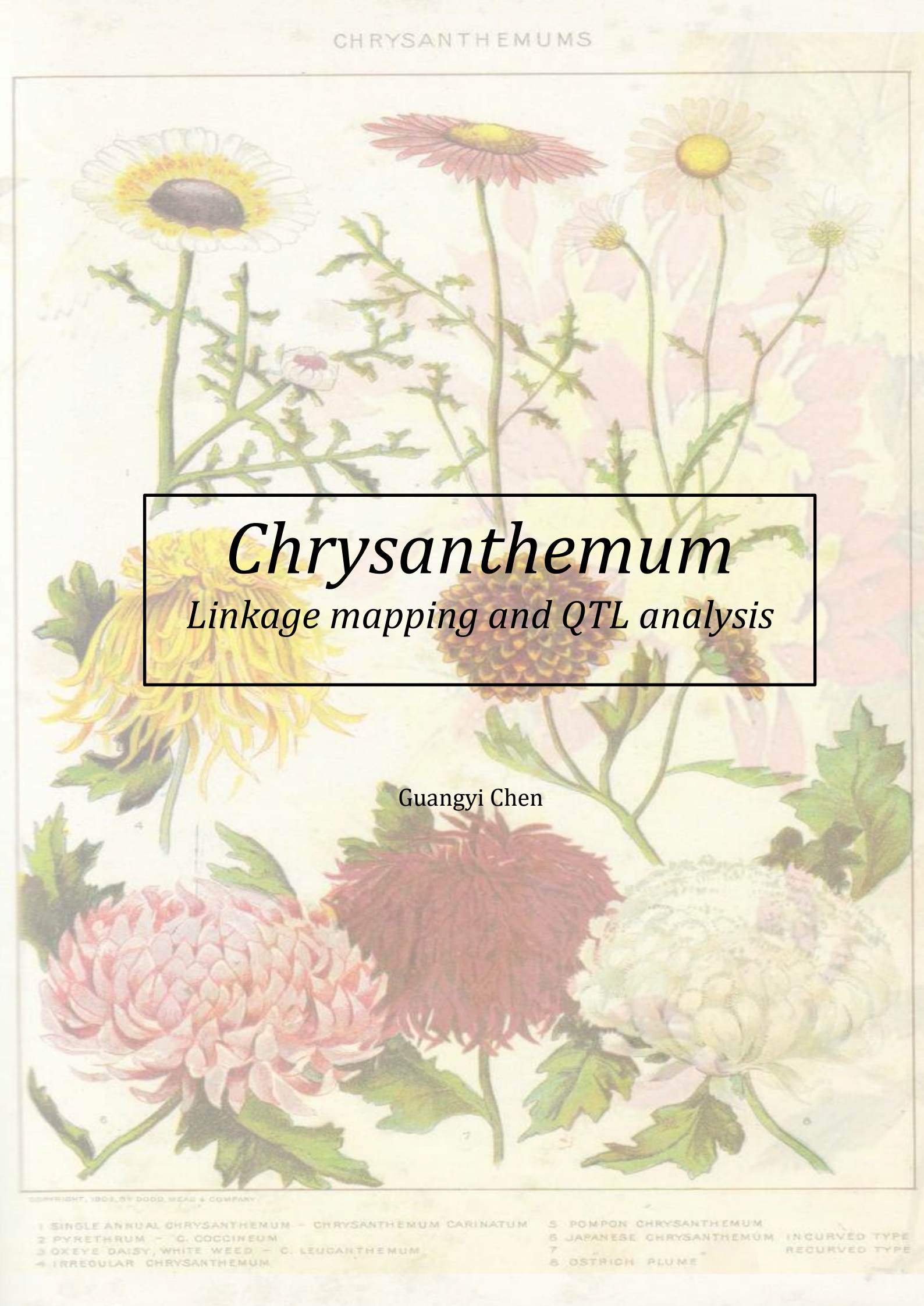

Chrysanthemum

Linkage mapping and QTL analysis

Guangyi Chen

Construction of an integrated linkage map of hexaploid chrysanthemum x *morifolium* and QTL analysis on leaf traits

Guangyi Chen

930802157090

M.Sc Major Thesis PBR-80436

Supervisor: Paul Arens, Chris Maliepaard, Geert van Geest,

May 2018

Laboratory of Plant Breeding, Wageningen University & Research

Wageningen, the Netherland

Source of cover picture: Dodd, Mead & Co. 1902

Abstract

This paper illustrates the construction of an integrated linkage map with 9 linkage groups of hexaploid chrysanthemum x morifolium by using a 183k Affymetrix Axiom SNP array. After that, QTL analysis were performed for four leaf traits: leaf width/length ratio, stem length, leaf stipule size and leaf number. The results indicate no significant QTL regions have been detected for traits leaf width/length ratio, leaf stipule size and leaf number. Two significant QTL regions have been found on linkage group 4 and 7 for trait stem length. A marker has been confirmed to be significantly linked to this trait. Also, a list of candidate genes was generated by performing BLAST with the sequences of contigs on the significant QTL regions of this trait. Under the circumstance that the genome of chrysanthemum has not yet been fully sequenced, these candidate genes could provide reference information for estimating genes in chrysanthemum.

Key words: chrysanthemum, SNP genotyping, linkage mapping, QTL, leaf traits

Acknowledgement

I would like to acknowledge the tremendous help from my supervisors: Dr. Geert van Geest, Dr. Chris Maliepaard and Dr. Paul Arens. I wouldn't finish this amazing project without your support.

I would like to thank Geert for teaching me from zero and showing me the beauty of polyploid world. I would like to thank you for giving me huge help on programming skills. Your modesty, your attitude and enthusiasm to scientific researching which make me learned a lot.

I would like to thank Chris for giving me valuable feedbacks and support. It is my honor to be your student in this project and in the statistic courses before.

I would like to thank Paul for all the support and encouragement. You're a spirit totem to me, sterkte!

I also would like to thank Yanlin and other students, friends who gave me huge help during my project.

Content

1. Introduction

1.1 General introduction of chrysanthemum

Chrysanthemum (*Chrysanthemum morifolium*), which belongs to family *Asteraceae*, is one of the most important ornamental flowers in the world. It is broadly used for cut-flower and potted plant. The diversity of its flower traits creates much enjoyment for people. Dating back to around 1600 years ago, the famous Chinese poet Tao Yuanming wrote the famous sentence '秋菊有佳色，裛露掇其英' (The chrysanthemum looks splendid in the autumn, I pick it up with some dew with it) to describe the beauty of chrysanthemum.

Chrysanthemum originally comes from China and has a long history of cultivation in Asia. In the 18th century it has been transported to Europe from China by a French merchant. Now it has become one of the most important ornamental flowers in the world markets.

The aim of chrysanthemum breeding is to combine good horticultural traits. The traditional way is to cross chrysanthemums with different traits and select the offspring with target traits. However, because chrysanthemum genome is very complex the traditional breeding can be very much hampered and slowed down. With the help of booming development of DNA markers, DNA-informed breeding can largely speed up the breeding process. The most important issue of DNA-informed breeding is to genotype large number of genotypes to get the polymorphisms information and associate them with certain phenotypes.

1.2 Genetic study of chrysanthemum

1.2.1 Polyploids and dosage

Polyploid plants are those plants containing more than two paired set of chromosomes. As a hexaploid plant, chrysanthemum has six homologues of each chromosome. This can be very important in generating phenotypic diversity and creating new cultivars because of huge segregations in the next generations. For each locus, offspring will receive 3 alleles from each of the parent by inheriting homologous chromosomes. It has been proved that chrysanthemum shows polysomic inheritance mode (see 1.2.2), consequently there are $c\binom{3}{6} * c\binom{3}{6} = 400$ different segregation scenarios for each locus during the meiosis. However, huge segregations can largely complicate the study of inheritance. Hence, it is of much difficulty to predict what will be outcome of crosses and to know if certain traits are present in young seedlings.

Unlike diploid plants, due to the multiple numbers of homologues in the chromosome set of polyploid plants, there may exist multiple alleles in one locus.

Although in some outcrossing diploids, there may also exist multiple kinds of alleles in one locus, e.g. in gerbera not uncommon to find 3-4 different alleles in a cross (Wernett *et al*, 1996), but for sure the situation goes more complex in hexaploids. There are four homologues chromosomes of each chromosome set in tetraploid plants and six in hexaploid plants. Here we call the relative allele composition in each locus dosage.

Inside the chrysanthemum genomes there exist large numbers of polymorphisms. Single-nucleotide polymorphism (SNP) is one of the most common polymorphisms. SNP is a variation in a single nucleotide occurring at a specific position in the genome. SNP loci are generally bi-allelic, which means there are 2 potential alleles in one SNP locus. As in the case of chrysanthemum, 7 possible SNP genotypes can be generated in one locus: AAAAAA, AAAAAB, AAAABB, AAABBB, AABBBB, ABBBBB, BBBBBB, representing dosage from 0 to 6 considering the B allele.

1.2.2 Inheritance mode of chrysanthemum

The inheritance mode of *chrysanthemum* has been proven to be hexasomic (van Geest, Voorrips, et al., 2017). However, it has been discussed for a long time that whether the inheritance mode of chrysanthemum should be hexasomic or disomic, or a mix with both. The differences of polysomic and disomic inheritance mode can be explained by Fig. 1 (van Geest, Voorrips, et al., 2017). In disomic inheritance, 6 alleles in each locus are divided into 3 groups, each group has 2 alleles. During meiosis one allele of each of the 3 groups is chosen to pass to the offspring. Hence in total there are $2^3 = 8$ kinds of possible gametes for every locus. In hexasomic inheritance, 3 alleles out of 6 alleles are randomly paired and delivered to the offspring. In total $c\binom{3}{6} = 20$ kinds of possible gametes can be generated in each locus.

Previous research has shown that polysomic inheritance mode occupies a leading role in *chrysanthemum*. But the disomic mode may also exist (Klie, Schie, Linde, & Debener, 2014). There are two methods that have been used to analyse the inheritance mode in *chrysanthemum*: cytological imaging and segregation analysis. Cytological imaging shows that polysomic inheritance is indicated in *chrysanthemum* according to the presence of multivalents, but disomic inheritance is not excluded. Segregation analysis also shows conclusive evidence of polysomic inheritance of flower color (Langton, 1989), segregation of multi-allelic SSR-markers (Klie et al., 2014; Park, Arens, Esselink, Lim, & Shin, 2015) in chrysanthemum. However, these studies are limited because they are restricted to a relative low number of specific loci. Further, disomic inheritance is indicated based on segregation of multi-dose dominant alleles (De Backer, 2012; De Jong & Rademaker, 1986; Klie et al., 2014)

In order to study the mode of inheritance of chrysanthemum conclusively, a genome-wide approach is needed. In 1.2.1 the dosage in polyploids has been discussed. It is of great use to do the dosage counting for polyploid plants. For analyses of inheritance we first need to establish the dosage type of individual

parental plants. It can help to make a clear sight for analysing the inheritance mode of polyploid plants. van Geest, Voorrips, et al (2017) desigened an Affymetrix Axiom array with 183K SNPs. With the help of this SNP array four bi-parental populations have been genotyped and the dosages have been counted. The duplex x nulliplex marker segregations in the population supported the evidence for hexasomic inheritance in chrysanthemum of genome-wide range.

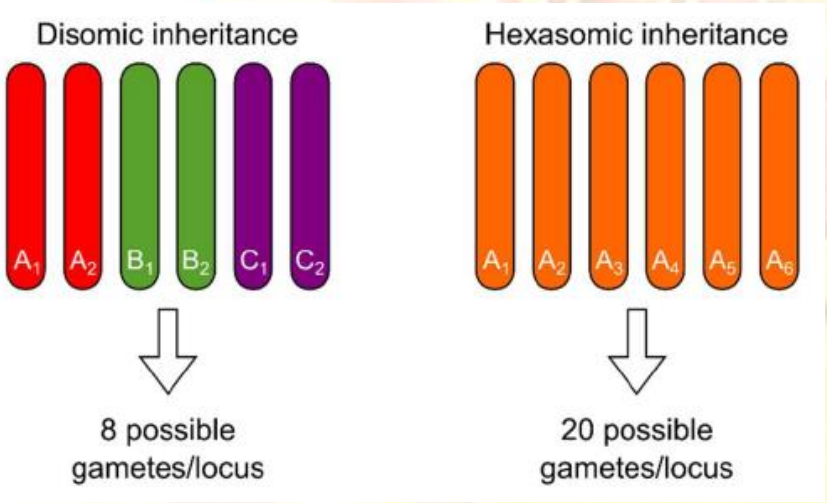

Figure 1. Graphical representation of disomic inheritance and hexasomic inheritance. Chromosomes represented by the same letter and colour can pair during meiosis I.

1.3 SNP genotyping and SNP array

It is important for research to discover markers linked to a trait. To achieve this a good even coverage of markers spreading all homologues is needed. Hence, it is important to genotype plant with different genetic constitutions (genotypes) to discover polymorphisms distributed over the entire genome. There are multiple kinds of polymorphisms, like Single-nucleotide polymorphism (SNP), Restriction Fragment Length Polymorphism (RFLP), microsatellites and so on. What makes SNPs the good marker of choice is their abundance in the genome but above all the possibility to use these markers in high throughput genotyping.

The genetic polymorphisms can be linked to certain locations of genes which are responsible for the trait of interest. Ideally, genotyping a set of polymorphisms that cover all homologues in the genome enables the finding of polymorphisms that are linked to a trait. The cost of detecting and genotyping large number of polymorphisms is dropping down. Hence it can be more and more well applied to the breeding of important plant species, including polyploids. The detection and analysis of polymorphisms in polyploids is more complex than in diploids. As an outcrossing hexaploid, the chrysanthemum genome has not been fully sequenced. The polymorphisms detection method of chrysanthemum is restricted to methods

using a reduced representation of the genome, like restriction enzyme based selection methods (RADseq, GBS, etc), bait capture and RNA sequencing.

SNP genotyping is the assessment of the allelic composition of individual genotypes. It can be very useful to detect the SNP polymorphisms at certain loci. Further these polymorphisms can be used to analyse the relationship with target traits. There are several high-throughput genomic technologies that have been put into use for SNP genotyping, like Genotyping By Sequencing (GBS), bait capture, Whole Genome Sequencing, SNP arrays, etc.

One of the most important technologies for detecting SNP polymorphisms are SNP genotyping arrays, which is a DNA microarray used to detect SNP polymorphisms within a population. The SNP arrays are very effective. There are hundreds of thousands of probes arrayed on the very small chip of SNP array, which can help to interrogate many SNPs at the same time. The SNP arrays are very efficient in detecting genome SNP polymorphisms among different members in one species. It has been shown that commercial probe-based SNP array platforms can now genotype about one million SNPs in an individual in one assay with a very high accuracy.

There are already many successful examples of SNP array applied in polyploids. A 68K rose SNP array by WagRhSNP Axiom has been developed for rose (Smulders et al., 2015). An Illumina Infinium array containing 9,277 SNPs has been developed for alfalfa (Li et al., 2014). An Axiom genotyping array with 183k SNP markers is designed and applied to chrysanthemum (van Geest, Voorrips, et al., 2017). A 20K SolSTW Infinium SNP array has been developed for tetraploid potato (*Solanum tuberosum*)(Vos, Uitdewilligen, Voorrips, Visser, & van Eck, 2015).

However, SNP arrays also have limitations. Because the SNPs are bi-allelic and the set of SNP loci is fixed, therefore other alleles that are not tagged in single SNP analyses will be ignored. Chrysanthemum crosses may have up to 12 possible different alleles/homeologs contributing to diversity in offspring whereas a SNP is mostly bi-alleleic. Also, as in the case of hexaploid chrysanthemum, discovering successful SNP markers and designing SNP arrays need laborious work. The large range of dosage sores in chrysanthemum complicates the dosage number prediction. Besides, there is no reference genome sequence available in chrysanthemum, so the SNP detection is based on de novo sequence assembly. This assembly for hexaploidy usually contains errors. Also, like tetraploids such as potato (Uitdewilligen et al., 2013), there is high SNP polymorphism densities in chrysanthemum. The neighbouring SNPs can have interference with each other.

1.4 Genetic mapping

Genetic mapping is also called linkage mapping. It is a method of constructing a map which shows the relative positions of genetic loci on chromosomes. The construction of genetic map is based on estimating the recombination frequency between alleles of loci, which are usually represented by genetic markers. The alleles of the

polymorphic loci that are close together on the same chromosome have a higher potential occurring together in the genome of offspring than random segregation. Recombinant frequency is the ratio that two alleles do not show up together in the offspring. The linkage between two polysomic loci can be estimated by calculating the recombination frequency.

The steps of constructing a linkage map are similar in diploid and polyploid crops. First a mapping population needs to be developed. The goal is to acquire enough genetic variation in terms of markers. A mapping population consists of large variation of phenotypes will increase the possibility of achieving large genetic variations in markers. There are different ways to perform this step. As in the case of chrysanthemum that is a strictly outcrossing hexaploid, usually we cross two parents that differ for the traits of interest distinctively. The next step is genotyping, which is to identify polymorphisms of markers in the mapping population. In the case of polyploids, there are multiple homologues for each chromosome. Next, dosage calling will be performed in order to find out the marker segregations in the offsprings. After that we need to do the linkage analysis of the markers to determine the degrees of associated segregation between markers. The construction of a linkage map will be done after clustering and ordering the markers.

There is already some progress in the construction of linkage maps in chrysanthemum. Zhang et al., (2010) developed a preliminary genetic linkage map of chrysanthemum cultivars by using RAPD, ISSR and AFLP markers but the map is limited to methods developed to diploids. Two separate genetic maps were constructed for each parent cultivar using SRAP markers (Zhang, Chen, Chen, Fang, Chen, et al., 2011). Van Geest, Bourke, et al., (2017) for the first time constructed an integrated genetic linkage map in a polysomic hexaploid chrysanthemum by using R package polymapR, which is of great usage to estimate inheritance of parental haplotypes in the offspring and detect multi-allelic QTLs afterwards.

1.5 QTL analysis

Quantitative trait locus (QTL) analysis is a statistical method that links phenotypic data (certain quantitative trait phenotype measurements) and genotypic data (usually molecular markers) in a mapping population. Many quantitative traits are regulated by more than one gene, these genes may act additively or interact with each other to a certain degree. The goal of QTL analysis is to investigate which chromosome regions on the genomes are associated with a certain quantitative trait and subsequently explain the genetic basis of variation in complex traits.
QTL analysis has been widely applied into diploids. Some polyploids like potato, rose and chrysanthemum are also the common targets for QTL analysis. Until very recently, the QTL analysis in hexaploids is quite limited, which still applies the simple segregation markers from diploid models but the actual situation is much more complex (Chang et al., 2009). This approach with only simplex segregating markers can only analyze one allele at a certain locus at a time, however other alleles at the same locus could also have effects on the phenotype. Consequently, multi-allelic QTL

analysis is needed for the estimation of presence or absence of each possible allele at one locus. Hackett, McLean, & Bryan, (2013) designed a method to infer the inheritance of each parental allele in each F1 individual for tetraploids. A simpler method is proposed to directly estimate the haplotype probabilities based on recombinant frequency (P. Bourke, 2014). This method can be applied to hexaploids.

For ornamental flowers, leaf architecture can have important influences on the beauty and can be very important to cultivation. In this thesis the target will be put on leaf traits: leaf number, stipule size, width/length ratio and stem length. These traits can be assumed quantitative traits because they have large variation among different individuals. Compared to flower traits, much less studies were focused on chrysanthemum leaf traits. Zhang et al., (2012) applied QTL analysis on three leaf traits: leaf length, leaf width and leaf length/width ratio. Another study focused on QTL analysis of trait plant height. These studies were based on genetic maps using sequence-related amplified polymorphism (SRAP) markers (Zhang, Jiang, Chen, Chen, & Fang, 2012). This map is not of high density, with the mean inter-marker distance of 6.9 cM. New QTL study is urgently needed to be operated based on a more marker-dense linkage map.

1.6 Research overview

Thanks to the development of DNA-informed technologies like SNP array, polyploids breeding can be proceeded much more precise than before. In this research, an integrated linkage map of chrysanthemum was constructed by using the genotyping data from a 183k SNP array. This map integrates the markers in all homologues of the both parents and provides basics for further genetic study like QTL analysis on chrysanthemum traits.

Leaf traits are very important traits in chrysanthemum, which have large influences on the shape of chrysanthemum. A few studies have been focused on chrysanthemum leaf traits but it is needed to be explored more in order to make improvements in breeding with the benefits of DNA-informed technology. Following the construction of an integrated linkage map, this thesis reached on performing QTL analysis for leaf traits in order to provide marker information for breeding chrysanthemum leaf traits. After that, candidate genes were provided by aligning contig sequences of the significant QTL regions with UniProt database, which provides reference information to estimate genes related to leaf trait stem length in chrysanthemum.

This research could give an overview of genetic analysis in chrysanthemum and provide methodologies of linkage mapping in chrysanthemum. In the long run, it also can give hints of genetic analysis in other polyploid species.

2. Materials and methods

2.1 Plant materials

The plant material used in this case is from Deliflor Chrysanten BV, which is a bi-parental population consists of 409 individuals representing different genotypes. Two individuals were served as parents and they were distinctively different from each other in the leaf traits of interest. The other 407 individuals were offspring after crossing. Considering the large number of individuals, this population had high possibilities of crossovers and was used for constructing integrated linkage map. Same individual grew in the same plot with multiple propagated replicates. For QTL analysis, 100 individuals and 3 replicates for each individual were randomly selected within the population for leaf traits phenotyping.

2.2 183k Affymetrix Axiom SNP array development and SNP genotyping

The method of developing a 183k Affymetrix Axiom SNP array used in this case has been explained in detail by Van Geest et al., (2017). The plant sources were chosen to be as diverse as possible to get as much genetic variations as possible. In total four bi-parental populations with were used as the materials for RNA sequencing. Also, three bi-parental populations were used for genotyping later. After the step of RNA extraction and RNA sequencing, the transcriptome data were achieved. These transcriptomes were assembled and the SNP variants were detected. After running the SNP filtering process, finally the array arrives in 183130 SNPs in total from the four bi-parental populations, among which 34068 could be tiled from both directions.

The mapping population was genotyped by an Affymetrix Axiom array with 183K SNPs as described by Van Geest et al., (2017). In order to better distinguish 7 genotype clusters maximumly in hexaploid, 4 replices were used in each probe of the Axiom SNP array instead of 2. The raw signals of the 4 replices of the probe were turned into normalized signal intensities per probe. The signal intensities are further prepared for dosage calling.
The genotyping data used in this paper was the same as the data from Van Geest, Voorrips, et al., (2017).

2.3 Dosage calling and marker quality filtering

Dosage calling is the step that assigns the marker to dosages from the genotyping array data. In this case, dosage calling was processed using R package fitPoly which is developed based on R package fitTetra (Voorrips, Gort, & Vosman, 2011).

After applying fitPoly, not all markers were qualified enough for further analysis. Markers and individuals were defined less qualified and were screened out by the following standards: non-segregating markers, markers with 5% missing values, individuals with 10% missing values, skewed markers (using $\chi 2$ test with $p < 0.01$) and duplicated markers. Duplicated markers were defined as for which all non-missing dosage scores were equal (van Geest, Bourke, et al., 2017b). A representative marker was created for these duplicated markers and added to the marker list.

2.4 Integrated linkage map construction

The linkage analysis of markers and the construction of linkage maps was performed using the R package polymapR. The polymapR package is a combination of a set of R functions, it relies on dosage data of molecular markers, which is produced by the package fitPoly.

The process of constructing a linkage map mainly consists of four parts: data inspection, linkage analysis, linkage group assignment and marker ordering.
The data inspection is the first step, which starts in reading the dosage-scored SNP data. The correct input dosage-scoring data should include a column of marker dosage for the mother, one for the father followed by a column for each of the offspring of the F1 cross. After that, all the marker dosages were assigned to their simplest forms in order to reduce the number of redundant marker segregation classe (P. M. Bourke et al., 2016). In this case, bi-parental SNP markers were used for genotyping. Hence for each locus 7 possible SNP genotypes can be detected: dosage 0 to 6. Here dosages x (0 – 6) basically has the same effects as dosage (6 – x), however the calculation is easier by using low dosages other than high dosages. For example, simplex x nulliplex (S x N), pentaplex x nulliplex (P x N), simplex x hexaplex (S x H) and pentaplex x hexaplex (P x H) have the same segregation ratio 1:1 hence they all can be assigned to the simplest S x N combination for the convenience of calculation. After the marker conversion, 49 marker segregation types were reduced to 19 in hexaploids (appendix 1).

The linkage analysis is based on the computation of logarithm of odds (LOD) score and of the recombination frequencies. For each combination of two marker types, there are multiple possible phases depending on the conformations in each of the parents: coupling, repulsion, mixed, coupling-coupling, coupling-repulsion, repulsion-coupling, repulsion-repulsion, coupling-mixed and repulsion-mixed. Before the linkage analysis, the phase between two markers is unknown. For every marker combination, recombination frequency (rf) was calculated using Maximum of the

Log-Likelihood (MLL). Phase with the lowest rf was selected and determined as the phase for this marker combination.

We focused on the homologues and integrated chromosomal groups during the linkage group assignment. First simplex x nulliplex (S x N) markers were used to perform the linkage group clustering at the LOD score 10. Multiallelic markers were used as bridge markers to the pairs of S x N markers in order to identify linkage groups (LG). In this case, mainly Duplex x Nulliplex (D x N) and simples x simplex (S x S) markers were used to separate out the coupling phase homologue clusters and reconnect these into chromosomal clusters. Uni-parental D x N markers acted as bridge markers between homologues within each LG. Bi-parental S x S markers provided bridging information between two parents. This was the backbone of the integrated linkage map. Based on this backbone, all the other marker types were assigned in to a LG and phased into a homologue based on linkages with S x N markers with LOD score over 5. The marker ordering was performed by using MDSMap_from_list function, which is a wrapper function of MDSMap (Preedy & Hackett, 2016. MDSMap is a rapid marker ordering approach designed for high density linkage maps. Configurations were set as defaults: Haldane's mapping function and LOD^2 as weights.

2.5 Map quality check

As it is described before, SNP markers were discovered from the assembly of RNA sequencing data. Multiple transcript contigs were sequenced and each SNP marker can be traced back to its original location in a transcript contig. For markers on the same contig should arrive in the same position in the integrated linkage map assuming the contigs were successfully assembled. This provided useful information for checking the quality of the integrated linkage map. All the markers were grouped according to what contigs they belong. For every LG, standard deviations (SD) of the positions of SNP markers on the same contig were calculated, which provided as a criterion to measure the difference of the positions of these markers in the integrated linkage map. A map with good quality should arrive in very few SDs of the marker positions for each contig.

2.6 QTL analysis

2.6.1 Phenotyping

The shape of chrysanthemum leaf can be very diverse and can largely influence the beauty of the plant. In this study we focused on four specific leaf traits: leaf number, stem length, total stipule size and leaf width/length ratio. In total 100 individuals with different genotypes were phenotyped.

Leaf number is the number of leaves on the main stem. For each individual, 3 replicates were randomly pulled out for each genotype/plot, the number of main-stem leaves was counted and the averages were taken.

For stem length, we picked 3 replicates for each individual and measured the length from the bottom to top and took the average.

Total stipule size is the summed-up area of the stipules at the 3rd node of chrysanthemum. We take out all stipules of 3 individuals at the 3rd node, scanned it by Canon CanoScan LiDE 220 scanner and calculated the total area using ImageJ. In each node there are three leaves, they grow against each other in 120 degrees. For each leaf there are two stipules growing at the bottom, at each side of the leaf. So usually we can get 6 stipules in one individual at the 3rd node and overall, 18 stipules for 3 individuals. But that's not all the cases. In some genotypes few or none stipule can be found. Also, there is a possibility that the stipules fall off from the plant during the phenotyping. This is a biological trait and biologists might be more interested about it. Breeders want to remove the stipules because they make chrysanthemum look furry and not beautiful to most of the people.

Leaf width/length ratio, as its name, is defined by the ratio of width divided by length of the chrysanthemum leaf. It indicates the leaf shape (Fig. 2). We took out 1 leaf at the 6th node for each plant, scanned them and calculated the parameter width/length through ImageJ for 3 plants for each genotype and took the average. For convenience, we put all the 3 main stem leaves and 18 stipules of 3 individuals in each plot together and scanned them on the same image (Fig. 3).

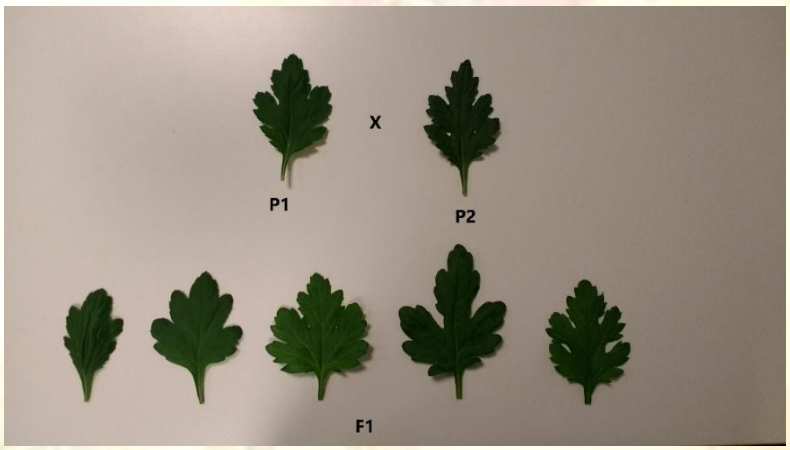

Figure 2. Different shapes of chrysanthemum leaves

Figure 3. Example of scanning

2.6.2 Correlations between the traits

In the next step correlations between any of the four target leaf traits were checked. Although we couldn't imagine any links from the surface, for example, leaf number looks totally unrelated to total stipule size, however, there might be some genetic links of their loci regions. The correlation of determinations between any of the two target traits were calculated. The outcome could be very informative: If the correlation of determinations between two traits is high, then we can assume that there is a high possibility that the QTL regions of these two traits can be overlapped for some degrees.

2.6.3 Marker dosage model

Marker dosage model is also called linear model. It is established in order to seek potential relationship between the dosages of markers and the phenotypical values of target traits:

$$\begin{Bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{Bmatrix} = \alpha + \beta \begin{Bmatrix} m_1 \\ m_2 \\ . \\ . \\ . \\ m_n \end{Bmatrix} + \varepsilon$$

Where:
y represents the phenotipic value for one sample plant
m reprensents the marker dosage number of marker M in different sample plants
α represents the interval

β represents the slope
ε represents the error

For every marker, an Analysis of Variance (ANOVA) was performed. We expect to get a slope value, its p-value and R-square value for every trait. The p-value indicates how well the marker dosage goes with the phenotypic value. 20 most important markers were picked up with the 20 most p-values. These markers have relatively high possibility to be involved in the QTL region for a target trait. R-square describes how much the phenotypic variation could be explained by the model.

2.6.4 Genotype probability model

Genotype probability model is also called IBD model. We first calculated the IBD probabilities to estimate parental haplotypes in the offspring (P. Bourke, 2014). By doing this we created a three-dimensional array for each linkage group with three parameters: marker, offspring individuals and homologue. We picked up the fully informative dosage scores to fill the array. Then, we used inter-marker distance to estimate the IBD probabilities of adjacent markers:

$$\text{If } P_j = 1 \text{ , } P_i = P_j - r_{ij}$$
$$\text{If } P_i = 0 \text{ , } P_i = P_j + r_{ij}$$

Where:
i and j represent the two markers
P represents the IBD probability
r represents the recombination frequency

QTL analysis was performed on mean phenotypic value using an IBD probability model by P. Bourke (2014) for tetraploid level. This model is modified by van Geest, Bourke, et al., (2017) in this case for hexaploid chrysanthemum:

$$Y = \mu + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 X_6 + \alpha_8 X_8 + \alpha_9 X_9 + \alpha_{10} X_{10} + \alpha_{11} X_{11} + \alpha_{12} X_{12}$$

Where: α_i and X_i represent the main effects and indicator variables for allele i.
The parameters of homologues 1 and 7 are set as references.

2.6.5 Model combination and improvement

The marker dosage model and genotype probability model were combined in this case to seek candidate markers associated with leaf traits. For every trait, 20 most significant markers in the Marker dosage model were selected out according to the p-value. For genotype probability model, 10 most significant markers in the detected QTL regions were selected. Markers showed significance in both of the two models were checked as candidate markers. For each detected

QTL region, homologue analysis was performed to check the correspondence with the dosage constitutions of the candidate markers.

2.7 Candidate genes

Multiple genes could be involved in affecting chrysanthemum leaf traits. These genes could be directly regulating or indirectly influencing these traits by other regulating processes. What we can do is to investigate a list of genes affecting the chrysanthemum leaf traits by other plants and see whether they are in chrysanthemum genome. By combining the results from the significant QTL regions in the previous chapter, it is more convenient and precise to see what's popping up in the QTL regions. There could be of higher possibility to find related genes in such regions. We aligned the sequences of the contigs from the significant regions against the Uniprot swissprot annotated database. The result indicated a list of candidate genes and was filtered with the bit-score over 100. The identifiers of these genes were extracted and retrieved for the gene function. The GO biological process terms were filtered by functions which have potentials to affect the leaf traits. A final list of candidate genes gives reference information on estimating genes affecting chrysanthemum leaf traits.

2.8 Software

To increase the efficiency of performing genetic analysis in polyploids, computer methods were used. Various software has been developed to convert the signal from e.g. SNP arrays into discrete dosage scores for polyploids, such as fitPoly, fitTetra (Voorrips et al., 2011), SuperMASSA (Serang, Mollinari, & Garcia, 2012) or ClusterCall (Schmitz Carley et al., 2017). In this case we use fitPoly because this R package can be used for polyploids with higher ploidy levels (6) compared to others.

For linkage map construction R package polymapR is applied. polymapR is an R package for constructing linkage maps for polysomic tetraploids and hexaploids. It consists of 4 parts, data inspection, linkage analysis, linkage group assignment and marker ordering. The input data of polymapR is dosage-scored SNP marker data, which can be generated in fitpoly.

Software ImageJ is used to analyse the scanning pictures and get the phenotypic value of the target traits. The custom R-scripts are used for QTL analysis.

3. Results

3.1 Dosage calling of SNP genotyping data

According to the data from van Geest et al., (2017), 67870 SNPs were successfully called from fitPoly in total from the ratio of signal intensities of the genotyping array. These SNPs were the starting point and were imported to polymapR for constructing linkage map.

3.2 Linkage map

3.2.1 Construction of integrated linkage map

After converting markers to their simplest form, 48022 markers were generated out of 67870 in total. They arrived in 19 marker types (Figure 4):

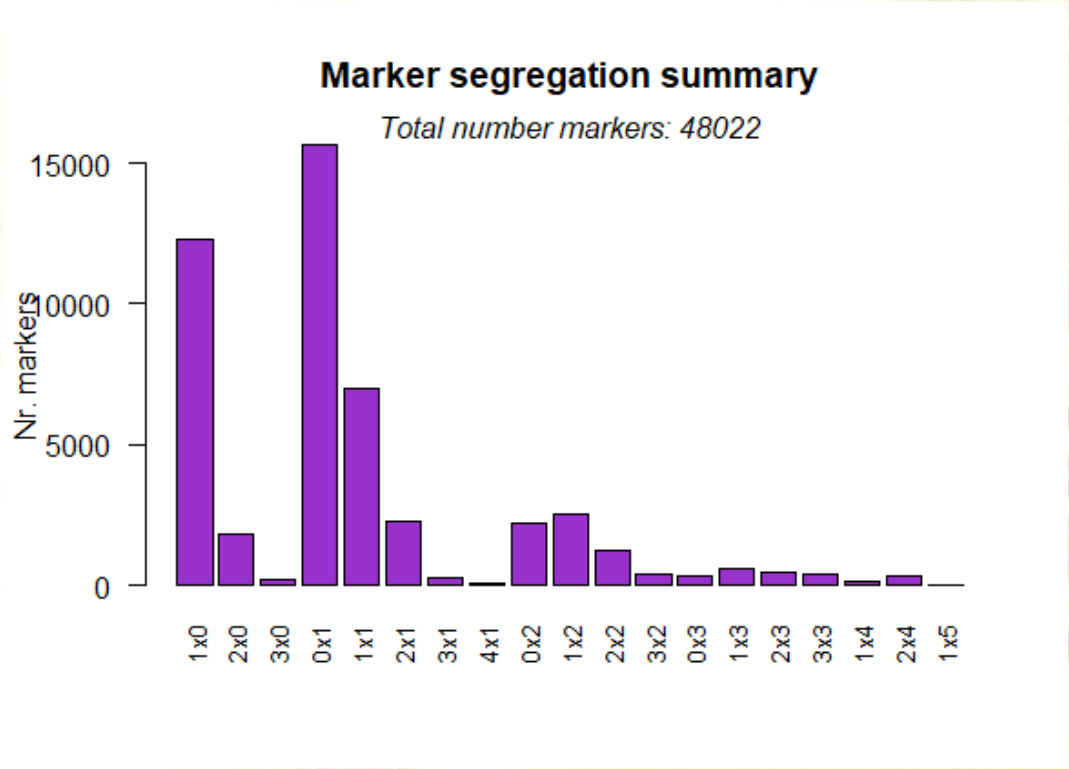

Figure 4. Marker segregation summary after converting markers to their simplest form

Not all these 48022 markers and 417 individuals were good enough to be used for the linkage mapping, hence we progressed the data inspection and selection. The selection steps were operated for markers and individuals separately. For individuals, 6 individuals were screened out for 10% missing values. 1 individuals were screened out for duplicated individuals with cut-off 90%. In the end there were 402 genotypes (400 F1 progenies and 2 parents) left. For markers, after removal of markers with 5% missing values, non-segregating markers, 30532 markers remained in the dataset. 9187 markers were screened out for duplicated markers. They were reduced to a unique representative marker for linkage determination. The rest duplicated markers were added to the integrated linkage map after construction. 21345 non-duplicated markers were prepared for linkage analysis.

Simplex x Nulliplex and Nulliplex x Simplex markers were used to identify and cluster the homologues. 54 homologues have been identified for Parent 1 and 53 for Parent 2. Simplex x Simplex markers were used to identify the linkage groups (LGs). As a result, 9 linkage groups have been identified for both parents. All the other markers were assigned to the LGs afterwards. The different marker types were assigned for each parent separately and combined to an integrated linkage map at the last step. The duplicated markers were removed from the map temporally and assigned to the map with the unique representing markers after map ordering.

The ordering of the markers was based on calculating pairwise recombination frequency, weighed by LOD^2. After calculating the recombination frequencies between these assigned markers in the map with Simplex x Nulliplex markers, In

total 30014 markers were able to be assigned and phased to their expected homologue(s) weighed by at least five significant coupling linkages at LOD 5. The integrated linkage map consists of 9 linkage groups whose lengths are ranging from 64.36 cM to 97.52 cM. Table 1 indicated the summary data of the integrated map.

Table 1. Summary of integrated linkage map

Linkage group	Length(cM)	Phased marker number	Contig number
1	82.36	2516	1259
2	74.88	3080	1161
3	64.36	2772	1054
4	85.02	3197	1265
5	90.20	3398	981
6	91.68	3521	1263
7	81.62	3422	1139
8	88.10	3479	1206
9	97.52	3051	1316
Sum	755.74	30014	10644

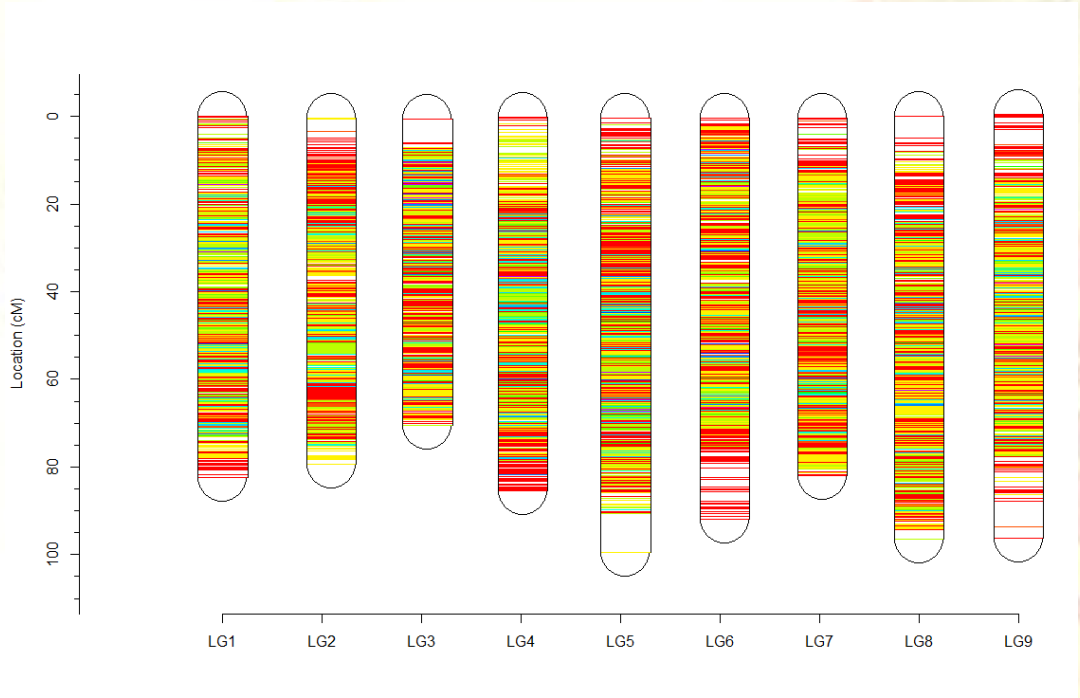

Figure 5. Integrated linkage map based on all marker types

Fig 5. shows the picture of integrated map based on all marker types. As indicated in the legend, the colors bands inside the chromosomes represent different marker types.

3.2.2 Linkage map quality check

According to the construction of the Affymetrix SNP array, the markers originated from the same transcript contig from the RNA-seq assembly should arrive in almost the position in the linkage map. We calculated the standard deviation(sd) of the positions of the markers from the same contig and took the average of these marker positions in the same contig for each LG. For contigs only containing 1 assigned marker in the map, the sd was set as 0. The result is shown in Fig 6.

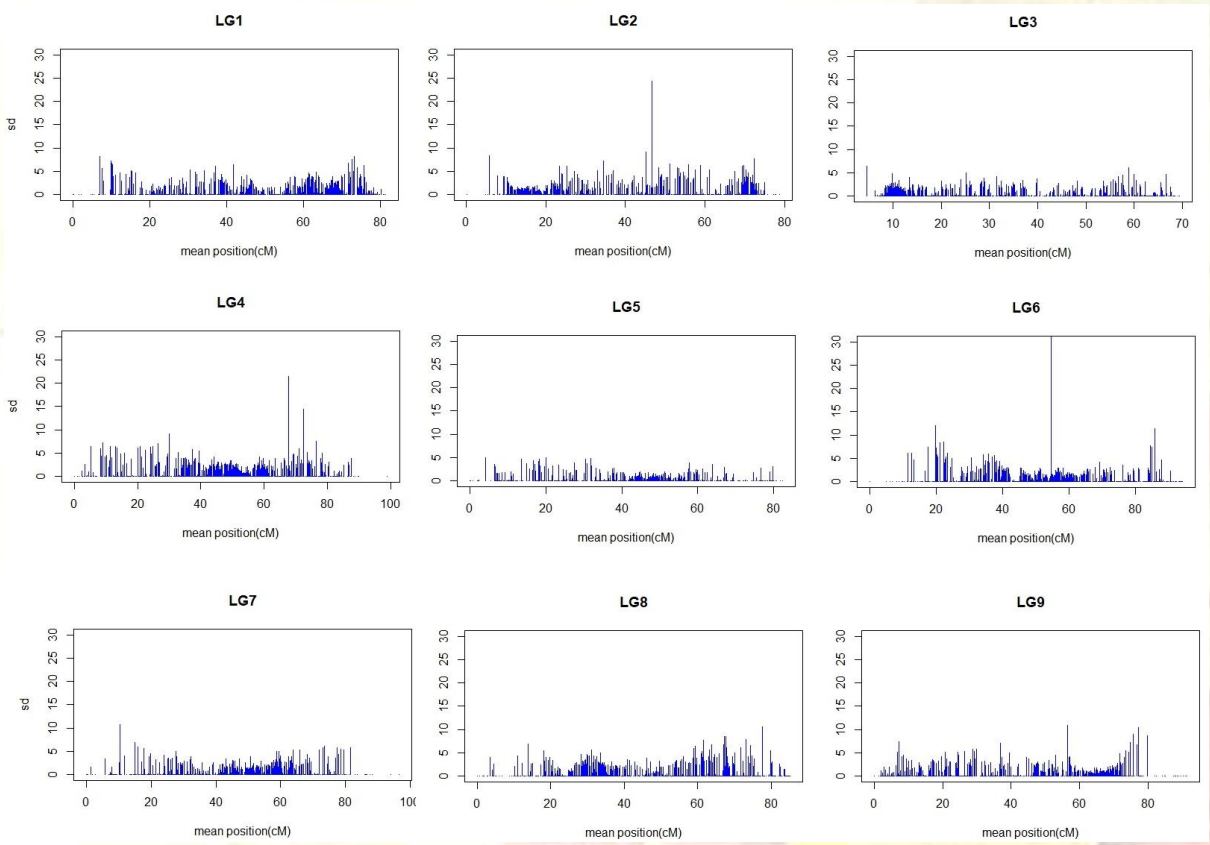

Figure 6. Standard deviations of contig positions for each Linkage Group

The X-axis stands for the mean position of markers from the same contig, the Y-axis stands for the standard deviation of the marker positions from the same contig. As we can see from the table, the linkage map is of good quality. Most of the standard deviations of contig positions in each Linkage groups are controlled well below 10, which indicates that the markers from the same contigs were mapped close together. However, there are 4 sd peaks on LG2, 4 and 6, which indicates markers were diversely spreading of these four contigs in the linkage map. For the markers belonging to this four contig groups, they are all SxN or NxS markers tagging homologues 1 and 12. It is inferred that the extreme dosage distribution hindered these markers mapped in same position on the integrated map.

3.3 QTL analysis on four leaf traits

3.3.1 Phenotyping

The QTL analysis was performed for four leaf traits in chrysanthemum: leaf number, stem length, stipule size and leaf width/length ratio. In total there were 100 individuals been phenotyped. The histograms of phenotyping data on the four leaf traits are shown in Fig. 7 - 10. As we can see from the figure, the phenotypic values of traits Leaf number, Stem length and Width/length ratio have the shape of normal distribution, which indicates the phenotyping data is of good quality. For trait Total stipule size, the phenotyping data is left-biased.

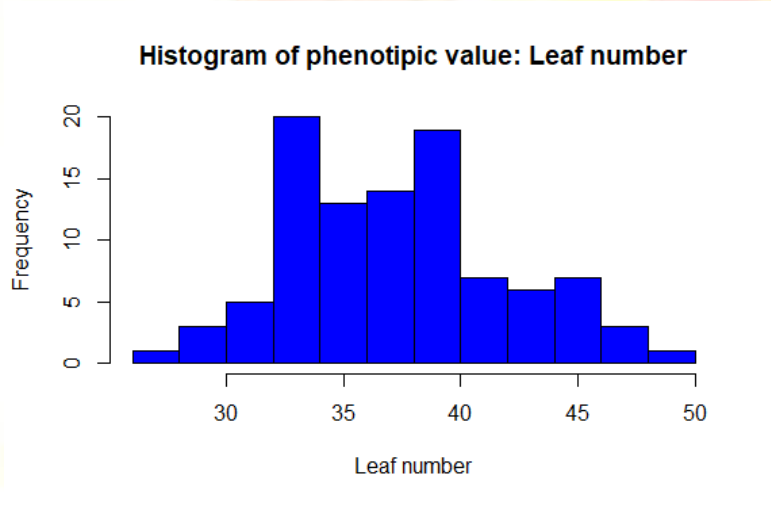

Figure 7. Distribution of leaf number trait value

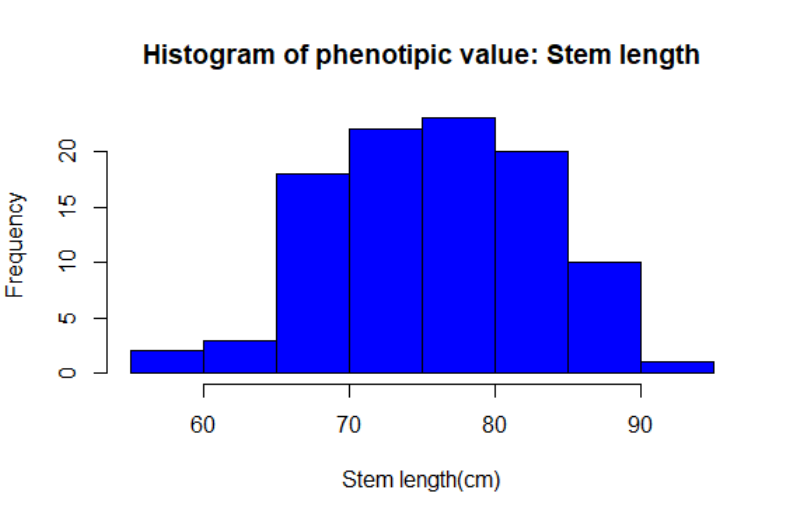

Figure 8. Distribution of stem length trait value

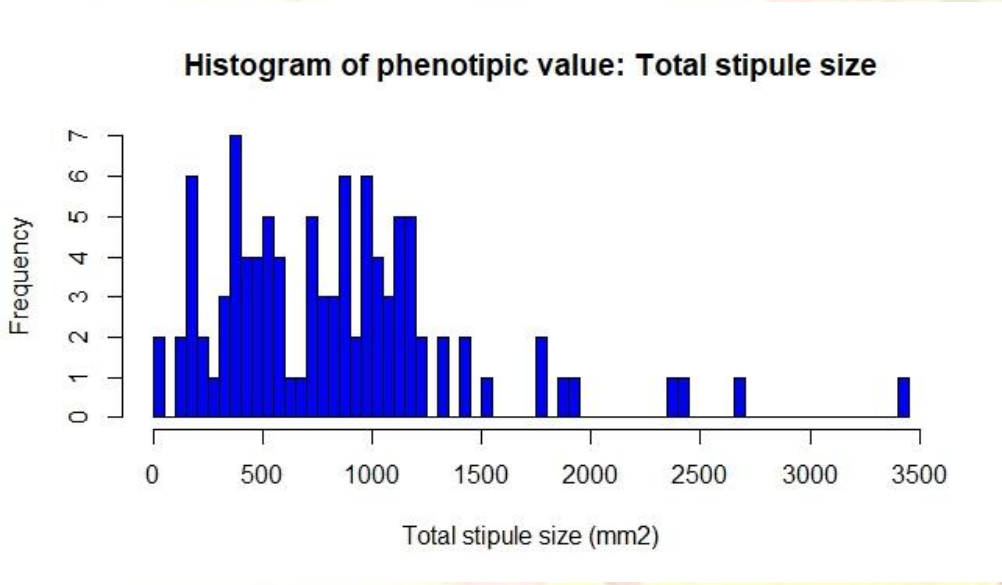

Figure 9. Distribution of total stipule size trait value

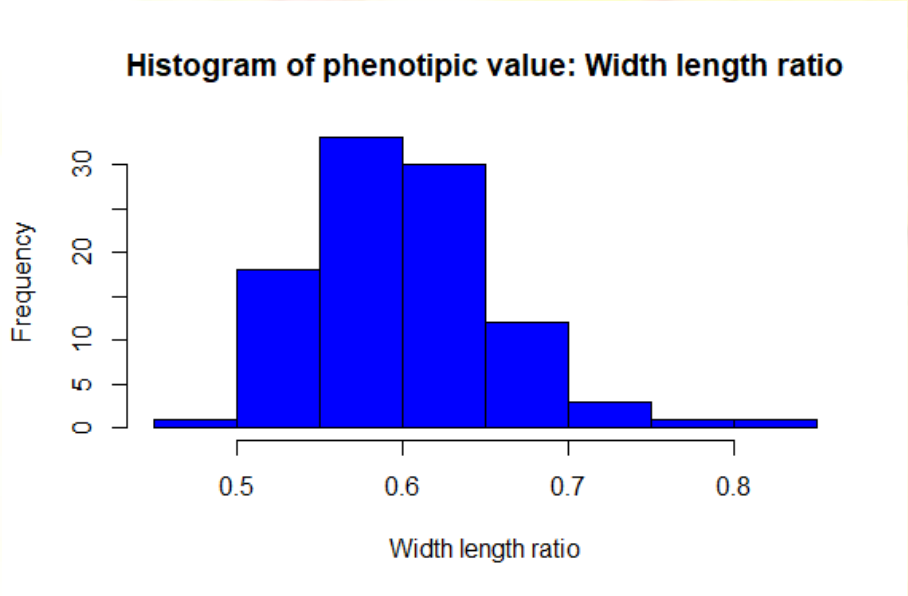

Figure 10. Distribution of width/length ratio trait value

3.3.2 Analysis of correlations between every two leaf traits

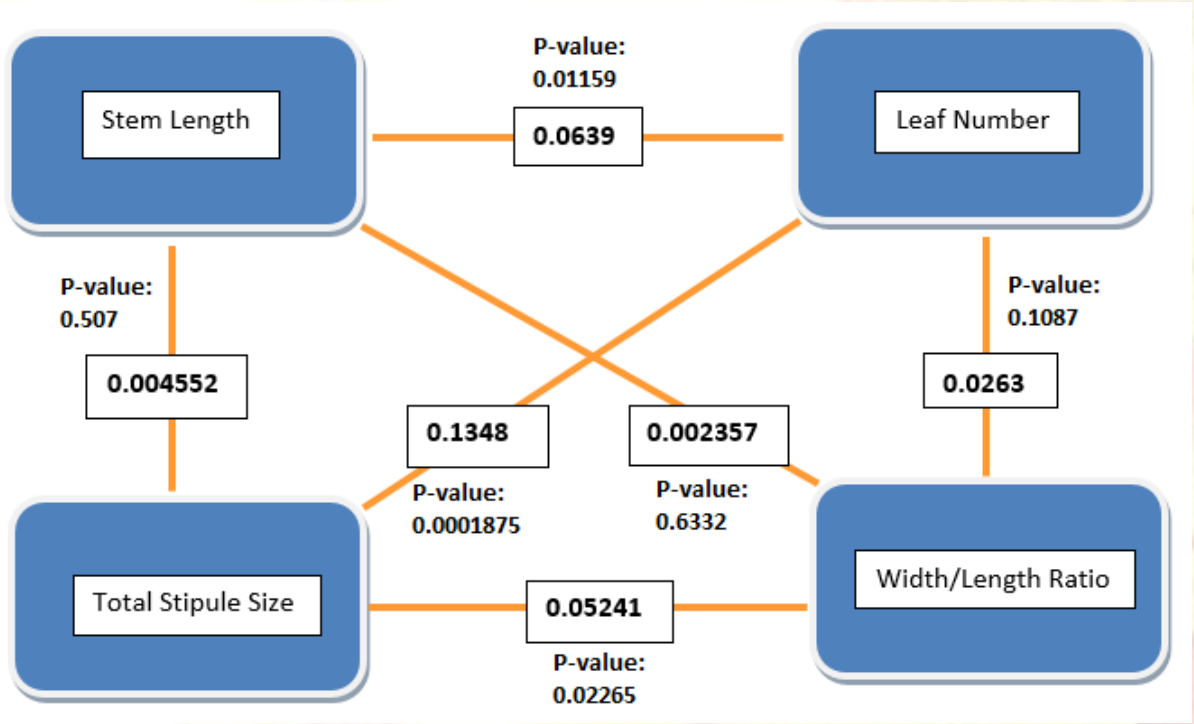

Figure 11. R-square values among the leaf traits

Fig.11 shows the correlation of determination (R^2) between every two traits and their p-values. In total there are six trait combinations of traits, neither of them shows a very strong correlation. The strongest correlation is between trait leaf number and total stipule size, which is 0.1348, which indicates 13.48% varience of trait leaf number could be explained by trait total stipule size. The distribution of leaf number with Total stipule size phenotypic value is shown in Figure 13.

It suggests that the QTL regions of trait leaf number may be partly overlapped with trait total stipule size. What have to be stressed is that correlation of determination does not logically infer whether one variable moves in response to another. It only gives a possible explanation and indicates no strict dependent relationship between trait leaf number and total stipule size (Mukaka, 2012).

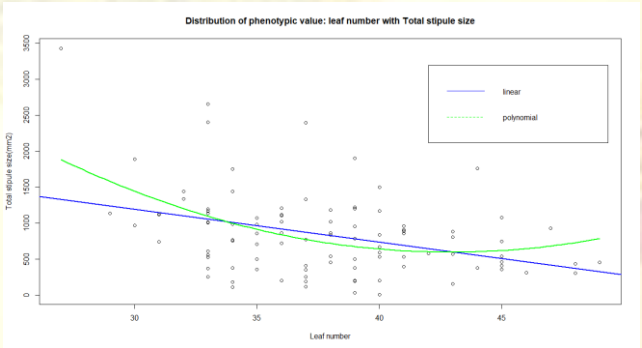

Figure 12. Distribution of phenotypic value: leaf number with Total stipule size.

3.3.3 Linear model of each marker with four leaf traits

A linear model was established for each marker in order to seek potential relationship between marker dosages and the target traits. For each leaf trait, we extracted the most significant 20 markers with highest p-value which are related to it. These markers have high possibilities to be involved in the QTL region for a target trait. The results are shown in Appendix 2.

3.3.4 QTL detection

Fig 13 - 16. shows the QTL plots on four leaf traits. Two models were combined together: the IBD model (indicated by blue line) and marker dosage model (indicated by grey dots). The thresholds are indicated by the dashed lines, the colors are coordinated with the color of the model. For setting QTL thresholds, 100 permutation tests were performed with ramdom phenotypic values.

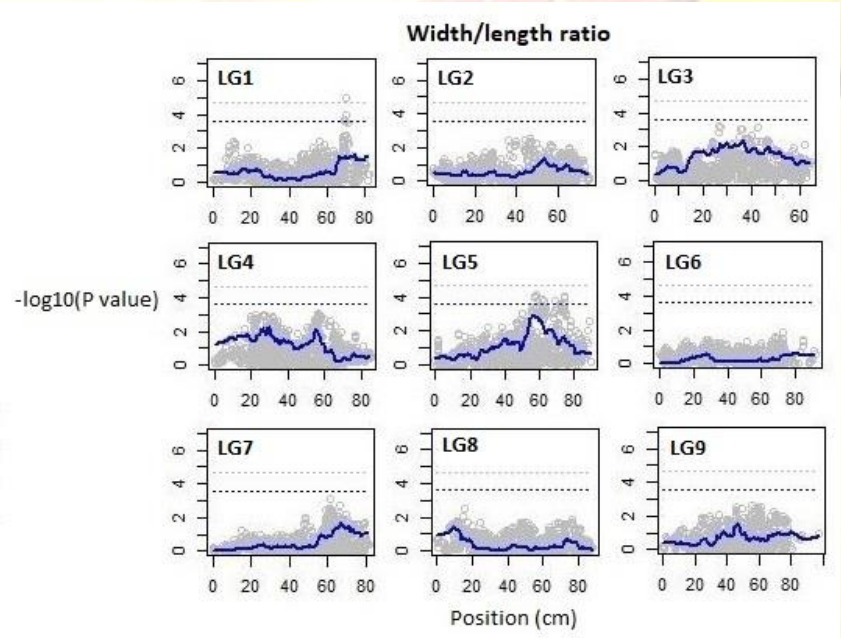

Figure 13. QTL analysis on four leaf traits: leaf width/length ratio.

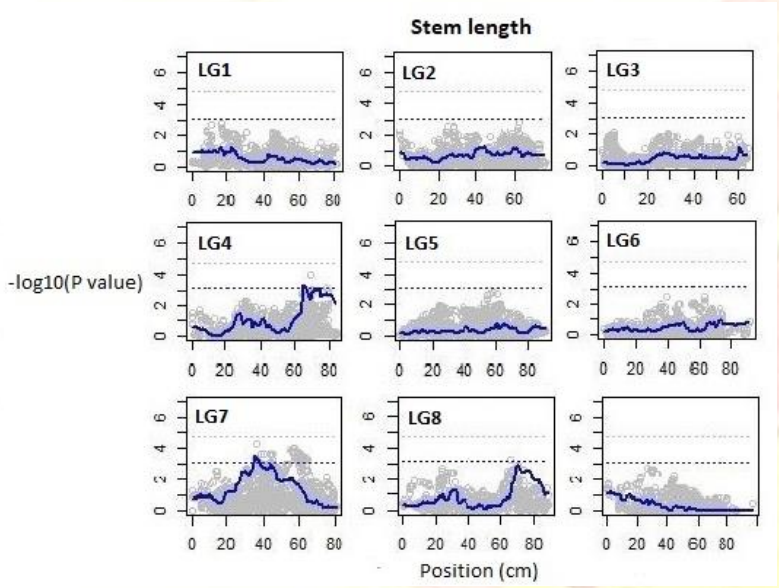

Figure 14. QTL analysis on stem length.

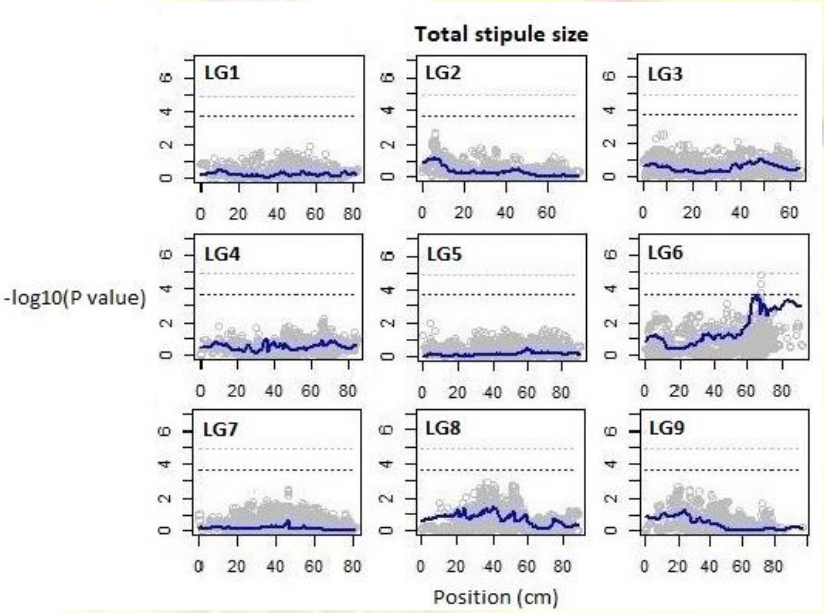

Figure 15. QTL analysis on total stipule size

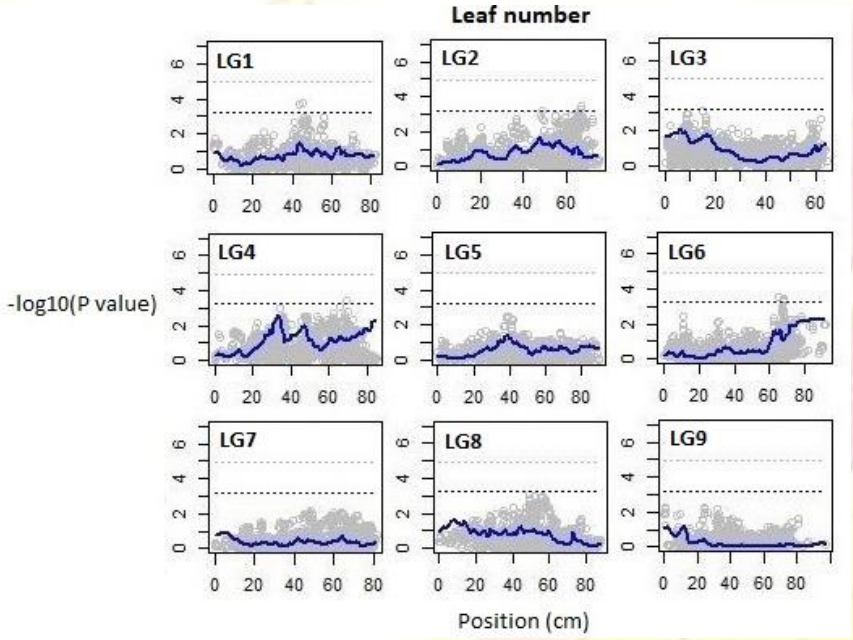

Figure 16. QTL analysis on leaf number.

For each trait, from left to right and from up to down lie LG 1 to 9 respectively. Two models are indicated with blue lines (IBD model) and grey dots (marker dosage model). Significant thresholds are indicated with blue dashed lines (IBD model) and grey dashed lines (marker dosage model).

As we can see from the plot, most of the LG regions are under the thresholds and are not significantly associated with the traits. For trait leaf width/length ratio and leaf number, clearly there is no QTL region has been found. In stipule size, on LG9 there seems to be a minor QTL region at position 65cM. However, the p-value (0.0002501646) is slightly below the threshold (0.0002233167).

For the trait stem length, two QTL regions were found: LG4 at location 35 cM and LG7 at location 65cM (with p-value 0.0002849656 and 0.0004309191), 29.82% and 22.59% phenotypic variation could be explained by these two QTL regions respectively. On LG7, homologue 4 and 8 were proved to have positive effects on stem length and homologue 5 has a negative effect. On LG4 homologue 1 and 9 denote positive effects and homologue 3 denotes a negative effect.

Table 2. Summer statistics of different QTL

LG	QTL peak position (cM)	p-value	p-value threshold	adjusted R^2	associated homologues*
Width length ratio					
	None		0.0002593693		
Stem length					

4	65	0.00043 09191	0.0007521 758	0.29817 27	1+, 3-, 9+
7	35	0.00028 49656	0.0007521 758	0.22589 29	4+, 5-, 8-
Stipule size					
*9	65	0.00025 01646	0.0002233 167	0.22879 08	
Leaf number					
	None		0.0005887 219		

*Homologues presenting a significant impact on phenotypes (p<0.01), "+" indicates a positive effect and "-" indicates a negative effect

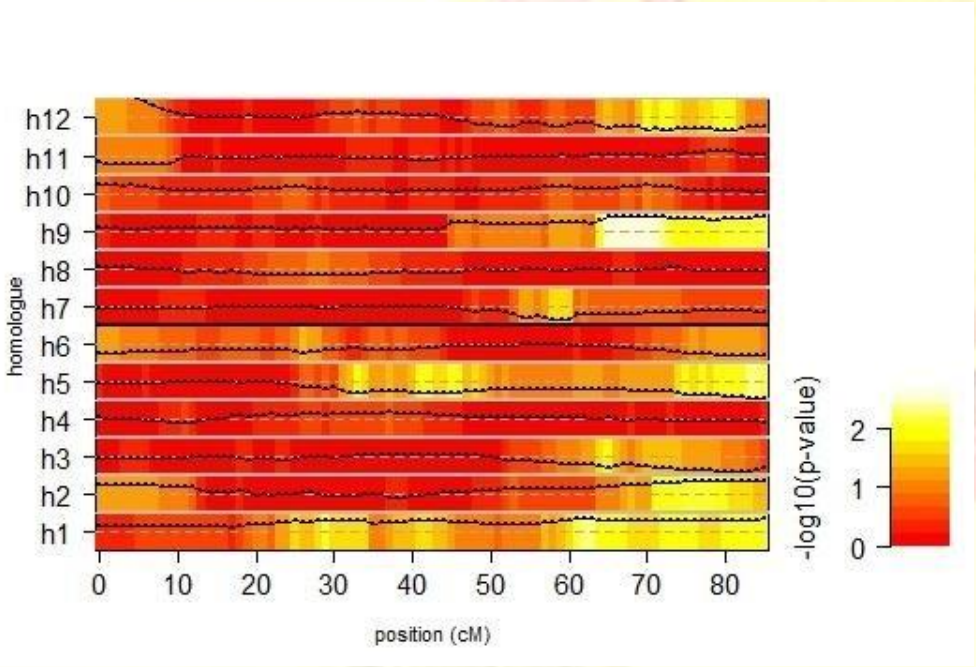

Figure 17. Analysis per homologue for Stem length QTL on LG4.

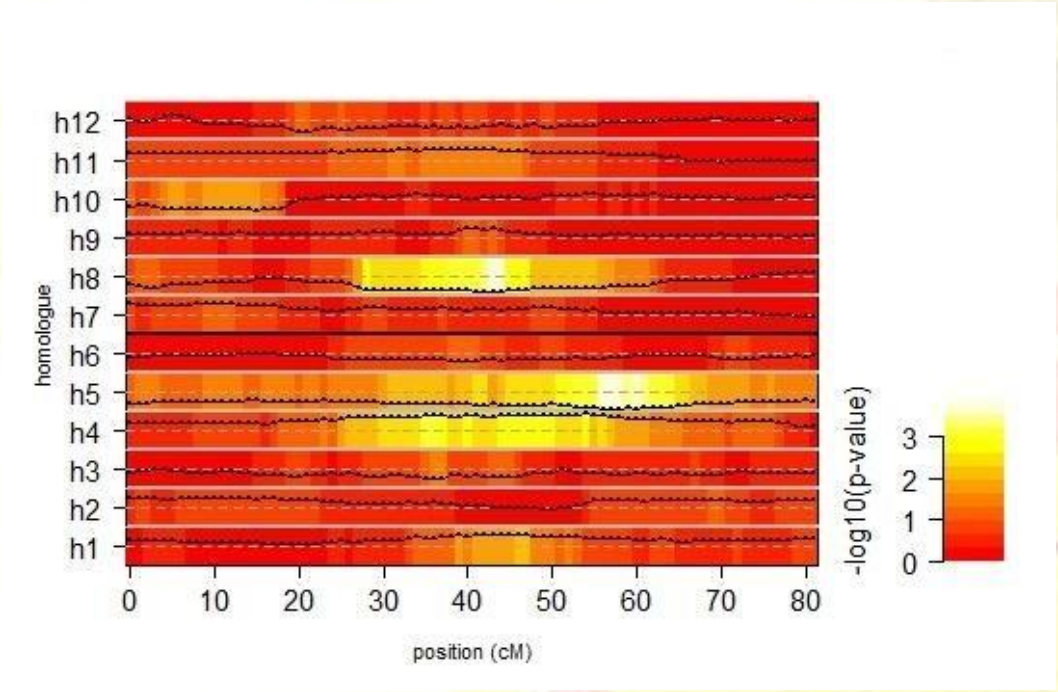

Figure 18. Analysis per homologue for Stem length QTL on LG7.

Figure 17 and 18 indicate the analysis per homologue for stem length on LG4 and LG7. The colours represent the range of the p values, which indicate the significance of explained variance of IBD probabilities of each homologue for trait stem length phenotypic values. The more yellow, the more significant, the more red, the less. Here we define homologue regions with -log10(p-value)>2 as significant. The black dotted lines represent the estimated effect of alleles on the phenotype. The horizontal grey dotted lines indicate zero effects of alleles on phenotype. The situation that black line is above the grey line indicates positive effect of the allele, *vice vera*. The black boxes of each homologue show the range of the positive and negative effect.

3.3.5 Candidate markers: markers significant in both models

On LG4 and LG7, the most significant 10 markers with peak p-values from genotype probability model were indicated in Table 3 and Table 4.

Table 3. Markers with peak p-values on LG4 of stem length.

Markers	Position	Pvals	Dosage
Cm54107_c0g1i1_454_Rnn	69.20389	0.0001081555	1x1
Cm39099_c0g2i3_409_Rnn	78.99168	0.0006423178	1x1
Cm11646_c0g3i1_277_Sn	69.78912	0.0008356710	2x0

Cm63043_c1g2i1_1075_Sn	79.16417	0.0011702272	1x1
Cm33966_c0g1i1_588_Pn	70.00404	0.0014190796	0x2
Cm40117_c2g1i1_1153_Qn	64.20679	0.0016191883	2x0
Cm14389_c0g1i1_427_Sn	79.54923	0.0016544507	1x1
Cm29986_c0g1i2_412_Sn	64.63341	0.0018432918	0x1
Cm34599_c0g2i1_591_Sn	64.60879	0.0021758025	0x1
Cm36834_c0g1i1_713_Rnn	68.40221	0.0023465271	2x0

Table 4. Markers with peak p-values on LG7 of stem length.

Markers	Position	Pvals	Dosage
Cm38566_c0g1i1_1004_Rnn	36.56636	5.453746e-05	2x1
Cm42710_c0g3i1_348_Sn	57.08772	9.168232e-05	1x0
Cm41731_c0g1i1_1546_Sn	56.95815	9.198393e-05	1x0
Cm41924_c0g1i1_534_Sn	54.88245	1.239819e-04	1x0
Cm58550_c1g1i4_425_Sn	57.38526	1.401005e-04	1x0
Cm30161_c0g1i1_236_Rnn	57.53419	1.493650e-04	1x0
Cm54019_c0g1i1_1098_Rnn	56.29650	1.588180e-04	1x0
Cm62939_c0g1i1_948_Sn	55.83534	1.588180e-04	1x0
Cm54019_c0g1i1_99_Sn	56.29378	1.588180e-04	1x0
Cm54019_c0g1i1_283_Sn	56.29650	1.588180e-04	1x0-

Previous study shows the top 20 markers whose dosages have significant influence on stem length according to the linear model (Appendix 2.). After integrated linkage map construction, some markers were not shown in the map. For top 20 significant markers for stem length in linear model, in total there were 10 markers remained in the linkage map, their positions are shown in table 5. The other 10 markers were wiped out because of containing more than 10% missing values during the linkage map construction.

Table 5. Positions of significant markers of linear model remained in the linkage map for trait stem length.

Marker name	LG	Position(cM)
Cm54107_c0g1i1_454_Rnn	4	69.20389
Cm4088_c0g1i3_841_Sn	7	44.72649
Cm36765_c0g1i2_1378_Sn	7	36.26472
Cm38566_c0g1i1_1004_Rnn	7	36.56636
Cm42659_c0g4i1_1038_Sn	7	44.65858
Cm5596_c0g4i1_1231_Sn	7	44.48929
Cm10191_c0g4i1_174_Sn	7	44.9161
Cm19355_c2g11i1_300_Sn	7	43.25204
Cm19355_c2g11i2_218_Sn	7	43.46249
Cm5596_c0g4i1_1338_Sn	7	44.17874

So far two lists of significant markers for trait stem length were generated from linear model and genotype probability model (Table 5. and Table 3&4). Markers occuring on both of these two lists could be considered of high chance associating with the detected QTL regions of trait stem length. After comparing the two lists, marker “Cm38566_c0g1i1_1004_Rnn” and “Cm54107_c0g1i1_454_Rnn” show significance in both of the two models, their phased dosage constitutions are shown in Table 6.

Table 6. The dosage constitutions of promising marker.

Marker	LG	Position(cM)	Marker type	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12
Cm54107_c0g1i1_454_Rnn	4	69.20389	S x S	1								1			
Cm38566_c0g1i1_1004_Rnn	7	36.56636	D x S					1	1		1				

Marker “Cm38566_c0g1i1_1004_Rnn” is a Duplex x Simplex marker tagging homologues 5, 6 and 8. It locates on LG7 at the position 36.57cM on the linkage map. At this position, homologues 6 shows little contributions for the phenotypic values (red), homologue 5 doesn’t show strong effect on phenotype (orange) and homologue 8 indicates strong significance (yellow) with negative effect. Consequently, the alleles tagged by this marker do not all affect the phenotype significantly. This marker should not be considered useful associating with the QTL regions of stem length.

Marker “Cm54107_c0g1i1_454_Rnn” is a Simplex x Simples marker tagging homologues 1 an 9 and locates on LG4, 69.20 cM. At this position, homologue 1 and 9 all indicate significant influence on the phenotype (yellow and white). In conclusion, marker “Cm54107_c0g1i1_454_Rnn” is of high probabilities for linking the QTL of stem length, hence it may be used to track this quantitative trait.

The boxplot of marker “Cm54107_c0g1i1_454_Rnn” with the phenotypic value in marker dosage model is shown in Fig,19. It explains 14.31% phenotypic variation.

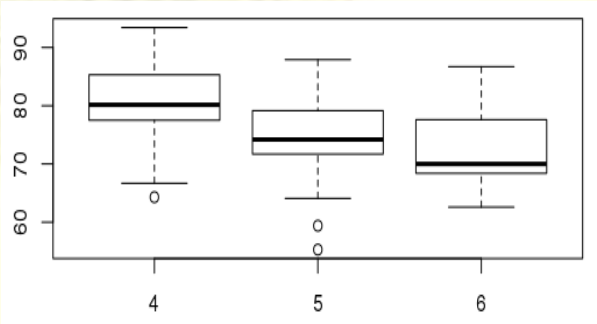

Figure 19. The boxplot of the phenotypical value of stem length with the dosage of marker “Cm54107_c0g1i1_454_Rnn”

3.3.6 Candidate markers: other markers

By combining the significant marker lists of linear model and genotype probability model we successfully acquired two markers occurring in both lists and one of them was proved to be significantly associated with trait stem length. Compared to these two markers, the rest markers on the significant marker lists of the two models could also be possible to associate with the trait. These markers were divided in three parts: 18 markers in genotype probability model, 8 markers in marker dosage model remained in the linkage map and 10 markers in marker dosage model missing in the linkage map.

For the rest 18 markers in the significant marker lists of genotype probability model apart from marker "Cm38566_c0g1i1_1004_Rnn" and "Cm54107_c0g1i1_454_Rnn", 9 markers are on LG4 and 9 on LG7. Their phased dosage constitutions were examined and compared to the per homologue analysis on LG4 and LG7 (Fig. 16 & 17). The result indicates that 6 markers on LG4 and all 9 markers on LG7 are tagging QTL regions for trait stem length. The results are shown in Table 7 and 8.

Table 7. Candidate markers of genotype probability model on LG4 of stem length.

Markers	Position	Dosage	Tagging homologues	R-square
Cm39099_c0g2i3_409_Rnn	78.99168	1x1	1,9	0.1135
Cm63043_c1g2i1_1075_Sn	79.16417	1x1	1,9	0.1065
Cm14389_c0g1i1_427_Sn	79.54923	1x1	1,9	0.09918
Cm29986_c0g1i2_412_Sn	64.63341	0x1	9	0.07949
Cm34599_c0g2i1_591_Sn	64.60879	0x1	9	0.07607
Cm36834_c0g1i1_713_Rnn	68.40221	2x0	1,2	0.1054

Table 8. Candidate markers of genotype probability model on LG7 of stem length.

Markers	Position	Dosage	Tagging homologues	R-square
Cm42710_c0g3i1_348_Sn	57.08772	1x0	5	0.1273
Cm41731_c0g1i1_1546_Sn	56.95815	1x0	5	0.1345
Cm41924_c0g1i1_534_Sn	54.88245	1x0	5	0.1245
Cm58550_c1g1i4_425_Sn	57.38526	1x0	5	0.1234
Cm30161_c0g1i1_236_Rnn	57.53419	1x0	5	0.1208
Cm54019_c0g1i1_1098_Rnn	56.29650	1x0	5	0.1187
Cm62939_c0g1i1_948_Sn	55.83534	1x0	5	0.1187
Cm54019_c0g1i1_99_Sn	56.29378	1x0	5	0.1187
Cm54019_c0g1i1_283_Sn	56.29650	1x0	5	0.1187

For the significant markers marker dosage model which are remained in the linkage map, 6 out of 8 markers were proved to be associated with the QTL regions, the result is indicated in Table 9.

Table 9. Candidate markers of marker dosage model remained in the linkage map for trait stem length.

Marker name	LG	Position(cM)	Tagging homologues	R-square
Cm42659_c0g4i1_1038_Sn	7	44.65858	8	0.1404
Cm5596_c0g4i1_1231_Sn	7	44.48929	8	0.1404
Cm10191_c0g4i1_174_Sn	7	44.9161	8	0.1404
Cm19355_c2g11i1_300_Sn	7	43.25204	8	0.1404
Cm19355_c2g11i2_218_Sn	7	43.46249	8	0.1404
Cm5596_c0g4i1_1338_Sn	7	44.17874	8	0.1385

For the 10 significant markers of the marker dosage model, they are missing in the linkage map. We need to reconstruct the linkage maps by adding these 10 markers manually after the screening process.

However, there are 6 markers that still cannot be arranged into the linkage map. Four of them have high dosages which couldn't be converted to their simplest form. One possessed too many missing values and it was screened out by linkage function at the threshold 3 (LOD>3) in polymapR. Besides, one marker is in the final maplist but it cannot be generated into the phased maplist because polymapR cannot distinguish significant homologues it may locate on.

Consequently, we focus on the remaining 4 markers. After performing QTL analysis, we compared the dosage constitutions of these markers with the analysis per homologue on LG7. The result indicates that none of these four markers coordinate with the regions on the homologues which significantly linked to the stem length trait.

By comparing the R-square of these candidate markers occurring in only one of the two models with the significant marker ('Cm54107_c0g1i1_454_Rnn') occurring in both models, we can see that the marker 'Cm54107_c0g1i1_454_Rnn' is able to explain the most phenotypic variations (14.31%). Consequently, it suggests that by using one model we could arrive in significant markers, however only by combining the two models together we could find the most significant one.

3.3.7 A possible solution on marker dosage model improvement: combined-marker analysis

The linear model has a feature: a marker can only be detected significant if they're tagging the homologues with the same directions of effect. For markers tagging homologues with opposite directions, although they are tagging significant regions of homologues, would still be regarded as not significant. In other words, we only associate the dosage information with the phenotypic value, the phased homologue consititution is not been considered. This could lower the explained variation on phenotype.

One possible modification would be combining two markers whose dosage are spreading on specific homologues. By applying linear model to the combined dosage information of two markers with phenotypes, explanation of phenotype by genotype could be possibly promoted.

In this study, it was more convenient to focus on LG7 than LG4 because the detected QTL regions are spreading less complicatedly. In LG7, from 30 cM to 50 cM three homologues were proved to have QTL regions located: homologue 4, 5 and 8. Among them homologue 4 has positive effect, homologue 5 and 8 have negative effects. In the previous study we have investigated that marker "Cm38566_c0g1i1_1004_Rnn" is tagging homologues 5, 6 and 8. We also picked out the most significant markers in the linear model which are tagging homologue 4. The information of these three markers was listed in Table 10.

Table 10. Information of three markers to be combined

Marker name	Marker type	Tagging homologues	Position
Cm38566_c0g1i1_1004_Rnn	D x S	h5 (-), h6, h8 (-)	LG7 36.57 cM
Cm17795_c0g1i1_253_Sn	S x D	h4 (+), h7, h10	LG7 61.28 cM
Cm13465_c0g2i3_318_Sn	S x D	h4 (+), h10, h11	LG7 61.59 cM

Here we could arrive in two marker combinations: 1) "Cm38566_c0g1i1_1004_Rnn" and "Cm13465_c0g2i3_318_Sn" and 2) "Cm38566_c0g1i1_1004_Rnn" and "Cm17795_c0g1i1_253_Sn". For both of the two combination, there are 4 (0,1,2,3) x 4 (0,1,2,3) = 16 possible dosage combinations. We reran the linear model by associating the combined dosages of these two marker combinations with the phenotypic value of trait stem length. The boxplots of the phenotypic value of stem length with the combined dosages of the two marker combinations are shown as Fig. 20 and 21.

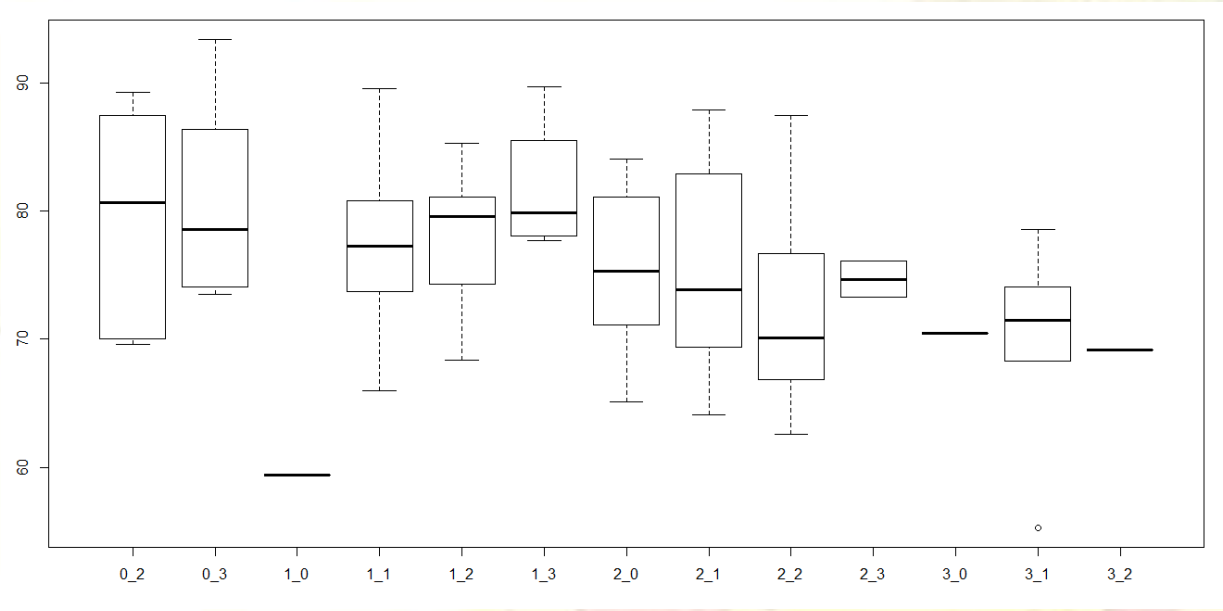

Figure 20. The boxplot of the phenotypical value of stem length with the dosage of combined marker "Cm38566_c0g1i1_1004_Rnn" and "Cm13465_c0g2i3_318_Sn".

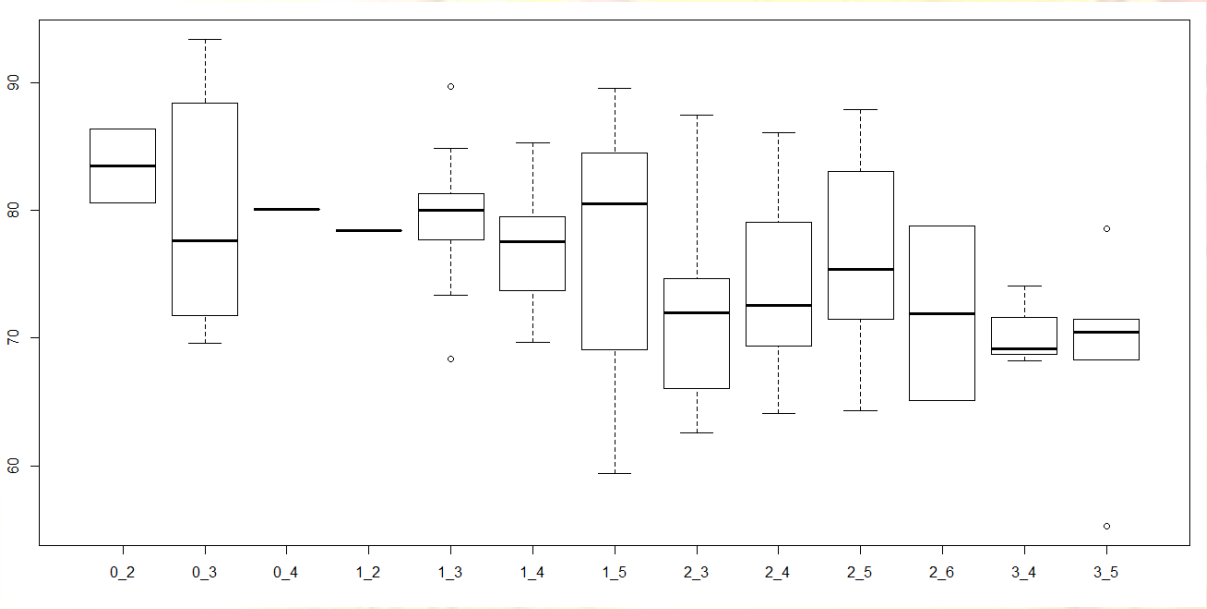

Figure 21. The boxplot of the phenotypical value of stem length with the dosage of combined marker "Cm38566_c0g1i1_1004_Rnn" and "Cm17795_c0g1i1_253_Sn".

The R-square value of each model shown in Table 11. The result indicates that the linear model based on combined marker dosage has the potential to explain the phenotypic variation better than the linear model based on single marker dosage (0.1271 > 0.1247). However, genotype probability model still excels the linear model with the R-square 0.2982 (LG4 65cM) and 0.2259 (LG7 35cM).

Table 11. R-square of each model

Model	R-square
Linear model: marker "Cm38566_c0g1i1_1004_Rnn"	0.1247
Modified linear model: Combined marker "Cm38566_c0g1i1_1004_Rnn" and "Cm13465_c0g2i3_318_Sn"	0.1272
Modified linear model: Combination: "Cm38566_c0g1i1_1004_Rnn" and "Cm17795_c0g1i1_253_Sn"	0.007619
genotype probability model	0.2981727 (LG4 65cM) 0.2258929 (LG7 35cM)

3.3.8 Investigation of prospect genes related to chrysanthemum stem length

In order to further investigate the prospective genes on the two QTL regions of trait stem length, a list of 321 contigs were picked out from these two regions out of the linkage map, representing markers locating on the significant QTL regions of trait stem length. Of them 194 contigs are located on LG4 and 127 are on LG7. These contigs have been aligned against the Uniprot swissprot annotated database. The results indicate 46946 aligned genes. Two parameters were considered for the alignment quality: E-value and bit-score. E-value is the expected number of local alignments with a given score that would be found in random sequences of the same length as the query and database. It reflects how likely a given alignment is due to an evolutionary relationship rather than chance similarity. The maximum E-value of the blast results is 1.000e-05, which can be regarded as significant still. The bit-score is the required size of a sequence database in which the current match could be found just by chance. The higher the bit-score, the better the sequence similarity. The aligned genes were filtered by bit-score cut-off 100. In total there were 21231 genes remained. Their identifiers were extracted and retrieved for the gene function. These identifiers were successfully mapped to 9501 UniProtKB IDs. Gibberellins are plant hormones that can promote the plant growth (Hedden & Sponsel, 2015). The GO biological process terms were filtered by 'gibberellin metabolic process' under 'Biological process'. The final candidate genes are shown in Appendix 3.

These genes can give very informative hints for genetic deduction on genes responsible for stem length on chrysanthemum. Before the realization of chrysanthemum sequenced genome, this can be very helpful to narrow down searching scope for genes responsible for stem length. Further, genes which are responsible for other quantitative traits can also be applied with this methodology.

4. Discussion

4.1 linkage map of chrysanthemum

With the development of DNA informed breeding, genetic analysis in polyploids has been achieving progress. In this study, an integrated linkage map was developed for chrysanthemum with the help of genotyping data from a 183k SNP array. Previously, three linkage maps have been published (Zhang et al., 2010; Zhang, Chen, Chen, Fang, Chen, et al., 2011; van Geest, Bourke, et al., 2017). A preliminary genetic linkage map was constructed by using dominant marker RAPD, ISSR and AFLP (Zhang et al., 2010). Two linkage maps for the two parents were established separately covered by 210 and 190 markers each. 64 intercross markers shared by two parental maps were used to construct the integrated map. In another study, two separate genetic linkage maps for chrysanthemum cultivars 'Yuhualuoying' and 'Aoyunhanxiao' were constructed with the same method, however based on using SRAP markers. For these two separate linkage maps, in total 500 SRAP markers were used covering 1900 cM in both maps. One map consists of 57 linkage groups and the mean inter-marker distance is 6.9 cM. The other map has 55 linkage groups with a inter-marker distance 6.6 cM. Compared to the preliminary map based on RAPD, ISSR, and AFLP markers and the two separate linkage maps using SRAP markers, our integrated map has the following advantages: 1) *More condensed.* With 30014 markers successfully saturated in 9 linkage groups and 54 homologues, our map arrives into an ultra-dense integrated linkage map. The preliminary linkage map only possesses 336 markers in total and is much less condensed. Integrated map with higher density will no doubt be beneficial to the further genetic studies. 2) *Better choice of marker type.* In this two studies, RAPD, ISSR, AFLP and SRAP markers were used. Compared to SNP marker, these markers are difficult to transfer and integrating different linkage maps is very difficult (van Geest, Bourke, et al., 2017a). SNP marker is abundantly saturated in the genome and it can be flawlessly applied in high throughput genotyping. 3) Providing better basis for QTL analysis. An integrated linkage map with high density would promote the accuracy of calculating the genotype probabilities and is therefore more precise at estimating the significant QTL region.

Van Geest, Bourke, et al., (2017a) constructed the first ultra-dense integrated linkage map for chrysanthemum. It arrived in 30312 markers saturated in 9 linkage groups with total length 752.1 cM. The SNP genotyping data used in this study is the same as the first ultra-dense integrated linkage map. This study arrived in an integrated linkage map with 30014 markers and total length 755.74 cM, which is almost the same to this map.

4.2 Chrysanthemum leaf traits and QTL analysis

For chrysanthemum, the success of the plant is mainly determined by its abundant diversity of flower types, colors and plant architectures (Zhang et al. 2010). Leaf traits are very important parts of the plant architecture. Not only the shape of the leaves can largely affect the beauty of chrysanthemum, but also the leaf architecture strongly affects suitability of a plant for cultivation (F. Zhang, J. Jiang, S. Chen, 2012). However, most studies were focused on flower traits and other traits that affect plant architecture like plant width, the ratio of plant height to plant width, internode length (Zhang, et al., 2010), initial blooming time, the duration of flowering (Zhang, Chen, Chen, Fang, Deng, et al., 2011), flower neck length (Zhang et al., 2012), flowering time (Zhang et al., 2013) and so on. Compared to these studies, less researches have been focused on chrysanthemum leaf traits. F. Zhang et al., (2010) performed single-locus and epistatic QTL analysis on plant height. Same measurement method was taken for plant height as stem length in this study, so we could describe them as the same trait. According to the single-locus QTL analysis, two QTLs were detected explaining 6% and 13.2% phenotypic variation. In the epistatic QTL analysis, 3 pairs of digenic epistatic QTLs were found, explaining phenotypic variation from 3.5% to 11.8%. In comparison, the phenotypic variation explained by the two QTL regions detected in this case are 22.59% and 29.82%. However, this study was based on the linkage map based on SRAP markers (Zhang, Chen, Chen, Fang, Chen, et al., 2011), in which the constitution of the map and the nomination of the linkage group and markers are completely different from the integrated linkage in this case. Hence the comparisons of detected QTL regions for trait plant height/stem length in these two studies needs more investigations. Another study focused on QTL analysis with chrysanthemum leaf traits leaf length, leaf width and the ratio leaf width to length, in which the last trait is also investigated in this study. Two QTL regions were found accounted for 8.8%, 11.1% and 6.3%, 12.0% of the phenotypical variation of year 2008 and 2009, a major gene was estimated to be responsible for this trait. Considering in this study no significant QTL region was detected for trait leaf width/length ratio, the reasons could be as follows: 1). *Insufficient phenotyping data.* In this previous study, in total 142 F1 hybrids and 2 parental lines were phenotyped and 3 replicate plants were applied in each F1 hybrid, 5 replicate plants were applied for the two parental lines. Comparing to that, this study phenotyped 98 F1 hybrids and 2 parental lines with three replicate plants in each hybrid/parental line. To achieve more accurate result of QTL detection, more phenotyping data is suggested. 2). *The differences between using dominant markers and co-dominant markers.* The detected QTL regions were based on genetic map using SRAP markers, which are dominant. Given the fact that alleles with the same sequences in a locus would be visualized once, the results from SRAP marker-based genotyping would neglect different dosage constitution scenarios in polyploids. It has been illustrated that chrysanthemum is a hexaploidy with 6 alleles and 0 – 6 dosages could be found for each locus, different dosage would have impact on the phenotypes as well. It was more accurate to use co-dominant marker like SNP than

dominant markers for genotyping. 3). *The ratio of leaf width to length could be a very complex trait.* In fact this trait can be seemed as contributed by two sub-traits: leaf length and leaf width. A significant positive correlation between this two sub-traits was proved. 5 QTL regions were found for each of the two sub-traits and 2 of them were co-localised, which suggest there could be same genes regulating both sub-traits
(F. Zhang, J. Jiang, S. Chen, 2012). These results were based on a SRAP genotyping linkage map, whose disadvantages have been discussed before. Consequently, QTL analysis for traits leaf width and leaf length is suggested further based on the integrated linkage map constructed in this study.

In this paper, QTL analysis were focused on four leaf traits. As shown in Figure 7, the correlation coefficient between trait Leaf number and Total stipule size is 13.48%, which was higher than the other trait combinations. One possible explanation is that in every shoot of chrysanthemum, generally each leaf is growing with two stipules on two sides. More number of leaves are grown indicates more vigor on the stem. Hence larger areas of stipules can be achieved.

In this study, two models (marker dosage model and genotype probability model) were combined to seek candidate markers associating with the detected QTL regions on chrysanthemum leaf trait stem length. Marker dosage model is a linear model associating different marker dosages with phenotypic values. Analysis of Variance (ANOVA) is performed to check the single marker dosage regression on phenotypic values. This is a very basic approach and treat the effects of marker dosages to the phenotypic values as purely additive. No dominant effects are considered (P. Bourke, 2014). Hence the biggest drawbacks of this model are: 1). *It doesn't take the dosage phasing scenarios into account.* Only the dosage number would be considered. 2). *Markers can only be detected significant if they're tagging the homologues with the same directions of effect.* Although some markers are tagging significant regions of homologues, however they are tagging homologues with opposite directions, would still be regarded as non-significant. One possible modified solution is using two markers tagging specific homologues whose effects on phenotypes are known. By combining the dosage information from two markers, the explained phenotypic variation is expected to be higher than with single marker dosage. In this case, the improvement was not very much (R-square: 12.72% > 12.47%). This is due to the complex dosage combinations of the two markers (DxS & SxD), maximumly 4 x 4 = 16 possible dosage combinations could be created. Also, only 3 out of 6 homologues which the two markers are tagging are associated with the QTL regions, too many irrelevant tagged homologues lowered down the explanation for phenotypes. The most ideal scenario is using markers with low dosages which are tagging all the homologues associating with the QTL regions, then the phenotypic variation explained by this model is expected to be higher.

Compared to Marker dosage model, IBD model is superior. The main reason is it checks the effect of each allele. The contributions made by each of the twelve alleles from the two parents are taken into account for each locus in chrysanthemum.

4.3 What's next for breeders

This paper presents candidate markers proved to be linked to the QTL regions for trait stem length. For further application of these markers into production, they need to be validated and tested in a new population. Single marker assays can be made and used to genotype the new population. One possible way is to use one of the two parents of the population to cross with a new variety. This parent can contribute the donor QTL and the offspring will inherit the QTL regions for stem length with higher chance. Hence co-segregation between the marker and the QTL regions can be checked.
It might also happen that under new crosses in different population, there are other alleles totally overruling the alleles which the markers are tagging. Hence epistatic QTL analysis is suggested.

Breeders may concern most about such questions: 'What specific markers can I use in different populations?', 'How to grow the tallest/shortest plant?'. To answer these questions, we need to check whether the alleles tagging QTL regions can be inherited to the next generation or are still working in other genetic background. In this case, combined marker analysis based on the marker dosage model could give a hint on estimating the phenotypes by selecting plants with markers which are tagging specific homologues. For example, for marker combination 'Cm38566_c0g1i1_1004_Rnn' and 'Cm13465_c0g2i3_318_Sn', in this study, 'Cm38566_c0g1i1_1004_Rnn' with dosage 2 (tagging homologue 5 and 8) and 'Cm13465_c0g2i3_318_Sn' with dosage 0 is proved to associate with the shortest stem length and 'Cm38566_c0g1i1_1004_Rnn' with dosage and 'Cm13465_c0g2i3_318_Sn' with dosage 4 (tagging homologue 4) is proved to associate with the longest stem length.

4.4 Genome and genes

In the long run, the chrysanthemum genome is foreseen to be achieved. However, the linkage map will still maintain its usage. Plant breeders are more interested about the genetic distance than the physical distance. They pay more attention to the recombination frequencies on the chromosomes in order to make better combinations of good traits in the offspring. Hence keep digging the information hidden from the linkage map is always of great use in breeding.

In this study, a list of candidate genes is provided to make estimations on the genes which are responsible for chrysanthemum growth. These genes were acquired by aligning the sequences of contigs on the detected QTL regions with the UniProt database. Given the fact that the larger the bit-score, the

better quality that alignments have and the choice of bit-score depends on the different results, the alignment result was filtered by bit-score 100, which sieved half of the data (21231/46946 genes).

5. Conclusion

In this study, an integrated linkage map was constructed for chrysanthemum by using genotyping data from a 183k SNP array. With the help of this linkage map, QTL analysis was performed on chrysanthemum leaf traits by combining two models: Marker dosage model and Genotype probability model. Two QTL regions were detected on linkage group 4 and 7 on leaf trait stem length. Candidate markers were listed and proved to be associated with the QTL regions. Also, a possible modification approach was proposed to modify Marker dosage model by combing dosage information of two markers which are tagging specific homologues. A list of candidate genes was shown to give reference for estimating genes which are responsible to regulate chrysanthemum stem length.

Reference:

Bourke, P. (2014). QTL analysis in polyploids QTL analysis in polyploids: Model testing and power calculations, 1–67.

Bourke, P. M., Voorrips, R. E., Kranenburg, T., Jansen, J., Visser, R. G. F., & Maliepaard, C. (2016). Integrating haplotypes-specific linkage maps in tetraploid species using SNP markers.

Chang, K. Y., Lo, H. F., Lai, Y. C., Yao, P. J., Lin, K. H., & Hwang, S. Y. (2009). Identification of quantitative trait loci associated with yield-related traits in sweet potato (Ipomoea batatas). *Botanical Studies*, *50*(1), 43–55.

De Backer, M. (2012). *Characterization and detection of* Puccinia horiana *on chrysanthemum for resistance breeding and sustainable control*.

De Jong, J., & Rademaker, W. (1986). The reaction of Chrysanthemum cultivars to Puccinia horiana and the inheritance of resistance. *Euphytica*, *35*(3), 945–952. https://doi.org/10.1007/BF00028604

F. Zhang, J. Jiang, S. Chen, F. C. & W. F. (2012). Detection of quantitative trait loci for leaf traits in chrysanthemum. *Genetics, Selection, Evolution : GSE*, *87*(6), 613–618. Retrieved from http://dx.doi.org/10.1080/14620316.2012.11512920

Hackett, C. A., McLean, K., & Bryan, G. J. (2013). Linkage Analysis and QTL Mapping Using SNP Dosage Data in a Tetraploid Potato Mapping Population. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0063939

Hedden, P., & Sponsel, V. (2015). A Century of Gibberellin Research. *Journal of Plant Growth Regulation*, *34*(4), 740–760. https://doi.org/10.1007/s00344-015-9546-1

Klie, M., Schie, S., Linde, M., & Debener, T. (2014). The type of ploidy of chrysanthemum is not black or white: a comparison of a molecular approach to published cytological methods. *Frontiers in Plant Science*, *5*(September), 1–8. https://doi.org/10.3389/fpls.2014.00479

Langton, F. A. (1989). Inheritance in Chrysanthemum morifolium Ramat. *Heredity*, *62*(3), 419–423. https://doi.org/10.1038/hdy.1989.57

Li, X., Han, Y., Wei, Y., Acharya, A., Farmer, A. D., Ho, J., … Brummer, E. C. (2014). Development of an alfalfa SNP array and its use to evaluate patterns of population structure and linkage disequilibrium. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0084329

Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3), 69–71. https://doi.org/10.1016/j.cmpb.2016.01.020

Park, S. K., Arens, P., Esselink, D., Lim, J. H., & Shin, H. K. (2015). Analysis of inheritance mode in chrysanthemum using EST-derived SSR markers. *Scientia Horticulturae*, *192*, 80–88. https://doi.org/10.1016/j.scienta.2015.05.009

Preedy, K. F., & Hackett, C. A. (2016). A rapid marker ordering approach for high-density genetic linkage maps in experimental autotetraploid populations using multidimensional scaling. *Theoretical and Applied Genetics*, *129*(11), 2117–2132. https://doi.org/10.1007/s00122-016-2761-8

Schmitz Carley, C. A., Coombs, J. J., Douches, D. S., Bethke, P. C., Palta, J. P., Novy, R. G., & Endelman, J. B. (2017). Automated tetraploid genotype calling by hierarchical clustering. *Theoretical and Applied Genetics*, *130*(4), 717–726. https://doi.org/10.1007/s00122-016-2845-5

Serang, O., Mollinari, M., & Garcia, A. A. F. (2012). Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. *PLoS ONE*, *7*(2), 1–13. https://doi.org/10.1371/journal.pone.0030906

Smulders, M. J. M., Voorrips, R. E., Esselink, G. D., Santos Leonardo, T. M., Vant' Westende, W. P. C., Vukosavljev, M., … Maliepaard, C. (2015). Development of the WagRhSNP AXIOM SNP array based on sequences from tetraploid cut roses and garden roses. *Acta Horticulturae*, *1064*, 177–184.

Uitdewilligen, J. G. A. M. L., Wolters, A. M. A., D'hoop, B. B., Borm, T. J. A., Visser, R. G. F., & van Eck, H. J. (2013). A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. *PLoS ONE*, *8*(5), 10–14. https://doi.org/10.1371/journal.pone.0062355

van Geest, G., Bourke, P. M., Voorrips, R. E., Marasek-Ciolakowska, A., Liao, Y., Post, A., … Arens, P. (2017a). An ultra-dense integrated linkage map for hexaploid chrysanthemum enables multi-allelic QTL analysis. *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-017-2974-5

van Geest, G., Bourke, P. M., Voorrips, R. E., Marasek-Ciolakowska, A., Liao, Y., Post, A., … Arens, P. (2017b). An ultra-dense integrated linkage map for hexaploid chrysanthemum enables multi-allelic QTL analysis. *Theoretical and Applied Genetics*, *130*(12), 2527–2541. https://doi.org/10.1007/s00122-017-2974-5

van Geest, G., Voorrips, R. E., Esselink, D., Post, A., Visser, R. G. F., & Arens, P. (2017). Conclusive evidence for hexasomic inheritance in chrysanthemum based on analysis of a 183 k SNP array. *BMC Genomics*, *18*(1), 1–12. https://doi.org/10.1186/s12864-017-4003-0

Voorrips, R. E., Gort, G., & Vosman, B. (2011). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics*, *12*. https://doi.org/10.1186/1471-2105-12-172

Vos, P. G., Uitdewilligen, J. G. A. M. L., Voorrips, R. E., Visser, R. G. F., & van Eck, H. J. (2015). Development and analysis of a 20K SNP array for potato (Solanum tuberosum): an insight into the breeding history. *Theoretical and Applied Genetics*, *128*(12), 2387–2401. https://doi.org/10.1007/s00122-015-2593-y

Zhang, F., Chen, S., Chen, F., Fang, W., Chen, Y., & Li, F. (2011). SRAP-based mapping and QTL detection for inflorescence-related traits in chrysanthemum (Dendranthema

morifolium). *Molecular Breeding*, *27*(1), 11–23. https://doi.org/10.1007/s11032-010-9409-1

Zhang, F., Chen, S., Chen, F., Fang, W., Deng, Y., Chang, Q., & Liu, P. (2011). Genetic analysis and associated SRAP markers for flowering traits of chrysanthemum (Chrysanthemum morifolium). *Euphytica*, *177*(1), 15–24. https://doi.org/10.1007/s10681-010-0239-3

Zhang, F., Chen, S., Chen, F., Fang, W., & Li, F. (2010). A preliminary genetic linkage map of chrysanthemum (Chrysanthemum morifolium) cultivars using RAPD, ISSR and AFLP markers. *Scientia Horticulturae*, *125*(3), 422–428. https://doi.org/10.1016/j.scienta.2010.03.028

Zhang, F., Chen, S., Jiang, J., Guan, Z., Fang, W., & Chen, F. (2013). Genetic mapping of quantitative trait loci underlying flowering time in chrysanthemum (Chrysanthemum morifolium). *PLoS ONE*, *8*(12), 1–9. https://doi.org/10.1371/journal.pone.0083023

Zhang, F., Jiang, J., Chen, S., Chen, F., & Fang, W. (2012). Mapping single-locus and epistatic quantitative trait loci for plant architectural traits in chrysanthemum. *Molecular Breeding*, *30*(2), 1027–1036. https://doi.org/10.1007/s11032-011-9686-3

Appendix

Appendix 1. Situations of marker dosage conversion

Marker segregation types before conversion	Marker segregation types after conversion
0x1 0x5 6x1 6x5	0x1
0x2 0x4 6x2 6x4	0x2
0x3 6x3	0x3
1x0 1x6 5x0 5x6	1x0
1x1 5x5	1x1
1x2 5x4	1x2
1x3 5x3	1x3
1x4 5x2	1x4
1x5 5x1	1x5
2x0 4x0 2x6 4x6	2x0
2x1 4x5	2x1
2x2 4x4	2x2
2x3 4x3	2x3
2x4 4x2	2x4
3x0 3x6	3x0
3x1 3x5	3x1
3x2 3x4	3x2
3x3	3x3?
2x5 4x1	4x1

Appendix 2. Top 20 sig markers whose dosages have significant influence on four leaf traits.

Stem length	Marker name	P-value	Leaf number	Marker name	P-value
	Cm63076_c0g1i1_327_Sn	4.952076e-05		Cm48278_c0g1i1_408_Ss	9.143220e-06
	Cm61216_c0g1i1_831_Sn	7.799028e-05		Cm51968_c0g3i1_247_Sn	4.413587e-05
	Cm28578_c0g1i2_1870_Sn	9.087962e-05		Cm2410_c0g2i1_305_Ss	8.229479e-05
	Cm54019_c0g1i1_240_Sn	9.563958e-05		Cm30361_c0g1i3_73_Sn	1.013483e-04
	Cm38566_c0g1i1_1004_Rnn	1.215021e-04		Cm26300_c0g1i1_1419_Sn	1.090006e-04
	Cm54107_c0g1i1_454_Rnn	1.225519e-04		Cm15004_c0g2i1_1461_Sn	1.157300e-04
	Cm4088_c0g1i3_841_Sn	1.242714e-04		Cm33457_c0g1i2_883_Qn	1.199550e-04
	Cm4088_c0g1i3_839_Sn	1.315800e-04		Cm10224_c0g2i1_405_Qn	1.441883e-04
	Cm36765_c0g1i2_1378_Sn	1.318080e-04		Cm62443_c0g4i1_1231_Pn	1.471396e-04
	Cm42659_c0g4i1_1038_Sn	1.325424e-04		Cm61282_c0g2i1_168_Sn	1.676177e-04
	Cm5596_c0g4i1_1231_Sn	1.325424e-04		Cm19252_c0g1i1_1364_Sn	1.734643e-04
	Cm10191_c0g4i1_174_Sn	1.325424e-04		Cm29284_c0g1i1_346_Sn	1.800957e-04
	Cm4088_c0g1i3_244_Sn*	1.325424e-04		Cm32716_c0g1i1_147_Sn	1.821049e-04
	Cm19355_c2g11i1_300_Sn	1.325424e-04		Cm37522_c0g4i1_844_Pn	2.116589e-04
	Cm19355_c2g11i2_218_Sn	1.325424e-04		Cm40144_c0g5i2_2101_Pn	2.170747e-04
	Cm5596_c0g4i1_1338_Sn	1.612022e-04		Cm3047_c0g1i1_75_Sn	2.414043e-04
	Cm55355_c1g4i2_244_Pn	1.698838e-04		Cm9155_c0g5i1_209_Qn	2.685209e-04
	Cm13713_c0g2i1_1456_Sn	1.753210e-04		Cm42234_c0g1i5_430_Sn	2.792401e-04
	Cm50757_c0g1i1_843_Sn	1.910821e-04		Cm21605_c0g2i1_434_Sn	2.793353e-04
	Cm30384_c0g2i1_280_Sn	1.916439e-04		Cm23784_c0g1i1_173_Sn	2.893092e-04
Total stipule size	**Marker name**	**P-value**	**Leaf width/length ratio**	**Marker name**	**P-value**
	Cm25292_c1g1i1_1467_Sn	2.233742e-06		Cm35770_c0g1i2_1493_Sn	5.056932e-06
	Cm62451_c0g1i1_649_Qn	9.679420e-06		Cm58661_c1g5i4_3563_Sn	1.046911e-05
	Cm23072_c0g4i1_745_Sn	4.001484e-05		Cm15477_c0g1i3_487_Sn	1.384440e-05
	Cm32716_c0g1i1_147_Sn	5.526876e-05		Cm35770_c0g1i2_3114_Sn	1.608983e-05
	Cm26300_c0g1i1_1419_Sn	9.723630e-05		Cm14409_c1g1i1_508_Sn	1.821006e-05
	Cm37674_c0g1i1_855_Pn	9.854494e-05		Cm35770_c0g1i2_2369_Rsn	1.918480e-05
	Cm22452_c0g2i1_505_Qs	1.174850e-04		Cm14409_c1g1i1_1540_Pn	1.930953e-05
	Cm46531_c0g6i2_577_Pn	1.655261e-04		Cm6667_c0g1i1_548_Sn	2.113084e-05
	Cm16543_c1g2i1_922_Sn	1.735090e-04		Cm30957_c0g1i1_997_Sn	2.244890e-05
	Cm38611_c0g1i2_921_Sn	2.155897e-04		Cm35770_c0g1i2_3822_Sn	2.598150e-05
	Cm50921_c0g1i1_1450_Sn	2.628442e-04		Cm58661_c1g2i1_921_Sn	2.688768e-05
	Cm55902_c0g1i1_1737_Sn	3.305492e-04		Cm35770_c0g1i2_1812_Pn	3.191721e-05
	Cm16726_c0g1i1_220_Sn	3.871894e-0		Cm35770_c0g1i2_1686_Qn	4.461112e-05
	Cm17954_c0g1i15_1606_Sn	4.035611e-04		Cm6667_c0g1i1_1511_Rnn	4.603984e-05
	Cm12991_c0g4i1_341_Sn	4.048498e-04		Cm35770_c0g1i2_2187_Sn	4.922717e-05
	Cm16930_c0g1i1_642_Qn	4.307635e-04		Cm25688_c0g1i1_263_Sn	6.852002e-05
	Cm28609_c0g1i1_614_Qn	4.628708e-04		Cm47666_c0g2i1_109_Sn	7.403321e-05
	Cm45927_c0g5i4_285_Sn	4.644224e-04		Cm57675_c2g3i1_124_Sn	8.396041e-05
	Cm44117_c0g1i1_247_Sn	4.669548e-04		Cm54766_c1g2i1_1064_Sn	8.442822e-05
	Cm40779_c0g1i1_414_Sn	4.964817e-04		Cm15098_c0g9i1_675_Sn	8.660405e-05

Appendix 3. Potential reference genes for QTL regions of chrysanthemum responsible for stem length.

Gene names	Organism
CYP701A8 KO1 Os06g0569500 LOC_Os06g37300 OSJNBa0062E01.27	Oryza sativa subsp. japonica (Rice)
GA3OX3 At4g21690 F17L22.150	Arabidopsis thaliana (Mouse-ear cress)
GA20OX1 20ox1 At2301 GA5 At4g25420 T30C3.90	Arabidopsis thaliana (Mouse-ear cress)
GA20OX5 At1g44090 F9C16.33 T7O23.20	Arabidopsis thaliana (Mouse-ear cress)
GA20OX2 20ox At2353 At5g51810 MIO24.5	Arabidopsis thaliana (Mouse-ear cress)
GA2OX2 At1g30040 T1P2.6	Arabidopsis thaliana (Mouse-ear cress)
GA2OX2	Pisum sativum (Garden pea)
GA2OX6 At1g02400 T6A9.9	Arabidopsis thaliana (Mouse-ear cress)
20ox1 GA20ox-1 Os03g0856700 LOC_Os03g63970 OSJNBa0059G06.22	Oryza sativa subsp. japonica (Rice)
GA20OX3 20ox3 YAP169 At5g07200 T28J14_140	Arabidopsis thaliana (Mouse-ear cress)
GA20OX4 At1g60980 T7P1.12	Arabidopsis thaliana (Mouse-ear cress)
GA2OX4 At1g47990 T2J15.10	Arabidopsis thaliana (Mouse-ear cress)
GA2OX1 SLN	Pisum sativum (Garden pea)
GA2OX3 At2g34555 T31E10.11	Arabidopsis thaliana (Mouse-ear cress)
GA2OX7 At1g50960 F8A12.18	Arabidopsis thaliana (Mouse-ear cress)
GA3OX4 At1g80330 F5I6.8	Arabidopsis thaliana (Mouse-ear cress)
GA2OX1 At1g78440 F3F9.5	Arabidopsis thaliana (Mouse-ear cress)
GA2OX1	Phaseolus coccineus (Scarlet runner bean) (Phaseolus multiflorus)
GA3OX1 GA4 At1g15550 T16N11.6	Arabidopsis thaliana (Mouse-ear cress)
GA2OX8 At4g21200 F7J7.140	Arabidopsis thaliana (Mouse-ear cress)
GA3OX2 GA4H At1g80340 F5I6.9	Arabidopsis thaliana (Mouse-ear cress)
LE	Pisum sativum (Garden pea)
KO CYP701A3 GA3 KO1 At5g25900 T1N24.23	Arabidopsis thaliana (Mouse-ear cress)
CYP714D1 EUI1 Os05g0482400 LOC_Os05g40384 OsJ_18961 OSJNBa0095J22.13	Oryza sativa subsp. japonica (Rice)
CTR1 At5g03730 F17C15_150	Arabidopsis thaliana (Mouse-ear cress)
CYP701A9 Os06g0568600 LOC_Os06g37224 OSJNBa0062E01.13	Oryza sativa subsp. japonica (Rice)
CYP701A6 D35 Os06g0570100 LOC_Os06g37364 OSJNBa0062E01.38	Oryza sativa subsp. japonica (Rice)
CYP701A19 Os06g0569900 LOC_Os06g37330 OSJNBa0062E01.34	Oryza sativa subsp. japonica (Rice)
WRKY71 OsI_06106	Oryza sativa subsp. indica (Rice)