DISENTANGLING HEXAPLOID GENETICS

Towards DNA-informed breeding
for postharvest performance in
chrysanthemum

Geert van Geest

Propositions

1. Cultivated chrysanthemum should be reclassified as autopolyploid.
   (this thesis)

2. Many deteriorative processes occurring after harvest of plant produce should not be classified as senescence.
   (this thesis)

3. The biggest threat of genome modifying techniques is a second diversity bottleneck in agriculture.

4. Convergence of subcultures catalyses the invention of new ways to express identity.

5. Preconception on gender-related behaviour is the largest threat to gender inequality.

6. Only in nations that are tolerant towards other cultures and discourage ethnocentrism there can be world-leading scientific progress.

7. The competitive advantages of not sharing data or knowledge are highly overestimated.

Propositions belonging to the thesis, entitled:

"Disentangling hexaploid genetics: Towards DNA-informed breeding for postharvest performance in chrysanthemum"

Geert van Geest
Wageningen, 15 September 2017

Disentangling hexaploid genetics

Towards DNA-informed breeding for postharvest performance in chrysanthemum

Geert van Geest

Disentangling hexaploid genetics

Towards DNA-informed breeding for postharvest performance in chrysanthemum

Geert van Geest

Table of contents

Chapter 1

General introduction

The first stock market crash in history was a herald for the beginning of a flourishing trade in ornamental plants and their diversity. This crash took place in the 1630s in the Netherlands and is commonly known as 'tulipomania'. It was caused by highly speculative trade in tulip bulbs. At its peak, bulbs were sold for high prices even though their flower traits had not yet been characterized. When these traits turned out disappointing, resell prices dropped far below the value of futures and the trade crashed. Since the start of tulip bulb trading in the 1600s, breeding and trade of ornamental plants has been a major economic activity. What used to be a domain of mainly elite hobbyists for centuries, grew out to a billion-euro business bringing colour to the homes of millions. Today, the ornamental industry plays a substantial part in the global economy and is driven by innovation.

Many ornamentals are clonally propagated. This means that plant material is propagated by division of plant parts and not seeds, which results in exact clones of the mother plant. In the top 10 cut flowers sold at the main flower auction in the Netherlands, Flora Holland (Flora Holland, 2015), only one (Eustoma) is propagated through seeds. The nine others, including chrysanthemum, are propagated through cuttings, bulbs, or tubers (Table 1.1). In ornamentals, polyploidy and clonal propagation often go hand in hand (Table 1.1). This also applies to food crops: most major clonally propagated food crops are also polyploid (e.g. potato, sweet potato, yam and plantain). Polyploidy can be very useful when it comes to generating phenotypic diversity, but the study of inheritance becomes severely more complicated.

Box 1. A short history of cultivated chrysanthemum

Chrysanthemum has a rich and long history of cultivation in Asia. The Chinese poems written by Tao Yuanming, who lived from 365 to 427, are the earliest written accounts of chrysanthemums as an ornamental plant (Spaargaren, 2015). Only while the Age of Discovery was at its full pace, in 1789 the first chrysanthemum plants rooted in European soil. Pierre-Louis Blancard brought chrysanthemums with him to France from China. More than 50 years later, in 1846, Robert Fortune brought cut flower type accessions from Japan, kick-starting the breeding for chrysanthemums as cut flower in Europe.

Breeding of ornamentals aims to combine the best traits in a genotype with high ornamental value. However, ornamentals are only valuable if they can reach the homes of consumers. Therefore, in a successful cultivar, ornamental value should be combined with at least an acceptable postharvest performance. Further improving postharvest performance can result in a profoundly more sustainable distribution of flowers. This is because products that are less perishable make logistic systems more flexible, and harvested flowers therefore have a smaller chance to end up as waste. Therefore, postharvest performance is of major interest for ornamental breeding. However, measuring performance during the postharvest phase is complex. Postharvest performance is affected by quality parameters that are often difficult to quantify. In addition, there is generally a large effect of the growth environment on the phenotype and often various postharvest problems play a role (Figure 1.1).

Table 1.1. Top 10 of cut-flowers according to turnover that passed the Flora Holland auction in 2015 (Flora Holland, 2015).

Cut flower	Scientific name	Turnover (million euro)	Main mode of propagation	Ploidy level of commercial cultivars
Rose	*Rosa x hybrida*	735	Grafting	4x
Chrysanthemum (spray)	*Chrysanthemum x morifolium*	300	Cuttings	6x
Tulip	*Tulipa* species and hybrids	295	Bulbs	2x-4x*
Lilium	*Lilium* species and hybrids	159	Bulbs	2x-4x†
Gerbera	*Gerbera x hybrida*	138	Tissue culture	2x
Chrysanthemum (disbud)	*Chrysanthemum x morifolium*	63	Cuttings	6x
Cymbidium	*Cymbidium* species	58	Plant division	2x-4x
Eustoma	*Eustoma grandiflorum*	52	Seeds	2x
Freesia	*Freesia refracta*	52	Bulbs	4x
Hydrangea	*Hydrangea macrophylla*	45	Cuttings	2x

*Most cultivars are diploid
†Most cultivars are triploid

DNA-informed breeding (Peace, 2017) can greatly improve breeding efficiency of traits that are difficult to measure. It is based on the co-segregation of DNA-markers and traits. Many important traits like postharvest performance cannot be assessed at seedling stage, and can only be measured after multiple rounds of propagation. By using a DNA-marker that is linked to a trait, these traits can nevertheless be assessed early in the breeding program. To discover such marker-trait associations, detailed knowledge on the inheritance of markers and traits is required. However, complex inheritance patterns in polyploids hamper the identification and localization of associated DNA-markers.

Of the top 10 ornamentals (Table 1.1), chrysanthemum is amongst the genetically most complex. It is a hexaploid that usually carries 54 chromosomes ($2n = 6x = 54$; Dowrick, 1953), it is strictly outcrossing and the mode of inheritance has been under discussion (Klie et al., 2014). The research area of chrysanthemum genetics is slowly becoming more elaborated with the tools developed in the recent era of DNA-based genetic analysis. A first genetic linkage map of DNA markers has been published (Zhang et al., 2010), and quantitative trait loci (QTL) were characterized (Klie et al., 2016; Zhang et al., 2013, 2011a). However, due to its genetic complexity, progress in chrysanthemum is slow compared to other ornamentals like rose (Debener and Mattiesch, 1999; Vukosavljev et al., 2016), lilium (Abe et al., 2002;

Shahin et al., 2011) and tulip (Tang et al., 2015). This lag is mainly caused by absence of high-quality datasets, knowledge on the mode of inheritance and methods that can work with hexaploidy. The generation of large genotyping datasets and understanding inheritance in hexaploid chrysanthemum enables the development of such methods. These are essential steps for the way towards DNA-informed breeding.

Figure 1.1. Different symptoms of a reduced postharvest quality in chrysanthemum: leaf yellowing and browning (A), disk floret degreening (B), flower wilting (C) and micro-organism infection of the vase water (D).

1.1 Phenotyping for postharvest performance

Postharvest performance is a major selection criterion for ornamentals. Assessing it can be done by mimicking the practical treatments of harvest, storage and vase life. However, this can only be done late in the breeding program, it needs space and resources, and it needs skilled labour to judge plants consistently on their quality. Most importantly, postharvest performance within a genotype can be highly variable. Therefore, improving phenotyping could strongly improve selection efficiency and enable DNA-informed breeding for postharvest performance.

1.1.1 Complexity of postharvest performance

Postharvest performance involves many different disorders that cause a reduced quality (Figure 1.1), which makes it a complex trait. In chrysanthemum, symptoms of those disorders would be e.g. dehydration symptoms (Van Ieperen et al., 2001; Van Meeteren and Van Gelder, 1999), leaf yellowing (Doi et al., 2003; Narumi et al., 2005; Reyes-Arribas et al., 2001), flower senescence or wilting (Bartoli et al., 1997, 1995), disk floret degreening and micro-organism infection of vase water (Put, 1990). Traits like postharvest performance can be described as heterogeneous, which means it has multiple underlying traits (Thornton-Wells et al., 2006). Many traits that are important for crop improvement fall within this definition; other examples are yield, taste, and tolerance against extreme environments, like heat, cold, drought and salinity. Such traits have in common that they are often strongly affected by the environment, and their phenotypes are difficult to predict from genotypic information.

The strong environmental effect on heterogeneous traits results in a low heritability. A reason is that different sub-traits can respond differently to specific environments. An example is postharvest performance in chrysanthemum; the sub-trait rehydration ability during vase life is affected by water availability during growth (Van Meeteren et al., 2005), but also by cut-height of the flower stem (Van

Meeteren and Van Gelder, 1999), and even the source of vase water (Van Meeteren et al., 2000). This is only one example of one sub-trait of postharvest performance. If any of this range of factors is not controlled in an experiment, this can have a profound effect on the phenotype, and therefore a large part of the phenotypic variability will be affected by environmental variability.

Heterogeneous traits are often challenging to phenotype, and phenotypic values are difficult to predict from genotypic information. This is because different sub-traits are likely regulated by several genetic loci. A good example are the sub-traits of the heterogeneous trait taste. In fruits, taste composes of a range of chemical attributes like sweetness, sourness and aroma (Klee, 2010; Stevens, 1986). Identification of loci associated with these attributes as sub-traits separately can be successful, but the main trait taste is build up out of hundreds of loci (Causse et al., 2001; Fulton et al., 2002; Paris et al., 2008; Saliba-Colombani et al., 2001). Thus, for the main trait less phenotypic variation can be explained per locus, and many loci explaining a small amount of variation will not be detected. Breeding for the sub-traits separately on the other hand is likely to be more successful.

1.1.2   Improving selection for heterogeneous traits

In order to investigate inheritance of heterogeneous traits with low heritability, a first step would be to reduce phenotypic variance within genotypes. One way to do this, is to control each known environmental factor. However, in most cases only a few factors are known to affect the trait. In addition, breeding has in general the aim to develop varieties that perform well in a range of environments. By selection in a highly-regulated environment, the performance of developed varieties in slightly different environments will be ignored. A more logical strategy to investigate inheritance of heterogeneous traits is to break up the trait of interest in its multiple underlying traits, and investigate those individually, similar to the examples on fruit taste. As less different processes are involved, probably less environmental factors will play a role. Those sub-traits will also be regulated by fewer loci, and therefore likely inherit in a less complex manner.

In order to divide heterogeneous traits into measurable sub-traits, knowledge is needed on the underlying processes that shape the phenotype. If those are known, one can test for genotypic differences in those processes, and with that find important determinants for which genetic variation exists. In addition to the examples of taste, successful examples come from research on another heterogeneous trait: salt tolerance. Underlying processes that are important to be able to gain acceptable yields under saline conditions are related to homeostasis of sodium and potassium ions (Yeo et al., 1990). Selection on yield or plant growth under saline conditions is restricted by low heritability and little explained phenotypic variation by the genotype (Flowers, 2004). However, identification of loci associated with different sub-traits related to ion homeostasis, has been much more successful (Lin et al., 2004; Munns et al., 2012; Thomson et al., 2010). It is therefore important to understand a heterogeneous trait physiologically to dissect a heterogeneous trait and enable selection.

Like salt tolerance, postharvest performance in cut-flowers is heterogeneous. Therefore, selection should be on lower-level traits individually. Such strategies have been proven useful for breeding for postharvest performance in several cut-flower species. In gerbera, selection on specific vase-life determining symptoms was proposed to be more effective than selection for vase life as a single trait (Wernett et al., 1996). In rose and carnation, variation in ethylene production and response have been used to aid selection for postharvest performance (Macnish et al., 2010; Onozaki et al., 2001). In rose, stomatal functioning has been used as a sub-trait of leaf wilting for QTL-mapping (Carvalho et al., 2015). In chrysanthemum, dimensions of xylem vessels were proposed as potential selection criteria for vulnerability to dehydration during vase life (Nijsse, 2001). After weeks of storage however, disk floret degreening (Figure 1.1B) is the most frequently occurring vase life-ending symptom in the current Dutch breeding germplasm of single (daisy-type) chrysanthemums (Figure 1.2). It also occurs (although less frequently) after shorter storage periods, and similar symptoms are found in other flower types (not shown). It is therefore a promising sub-trait of postharvest performance. Understanding the

physiology and inheritance of disk floret degreeening can have a large impact on chrysanthemum postharvest performance as a whole.
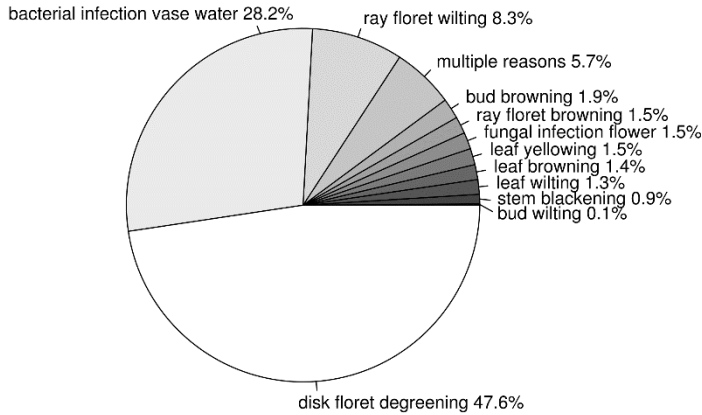
Figure 1.2. Pie chart of the distribution of the main symptom for ending vase life in single (daisy-type) flowers after three weeks of storage. Data is obtained from 3355 vase life tests performed at Deliflor chrysanten B.V for breeding purposes.

1.2   Towards DNA-informed breeding in chrysanthemum

In addition to precise phenotyping, DNA-informed breeding requires information on associations between DNA polymorphisms and the trait of interest. To reach this, three main steps need to be taken: discovery of polymorphic loci, localization of those loci in the form of a genetic linkage map and identification of loci associated with phenotypes of interest. The fact that chrysanthemum is an outcrossing and highly heterozygous hexaploid complicates polymorphism discovery and scoring, but also mapping of polymorphic loci and traits. In addition, analysis of inheritance is strongly influenced by the mode of inheritance; chromosomal pairing may occur only within sub-genomes or between all homologous chromosomes. In the following paragraphs, I describe the state-of-art, and what is needed to take the steps from DNA-polymorphism discovery to identification of genetic loci in hexaploid chrysanthemum in order to enable DNA-informed breeding.

Box 2. Definitions

Locus: A position on a chromosome (*i.e.* a set of homologous chromosomes).

Allele: A variant of a locus.

Polymorphism: Occurrence of two or more alleles at one locus (based on: Cavalli-Sforza and Bodmer, 1971).

(Genetic or DNA) marker: A single piece of DNA or a set of pieces of DNA that mark one or more specific alleles and can be detected through a single assay.

Heterozygosity: The state of an individual in which a locus carries at least two alleles.

Homozygosity: The state of an individual in which a locus carries only one single allele in the number of copies equal to the ploidy level.

Homologous chromosomes: chromosomes that can recombine and harbour the same loci in a generally same order of arrangement.

Polyploidy: a feature of organisms that have more than one pair of homologous chromosomes for each chromosome set.

Autopolyploidy: Polyploidy that originated from a whole genome multiplication within one species

Allopolyploidy: Polyploidy that originated from the combination of multiple sets of chromosomes from different species

Homoeologous chromosome: chromosomes that do not recombine but do harbour the same loci in generally the same order of arrangement. Homoelogous chromosomes usually originate from different species.

Polysomic inheritance: The mode of inheritance that is caused by random pairing during meiosis of homologues within a set. Usually (but not exclusively) a feature of autopolyploids.

Disomic inheritance: The mode of inheritance that is caused by preferential pairing during meiosis of homologue pairs within a set. Usually (but not exclusively) a feature of allopolyploids.

1.2.1 Polymorphism discovery and dosage scoring

In order to acquire a genome-wide representation in the form of genetic linkage map, each of the 54 chromosomes per individual should be saturated with markers for genetic polymorphisms. Currently, single-nucleotide polymorphism (SNP) assays are frequently used to characterize such polymorphisms. Very large numbers can be discovered with use of next generation sequencing approaches, and accurately assessing them is possible in a high throughput by the use of genotyping arrays (e.g. Koning-Boucoiran et al., 2015; Li et al., 2014a; Vos et al., 2015).

SNP loci are generally bi-allelic, meaning that two possible alleles are characterized per locus. Unlike diploids, polymorphisms in polyploids have a genomic dosage. Possible genotypes for one SNP in a diploid are AA, AB and BB. In speaking, these are homozygous A, heterozygous, and homozygous B. For a hexaploid, possible SNP genotypes are AAAAAA, AAAAAB, AAAABB, AAABBB, AABBBB, ABBBBB and BBBBBB. Considering the B-allele, those genotypes have dosages ranging from 0 to 6 respectively.

In speaking, they are addressed ranging from nulliplex B to hexaplex B (Table 1.2). These dosages are important to quantify for two main reasons. First, they have implications for segregation of an allele in the offspring. A SNP-allele for which one parent carries a dosage of 1 (simplex) and the other a dosage of 0 (nulliplex), later referred to as 1x0 marker, would be expected to segregate in a 1:1 ratio for the dosages 1 and 0. In a 2x0 situation, there would be a segregation of 1:3:1 (for the dosages 0, 1 and 2, assuming polysomic inheritance). Higher dosage combinations even have more complex segregation patters (The 3x3 situation would segregate as 1:18:99:164:99:18:1). Second, allele dosage can affect the phenotype. A duplex dosage of an allele might have a stronger phenotype than a simplex dosage (Hackett et al., 2014; Rosyara et al., 2016). It has therefore, next to linkage, important implications for identification and analysis of trait loci.

Table 1.2. Allele dosage scores in polyploids and their nomenclature. In this thesis, dosages are represented by numbers.

Genotype	Dosage	Spoken	Abbreviation
AAAAAA	0	Nulliplex	N
AAAAAB	1	Simplex	S
AAAABB	2	Duplex	D
AAABBB	3	Triplex	T
AABBBB	4	Quadruplex	Q
ABBBBB	5	Pentaplex	P
BBBBBB	6	Hexaplex	H

Assays to characterize SNP markers, like genotyping arrays or KASP assays, enable dosage scoring of detected SNPs. Genotyping arrays provide possibilities for massive throughput of SNP markers. Those assays are based on fluorescence of two probes; one for each SNP allele. The ratio between the fluorescence of these probes gives information on the allele dosage. The number of possible clusters correspond to the number of possible dosage scores. In a diploid, this number is three, in a tetraploid five, and in a hexaploid seven (Table 1.2; Figure 1.3). Separation of such clusters highly depends on the quality of the SNP and the assay. To build a high-throughput polymorphism detection platform, SNPs would need to be discovered from sequencing data, but this comes with several challenges. The first challenge is the large range of dosage scores, which complicates prediction of dosage out of sequence data. Second, in the case of chrysanthemum, there is no reference genome sequence, so SNP discovery is based on a de novo sequence assembly. Such an assembly from a hexaploid is most likely imperfect. Therefore, pseudo-assemblies of gene families, separated haplotype assemblies and chimeric assemblies will occur frequently and should be taken into account during SNP discovery. Lastly, similar to other polyploids like potato (Uitdewilligen et al., 2013), high SNP densities would be expected in chrysanthemum. SNPs neighbouring the SNP of interest can interfere with the assay, and should be dealt with during assay design. These three reasons indicate that successful SNP marker discovery and design of SNP assays in hexaploid chrysanthemum needs elaborate selection methods.
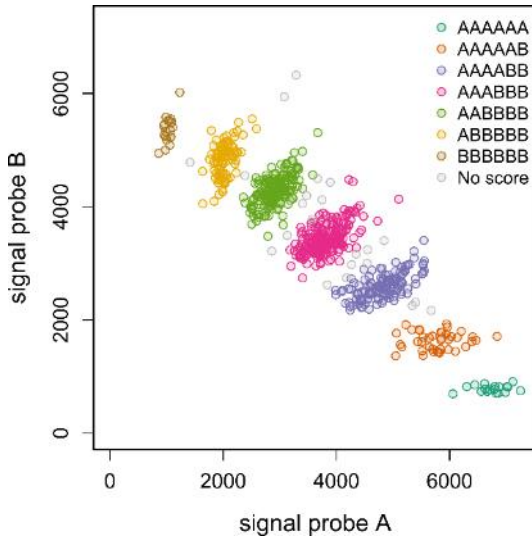
Figure 1.3. An example of output of a single SNP genotyping assay that could detect all seven different dosages in a hexaploid. Each dot represents an individual genotype. Dots with the same colour have the same dosage ranging from 0 to 6.

1.2.2   The mode of inheritance

A major issue for genetic analysis in polyploid crops is the mode of inheritance. Inheritance in polyploids comes in two modes: disomic and polysomic inheritance (Figure 1.4), but a mix of these is also possible (Allendorf and Danzmann, 1997; Bourke et al., 2017; Leal-Bertioli et al., 2015; Stebbins, 1947; Stift et al., 2008). The mode of inheritance has important implications for allelic segregation, recombination and with that; genetic analysis. In chrysanthemum, the mode of inheritance has been under discussion (Klie et al., 2014). Literature on this topic provides two types of analysis to draw conclusions on the type of inheritance: cytological imaging and segregation analysis.

Cytological imaging points to presence of polysomic inheritance. In polyploids, chromosomes can pair in bivalents, which is similar to diploids, or in multivalents, in which multiple homologous chromosomes pair in a single structure. With cytological imaging of metaphase I, the type of pairing of homologous chromosomes can be visualized. Presence of multivalents indicate polysomic inheritance. However, absence does not indicate disomic inheritance, because bivalents do not necessarily represent two specific chromosomes on every meiotic event, and can therefore result from random pairing of chromosomes (Stebbins, 1940). In chrysanthemum, multivalents are rare (Chen et al., 2009; Dowrick, 1953; Roxas et al., 1995). However, bivalent formation seems to be genetically regulated. Colchicine doubled diploid *Chrysanthemum boreale* pollen mother cells had much lower frequencies of multivalents than would be expected based on chance from these completely homologous chromosomes (Watanabe, 1983). Existence of genetic regulation of multivalent formation has been demonstrated in an autotetraploid wheat species (Avivi, 1976). The frequently observed bivalents therefore do not provide evidence for disomic inheritance, but existence of multivalents do indicate polysomic inheritance in at least some regions.
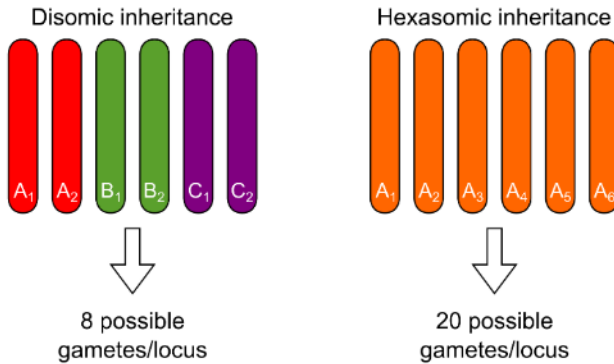
Figure 1.4. Graphical representation of disomic inheritance and hexasomic inheritance. Chromosomes represented by the same letter and colour can pair during meiosis I.

Conforming the cytological observations, studies on allelic segregation provide mainly evidence for polysomic inheritance in chrysanthemum. Conclusive evidence is based on segregation of flower colour (Langton, 1989), segregation of multi-allelic SSR-markers (Klie et al., 2014; Park et al., 2015) and absence of repulsion linkage between 1x0 markers (Klie et al., 2014). Despite the fact that these studies show clear hexasomic inheritance for the investigated loci, they are limited to these loci, and therefore do not exclude preferential pairing at other locations in the genome. Some indications for disomic segregation in chrysanthemum are based on segregation of multi-dose (dosage higher than one) dominant alleles (De Backer, 2012; De Jong and Rademaker, 1986; Klie et al., 2014). Markers that have a dosage of two (duplex) in one parent and zero (nulliplex) segregate in a 3:1 ratio with disomic inheritance and in a 4:1 ratio with hexasomic inheritance. There are reports of deviations from a 4:1 ratio towards a 3:1 ratio in some loci, indicating disomic inheritance (De Jong and Rademaker, 1986; Klie et al., 2014; Park et al., 2015). However, such analyses become inaccurate with small population sizes.

Most findings point to polysomic inheritance in chrysanthemum, but existence of disomic inheritance is not excluded. Remarkably, many recent publications treating this topic consider chrysanthemum as (segmental) disomic (Anderson, 2007; Klie et al., 2014; Li et al., 2011; Zhang et al., 2011a, 2014). To find conclusive evidence on the entire genome, a genome-wide approach is necessary. This will require the scoring of a large number of DNA polymorphisms on a large population in order to investigate expected segregation that fit disomic or hexasomic inheritance.

1.2.3   Linkage map construction

The information on allele dosage and the mode of inheritance enable the study of inheritance of alleles of polymorphic loci and their associated traits. Information on the dependence of alleles and trait segregation can be stored in a genetic linkage map, which provides an extensive and integrated representation of genetic linkage between alleles, loci and traits. The construction of a linkage map is based on estimating recombination frequency between alleles of adjacent polymorphic loci that are represented by genetic markers. Alleles of polymorphic loci that are close together on the same physical chromosome occur more frequently together in an offspring than would be expected from random segregation. A genotype in which the two alleles do not appear together, is called a recombinant. By calculating recombination frequency, one can estimate linkage between two polymorphic loci.

Because of the seven possible genomic dosages and recombination between six homologous chromosomes, genetic linkage analysis in polysomic hexaploids is more complex than for diploids. The study of hexaploid genetics is most advanced in sweet potato. Recent published maps are constructed with 1x0 and 1x1 markers, and identification of homologous chromosomes is based on linkages to

multi-dose markers (Cervantes-Flores et al., 2008; Zhao et al., 2013) or by aligning the homologue maps to a reference genome (Shirasawa et al., 2017). Fully integrating these homologue specific maps into linkage maps per chromosome is useful for further genetic analysis. First, it enables identification of trait loci that carry multiple alleles (Hackett et al., 2013). Second, linkage maps can be compared, which would otherwise not be possible because alleles are interchanged between homologues by crossing over. One way to integrate and order homologue maps, is to estimate linkage between simple segregating markers in repulsion. However, in hexaploids, this would require very large populations (n > 450; Wu et al., 1992). Another method to construct an integrated map is to estimate linkages between higher dose markers. These linkages can form bridges between homologous chromosomes and integrate the linkage map. However, such a method has not been applied in hexaploids before, but has been applied in tetraploids (Bourke et al., 2017; Hackett et al., 2013, 1998; Luo et al., 2001; McCallum et al., 2016). Hexaploids lag behind tetraploids mainly because high-quality datasets with a large number of multi-dose markers have not yet been described, and methods to analyse such data are not available.

1.2.4   QTL mapping

Analysis of quantitative trait loci (QTL) aims to locate regions on the genome that are linked to quantitative traits. For DNA-informed breeding, the locations of quantitative traits are used to identify associated DNA-polymorphisms. The results of a QTL analysis also shed light on the genetic architecture; it gives information on the number of loci involved, from which parent the alleles segregate and whether there are interactions between alleles. Currently, QTL analysis in hexaploids is limited to per-marker statistical tests or to the use of diploid models on simple segregating markers located on homologue specific maps (Chang et al., 2009; Shirasawa et al., 2017; Zhang et al., 2013, 2011a; Zhao et al., 2013). These analyses do often not represent the actual situation, as they only carry information on one allele at a time (Figure 1.5).

Estimation of presence or absence of each possible allele at a locus enables multi-allelic QTL analysis. Hackett et al. (2013) proposed a method to reconstruct the allelic combinations per locus for tetraploids using bi-allelic SNP markers. With this method, inheritance of each parental allele in each F1 individual can be inferred at each marker position (A situation approaching Figure 1.5B). For QTL analysis, a model is fitted over those genotype probabilities to test for explained variance of a phenotype. Recently, a similar but more elaborate and robust method has been published (Zheng et al., 2016). Bourke (2014) describes a simpler method to directly estimate these haplotype probabilities based on estimated recombination frequency. This method is ploidy independent and could therefore be applied to hexaploids. The improvement of QTL analysis in tetraploids (Hackett et al., 2014) indicates potential of the use of haplotype probabilities in polyploids. However, such methods have not yet been applied to hexaploids with polysomic inheritance.
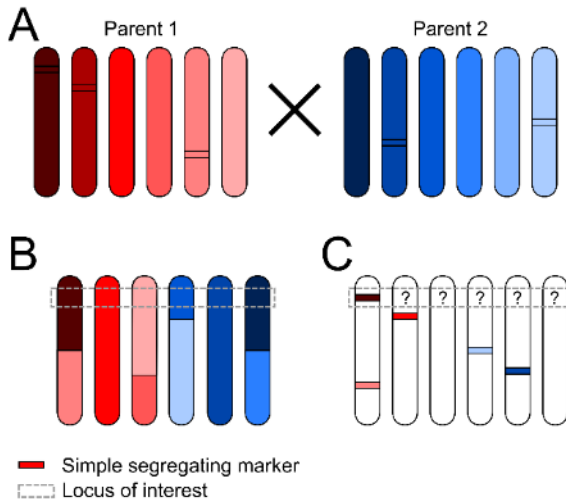
Figure 1.5. Example of a cross of two parents with each six different alleles (A), and an example of an F1 progeny (B), and the information on the segregation based on simple segregating markers (C). Considering single simple segregating markers, one can only characterize absence or presence in the F1 offspring. However, the actual situation is much more complex. On that same locus there are five other alleles that all could have specific effect on the phenotype. Genotype probability models try to reconstruct the actual situation by estimating absence and presence for each allele per locus that is approaching (B).

1.3   Scope of this thesis

The development of tools that enable DNA-informed breeding in hexaploid chrysanthemum comes with significant hurdles, among which SNP discovery, the unclear mode of inheritance, construction of an integrated linkage map and reconstruction of parental haplotypes in the progeny. One of the most important traits in chrysanthemum, postharvest performance, is a heterogeneous trait that needs to be broken down into sub-traits before DNA-informed breeding can be used to improve it. To fill these knowledge gaps, this thesis reaches two main objectives:

1. Understanding and dissecting some important characteristics of postharvest performance in order to develop precise phenotyping methods.
2. Understanding inheritance and performing genetic analysis in hexaploid chrysanthemum in order to pave the way towards DNA-informed breeding.

These two objectives are addressed in the chapters that follow the introduction. Chapter 2 and 3 describe the first objective of postharvest performance, chapter 4 and 5 the second objective on genetic analysis. In chapter 6 the two objectives come together in amongst others a QTL analysis for postharvest performance. A more detailed outline can be found below:

-   **Chapter 2**: Disk floret degreening is an important sub-trait of postharvest performance for long-stored chrysanthemum flowers. To understand the physiological processes underlying genotypic differences in disk floret degreening, in this chapter we investigate the metabolome of disk florets in different states of carbohydrate starvation and degreening. This shows that disk floret degreening is related to carbohydrate starvation, but complex interactions between carbohydrate content and cultivar-specific responses define genotypic differences.
-   **Chapter 3**: To genetically dissect the trait postharvest performance and its sub-trait disk floret degreening, we investigate inheritance of disk floret degreening and carbohydrate content in a large bi-parental population. We phenotype disk floret degreening with a method that uses detached capitula, which is cheap, fast and is highly correlated with disk floret degreening as observed during vase life after long storage of the flowers.

- **Chapter 4**: To perform genetic analysis in chrysanthemum, one needs to discover and assay DNA polymorphisms. In this chapter, we describe the development of a high-throughput genotyping platform for hexaploid chrysanthemum. The genotyping platform is validated in a genotype panel and multiple bi-parental populations. By analysis of the resulting genotypic data we deliver compelling evidence for hexasomic inheritance in chrysanthemum.
- **Chapter 5**: Software to perform genetic linkage analysis in hexaploids with polysomic inheritance was not available at the start of this thesis. We developed methods to perform such analyses. In this chapter, we present a software package that provides every step from SNP marker dosage data to an integrated genetic linkage map.
- **Chapter 6**: All aspects of the thesis come together in this chapter. We describe the construction of an integrated genetic linkage map from the data described in chapter 4. The map and phenotypic data described in chapter 3 is used to perform multi-allelic QTL analyses on four traits, among which disk floret degreening.
- **Chapter 7**: The general discussion is used to evaluate the obtained progress described in the experimental chapters in the broader context of breeding for complex traits in ornamentals and polyploid crops.

Chapter 2

Genotypic differences in metabolomic changes during storage induced-degreening of chrysanthemum disk florets

Geert van Geest, Young Hae Choi, Paul Arens, Aike Post, Ying Liu, Uulke van Meeteren

Abstract

Selecting chrysanthemum cultivars with long storability and vase life is a major challenge for breeders. The rate of degreening of disk florets during the postharvest phase is an important determinant of vase life. There is large genotypic variation in susceptibility to disk floret degreening. Our aim was to understand these genotypic differences at the physiological level. Carbohydrate starvation seemed to play a role, since application of sugars prevented degreening and degreening only occurred if florets had a long-term low carbohydrate content. In order to find out which metabolic processes could explain genotypic differences, we used [1]H Nuclear Magnetic Resonance (NMR) spectroscopy profiling, High Performance Anion Exchange Chromatography (HPAEC) and respiration measurements to compare metabolic responses of three genotypes to carbohydrate starvation. HPAEC and NMR measurements showed that carbohydrate content could not fully explain genotypic differences. A genotype with intermediate sensitivity to degreening showed similar carbohydrate content compared to an insensitive one. However, respiration rate declined faster under carbohydrate starvation in a sensitive and intermediate sensitive genotype compared to an insensitive genotype, suggesting a more abrupt constraint on the mitochondrial electron transport chain and with that oxidative stress. Changes in the metabolic profile under carbohydrate starvation were diverse and revealed candidate processes associated with disk floret degreening. Camphor content increased and correlated positively with degreening insensitivity. Phenylpropanoids and flavonoids also increased upon carbohydrate starvation and the response was genotype specific. We propose the upregulation of the phenylpropanoid metabolism as important source of nitrogen in the form of harmful ammonia during carbohydrate starvation. Our results provide a framework to identify processes associated with genotypic differences in the response to carbohydrate starvation and susceptibility to floret degreening.

2.1   Introduction

The environment to which plants and their products are exposed during post-harvest life is often quite different from their original growth environment in terms of light, temperature, and availability of nutrients. During production in greenhouses, light intensity and temperature are optimized for fast growth and high yields, whereas the environment during post-harvest storage and transport is aimed at keeping the quality and appearance of the flower product as it was at harvest. Light deprivation often means absence of photosynthesis and therefore no *de novo* production of carbohydrates. However, respiration continues in the dark, especially in products that are harvested in a developmental stage characterized by fast growth and high respiration rate, like flowers and sprouts. Eventually, this leads to a state of carbohydrate starvation, which is often detrimental for plant quality, since carbohydrates are the main source for respiration in plants (Plaxton and Podestá, 2007). Carbohydrate starvation leads to oxidative stress (Couée et al., 2006; Morkunas et al., 2003), which eventually results in programmed cell death (Tiwari et al., 2002). The exact process leading from carbohydrate starvation to cell death remains largely unknown.

Chrysanthemum (*Chrysanthemum x morifolium*) is an economically important cut flower, known for a relatively high postharvest performance compared to other cut flower species. However, after dry, dark, and cold storage during two weeks or more, some genotypes show quick degreening of disk florets, and eventually turn from green to yellow to brown, thereby losing market value. Yellowing is most likely due to loss of chlorophyll, while browning is likely the result of enzymatic oxidation of phenolic compounds into brown polymers, as is the case in many browning processes in plants (Ke and Saltveit, 1988; Pourcel et al., 2007). In unstressed cells, this oxidation is prevented by differential subcellular compartmentalization, but enzymes and substrates come together if membranes are leaky (Pourcel et al., 2007). Membrane damage can occur upon physical damage as in potato bruising (Bachem et al., 1994), during senescence (Thompson et al., 1998), or induced by stress, like carbohydrate starvation (Aubert et al., 1996).

Feeding flower stems with carbohydrates often increases vase life (Ichimura et al., 2005; Yakimova et al., 1996). In addition, feeding florets with sugars prevents degreening in broccoli (*Brassica oleracea*

var *italica*) and *Arabidopsis thaliana* (Irving and Joyce, 1995; Trivellini et al., 2012). Finally, feeding chrysanthemum flower stems with sugars also reduces degreening (Van Geest *et al.,* unpublished results). These results indicate that carbohydrate metabolism is very relevant for postharvest performance, specifically for problems related to degreening. There are large genotypic differences in sensitivity to degreening in chrysanthemum. Under the assumption that carbohydrate content plays a role in degreening sensitivity, these genotypes respond differently to a low availability of carbohydrates. Knowledge about genotypic variation in the response to carbohydrate starvation is generally absent. However, it has large potential to aid breeding for postharvest performance.

[1]H Nuclear Magnetic Resonance (NMR) allows simultaneous quantitation of both primary and secondary plant metabolites (Kim et al., 2010a), and therefore can be used to investigate a wide range of phenomena at the metabolomic level. Characterization of the carbon starved metabolome using ^{1}H NMR of *Arabidopsis thaliana* cell cultures (Kim et al., 2007) and ^{31}P-NMR of *Acer pseudoplatanus* cell cultures (Aubert et al., 1996), demonstrated marked increases of free amino acids, malic acid and phosphorylcholine. These observed metabolic changes are typical for carbohydrate starvation, but their amplitude or occurrence might differ strongly between genotypes that differ for sensitivity to carbohydrate starvation.

In order to investigate the genotypic differences in the metabolic response to carbohydrate starvation, we used a metabolome-wide approach to detect compounds that quantitatively vary during carbohydrate starvation. By constructing metabolomic profiles of disk florets using ^{1}H-NMR and High Performance Anion Exchange Chromatography (HPAEC) at different time points in three chrysanthemum genotypes during carbohydrate starvation and feeding, we identified genotype specific and general metabolites associated with starvation. This paper discusses the differential metabolic response to carbohydrate starvation between genotypes, and proposes hypotheses that explain how a genotype can affect the tolerance to starvation of carbohydrates by altering its metabolome.

2.2   Materials and methods

2.2.1   Plant material and pre- and postharvest conditions

Three chrysanthemum genotypes were obtained from Deliflor chrysanten B.V. (Maasdijk, the Netherlands): DB39287, which is a single (daisy-like) white type and sensitive to degreening of the disk florets after long storage (S), DB32030, a single purple type which is medium sensitive to degreening (MS), and DB36451, a single white type which is insensitive to degreening (I). Sensitivity was previously assessed by Deliflor Chrysanten B.V. (Maasdijk, the Netherlands) in commercial vase life tests. Plants were grown in a greenhouse in Maasdijk, the Netherlands using commercial growing practices. For initial respiration measurements, plants were grown from April to June 2014, harvested from large fields in the same greenhouse. For respiration measurements over time, plants were grown from August to November 2014 in a randomized block design. For both experiments, flowers were transported to Wageningen (the Netherlands) on the same day. For NMR, HPAEC and protein measurements, plants grown from July to September 2013 were used. Plots were planted in a randomized block design with five blocks representing five replicates. For each time point, one stem was harvested per plot. For all experiments, harvest took place in the morning, and stems were transported dry in cardboard boxes to Wageningen, the Netherlands. Except for respiration measurements, the combination of the disk florets of the upper three capitula of a flower stem was considered as one biological replicate. For all treatments, per stem the upper three capitula were cut in an angle of approximately 45°, and ray florets were removed. Disk florets sampled at day 0 were frozen in liquid nitrogen upon arrival at the lab. The cut ends of the three capitula were placed in a 35 mL pot containing 25 mg L^{-1} sodium dichloroisocyanurate (DICA) in demineralized water as bactericidal compound for the carbohydrate starvation treatment, or the same amount of DICA with 50 mM sucrose for the sugar feeding treatment. Solutions were replaced on every day on which sampling was performed. Pots with capitula were placed in a dark temperature controlled cabinet at 20 ± 1 °C. Air humidity was buffered using a 2 L saturated sodium chloride solution in the cabinet resulting in a relative humidity of approximately 75%. Samples

of disk florets were taken at 0, 5, 10 and 14 days after harvest. When sampled, florets were flash frozen in liquid nitrogen and stored at -80 °C until freeze-drying.

2.2.2   Colour measurements

To determine the change in degreening of disk florets over time, pictures of capitula were taken in a standardized light environment created by fluorescent light scattered by a Perspex plate bent in a half cylinder. A Hitachi HV-C20 video camera with a Tamron SP 35-80 mm objective was used for imaging. The average intensity of red and green of the disk florets was quantified using ImageJ (Schneider et al., 2012), using a custom-made macro.

2.2.3   Respiration measurements

Nine capitula were placed in three 35 mL pots (three per pot) containing 25 mg L^{-1} DICA per biological replicate. For initial respiration measurements, flowers were stored dry at 4°C overnight before excising the capitula and respiration measurements. For measurements over time, capitula were excised upon arrival in Wageningen, and placed in the 35 mL pots. Respiration measurements took place the next day. For respiration measurements, three 35 mL pots were placed in an airtight 1 L jar at 20 ± 2 °C. A 1 L jar was considered as biological replicate. Jars carried a septum and contained 40 mL saturated sodium chloride solution to buffer air humidity during the respiration measurement. Three jars per genotype were used. Jars were closed after one hour of temperature acclimation. After closure, 3 mL of air were withdrawn every 30 minutes for three hours. Oxygen concentration of the air sample was measured using a Dansensor CheckMate 3 headspace gas analyser (Dansensor A/S, Ringsted, Denmark). Dry weight was measured after respiration rate measurements by weighing samples after oven-drying at 80 °C overnight. The respiration rate was calculated using the slope of the regression between time and oxygen concentration corrected by total dry weight of all capitula per jar. In order to obtain reliable constant respiration estimates, only regression lines with an R^2 higher than 99 % were considered.

2.2.4   Carbohydrate analysis

Carbohydrates were extracted from a 10-20 mg freeze-dried sample with 5 mL of 80% ethanol for 20 minutes at 80 °C. After extraction, tubes were centrifuged at 8,800 g. One millilitre of supernatant was vacuum dried using a vacuum centrifuge (Savant SpeedVac SPD2010, Thermo Fisher Inc., Waltham, MA USA) at 45 °C and 5.1 mbar for 150 minutes and re-eluted in demineralized water using an ultrasonic water bath (Branson 2200, Branson Ultrasonics, Danbury, CT, USA) at 50 Hz. The remaining pellet was stored in ethanol at -20 °C for determination of starch concentration. Glucose, fructose and sucrose were quantified from the ethanol extract using High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD; Dionex ICS5000, Thermo Fisher Inc.), equipped with a CarboPac1 column (250 x 2 mm) eluted with 100 mM NaOH. Fructans were quantified as described by Verspreet et al. (2012) with some modifications. Of the extract, 150 μL were hydrolyzed with 150 μL 120 mM HCl for 90 minutes at 70 °C. Fructans were quantified and average degree of polymerization (DP) was determined by comparing the glucose, fructose and sucrose concentration of the hydrolysed with non-hydrolysed extract as described by Verspreet et al. (2012). All fructans were present in the 80% ethanol extract since no detectable fructose or glucose was found after hydrolysis of water extracts of the remaining pellet after ethanol extraction. For starch quantitation, the pellet was washed three times with 80% ethanol. The pellet was vacuum dried, resuspended in 2 mL 1 g L^{-1} thermostable α-amylase (SERVA Electrophoresis GmbH, Heidelberg, Germany) and incubated for 30 minutes at 90 °C. After that, 1 mL of 0.5 g L^{-1} amyloglucosidase from *Aspergillus niger* (Sigma 10115, Sigma, St Louis, MO, USA) in 50 mM citrate buffer (pH = 4.6) was added and tubes were incubated at 60 °C for 15 minutes. Tubes were centrifuged at 8,800 g and glucose and fructose were quantified as described above.

2.2.5   Protein measurement

Proteins were extracted in an ice-cold ultrasonic water bath (Branson 2200, Branson Ultrasonics, Danbury, CT, USA; 50 Hz) for 15 minutes from 12-20 mg freeze-dried tissue with 1 mL 100 mM Tris-HCl (pH = 7.6), 10 mM $MgSO_4$, 10 mM dithiothreitol and 0.1% Triton-X 100 in demineralized water (Eason et al., 1997) in a 2 mL microcentrifuge tube. After extraction, tubes were centrifuged at 12,000 g. Of the supernatant, 500 µL were centrifuged at the same speed. Hundred microliters were added to 3 mL Bradford reagent (Bradford, 1976). Absorption was measured at 466 and 594 nm (Varian Cary 4000 UV-VIS Spectrophotometer, Agilent Technologies Inc., Santa Clara, CA, USA). Protein was quantified as described by Zor and Selinger (1996) using bovine serum albumin standard series.

2.2.6 ^{1}H-NMR profiling

Freeze-dried samples were transferred to a 2 mL-micro tube to which 1 mL of the mixture of CH_3OH-d_4 and 90 mM KH_2PO_4 in D_2O (1:1) containing 0.05% (w/w) TMSP (trimethyl silyl propionic acid sodium salt, w/v) was added. The mixture was vortexed at room temperature for 30 s, ultrasonicated for 1 min, and centrifuged at 30000 rpm at 4 °C for 20 minutes. Of the supernatant, 300 µL were taken into a 3 mm-NMR tube for NMR analysis.

^{1}H NMR spectra, 2D J-resolved spectra as well as ^{1}H-^{1}H homonuclear and inverse detected ^{1}H-^{13}C correlation experiments were recorded at 25 °C on a Bruker 600MHz DMX NMR spectrometer (600.13 MHz proton frequency) equipped with TCI cryoprobe and Z-gradient system. CD_3OD was used for internal lock purposes. The parameters followed previous literature (Kim et al., 2010b) with some modifications. For ^{1}H NMR spectra a total of 32768 data points were recorded covering a spectral window of 9615 Hz. Sixty-four scans of a standard one-pulse sequence with 30° flip angle for excitation and pre-saturation during 2.0 s relaxation delay with an effective field of γB_1 = 50 Hz for suppression of the residual H_2O signal was employed. An exponential window function with a line-broadening factor of 0.3 Hz was applied prior to Fourier transformation. The resulting spectra were manually phased and baseline corrected.

2.2.7   Data analysis

The ^{1}H NMR spectra were automatically reduced to ASCII files using AMIX (v. 3.7, Bruker Biospin). Spectral intensities were scaled to TMSP and reduced to integrated regions of equal width (0.04 ppm for ^{1}H NMR spectra; further referred to as 'bin'), corresponding to the region of δ 0.3 – δ 10.00. The region of δ 4.7 - δ 5.0 and δ 3.28 - δ 3.34 were excluded from the analysis because of the residual signal of H_2O and CH_3OH-d_4, respectively. Missing values were imputed by half of the minimum value recorded for that bin. Intensities of NMR profiling and concentrations from HPAEC analysis were normalized by sample dry weight corrected for carbohydrate content (see Discussion):

Equation 1: $cdw_j = dw_j - c_j$

Equation 2: $\bar{x}_{ij} = \frac{x_{ij}}{cdw_j}$

Where x is a datapoint in the ith variable and jth sample, dw represents the dry weight, c carbohydrate content and cdw carbohydrate corrected dry weight. Variables (including compounds measured with HPAEC) were selected that showed significant change in time in any treatment-genotype combination using a Kruskal-Wallis test with p-value cut-off of 0.01. This left 122 variables out of 250. A correlation matrix using Pearson correlation was calculated, and metabolites were clustered using Euclidian distance. Principal component analysis (PCA; Mardia et al., 1979) was performed with the prcomp function from the stats package in R (R Core Team, 2014). Default values were used, except for the argument 'scale.', which was set to TRUE, resulting in unit variance scaling prior to PCA. For the respiration measurements, significant differences were identified using a one-way analysis of variance followed by a Tukey's honest significant difference test. Linear relations between variables were tested using a simple linear model.

2.3 Results

2.3.1 Colour and carbohydrate metabolism

We tested whether isolated flower heads in water showed the same degreening compared to commercial vase life trials as executed by Deliflor Chrysanten B.V. Our system revealed the expected genotypic differences (Figure 2.1). The increase in red/green (R/G) intensity represented visually observable degreening in the disk florets (Figure 2.2). Significant increase in R/G occurred at day 10 in the sensitive genotype, and at day 14 in the MS genotype. It did not occur during the experiment (stopped after 14 days) in the insensitive genotype. Application of sugar to the medium in which the flower heads were kept prevented degreening in all genotypes and facilitated development of disk florets towards maturation (Figure 2.1).
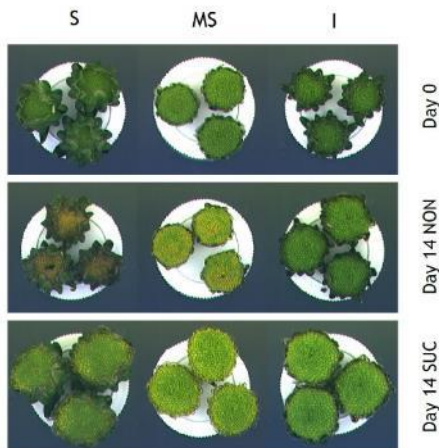

Figure 2.1. Appearance of flower heads of three genotypes before and at the end of two treatments. First column: sensitive (S), second column: medium sensitive (MS) and third column: insensitive (I) genotype. -SUC: carbohydrate starvation, +SUC: treatment with 50 mM sucrose.
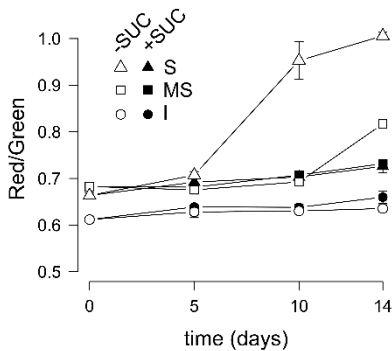

Figure 2.2. Average intensity of red divided by average intensity of green over time of disk florets with (closed symbols) or without (open symbols) 50 mM sucrose feeding for three genotypes: sensitive (S; triangles), medium sensitive (MS; squares) and insensitive (I; circles). Error bars represent standard errors of the mean, n=5.

Carbohydrates provided a significant part of the total dry weight, and carbohydrate content varied strongly during sucrose feeding and carbohydrate starvation (Figure 2.3A). Therefore, we normalized measured variables by the carbohydrate corrected dry weight (Equation 1 and 2; Additional file 2.1).

Sucrose feeding increased carbohydrate corrected dry weight, whereas during starvation carbohydrate corrected dry weight remained stable initially, and eventually declined.

Sucrose feeding resulted in an altered carbohydrate content and distribution between mono-, oligo-, and polysaccharides. This distribution was also highly dependent on genotype. The relative amount of fructans was much higher in the MS genotype compared to the S and I genotypes. The S genotype contained 73% less carbohydrates at harvest compared to the other genotypes (Figure 2.3). Starch was not very abundant; it made up at most 5% of total carbohydrates and declined quickly within the first five days irrespective of sucrose treatment. Sucrose, glucose, and fructose (Figure 2.3B, C and D respectively) content decreased only upon carbohydrate starvation. The decrease in glucose was faster than that of fructose. Fructose and glucose increased strongly with addition of sucrose, whereas sucrose increased slightly. Similar to the effect of sucrose on starch content, sucrose treatment only slightly increased fructan content. Protein loss only occurred after major losses of carbohydrates, and occurred concomitantly with degreening (Additional file 2.2).
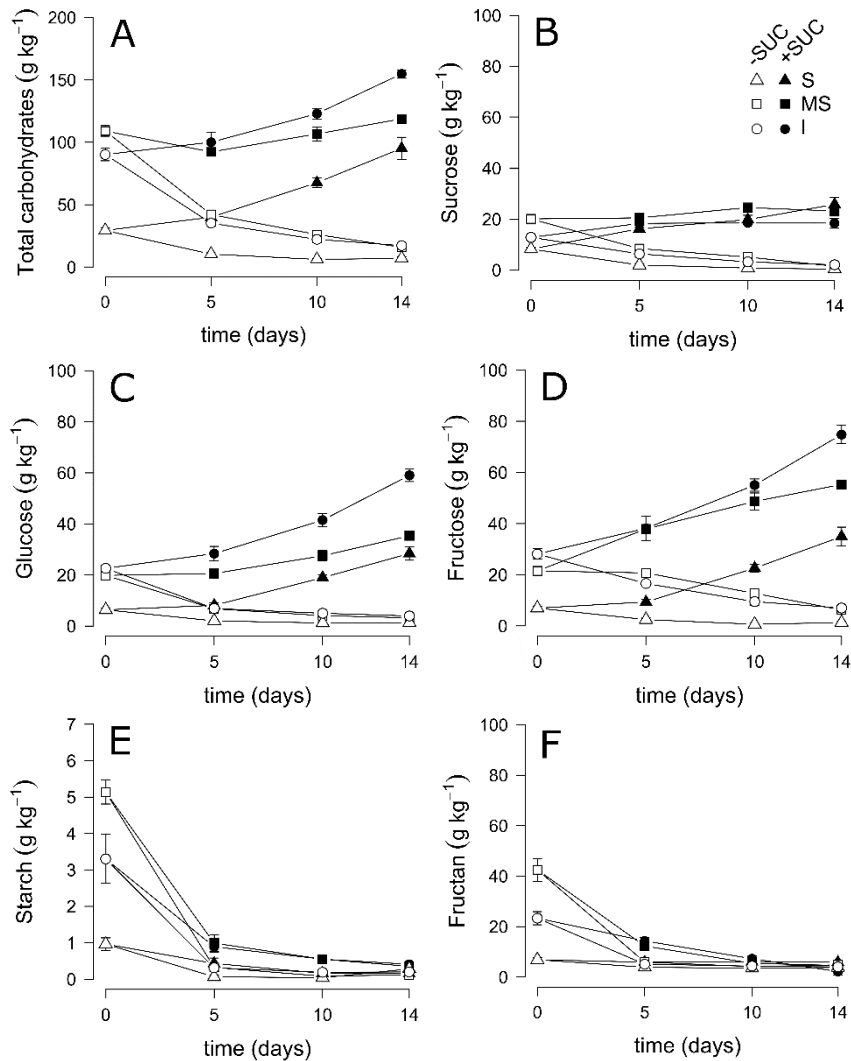
Figure 2.3. Disk floret total soluble carbohydrate (A), sucrose (B), glucose (C), fructose (D), starch (E) and fructan (F) content over time as measured by HPAEC of disk florets with (closed symbols) or without (open symbols) 50 mM sucrose feeding for three genotypes: sensitive (S; triangles), medium sensitive (MS; squares) and insensitive (I; circles). Error bars represent standard error of the mean, n=5.

Despite having a similar carbohydrate content, respiration rate was highest in the MS genotype and lowest in the I genotype at harvest (Additional file 2.3). Respiration rate decreased during the carbohydrate starvation and increased if sucrose was fed (Additional file 2.4A). Within genotype there was a strongly significant correlation between carbohydrate content and respiration rate in the S and MS genotype (p<0.001) and a weak but significant (p=0.02) correlation in the I genotype (Additional file 2.4B). The MS and S genotype had a much stronger decrease in respiration rate with lower carbohydrate contents compared to the I genotype.

2.3.2 Multivariate analysis of the metabolic response

Carbohydrate starvation drastically changed the disk floret metabolome. All carbohydrates measured with HPAEC changed significantly over time for at least one genotype-treatment combination according to a Kruskal-Wallis test at p=0.01 (Figure 2.4). NMR spectral intensities were binned in integrated regions of 0.04 ppm. From these bins, 116 out of 253 changed significantly over time. Under sugar feeding the metabolome remained rather stable, with eleven, twelve, and one bin for the I, S and MS genotype respectively changing significantly over time (not shown). Most bins that changed over time changed only in the carbohydrate starvation treatment.

1D and 2D based NMR allowed identification of 21 compounds (Table 2.1). Based on correlation in samples, we identified two main groups of metabolites using Euclidian distance (Figure 2.4). The first, group A, contained the carbohydrates, in which the soluble carbohydrates were highly correlated with each other. A larger second group, group B, generally correlated negatively with group A. Group B was metabolically diverse. It contained most identified amino acids, some organic acids, and molecules containing aromatic rings among which flavonoids, phenylpropanoids, and aromatic amino acids. Of the identified compounds, the strongest negative correlation with group A was found for asparagine (Figure 2.4).

Table 2.1. Identified compounds using NMR and HPAEC-PAD

Metabolite	Chemical shift (NMR)	Retention time (HPAEC-PAD) minutes	Significant change?*
Polysaccharides†			
Starch	-	-	yes
Fructans	-	-	yes
Sugars			
Glucose	δ 5.20	5.067	yes
Sucrose	δ 5.40	10.033	yes
Fructose	δ 4.08	5.65	yes
Amino acids			
Valine	δ 1.04	-	yes
GABA	δ 3.04	-	no
Glutamine	δ 2.44	-	yes
Arginine	δ 1.72	-	yes
Threonine	δ 1.32	-	yes
Adenine	δ 8.20	-	no
Asparagine	δ 2.84	-	yes
Alanine	δ 1.48	-	yes
Small organic acids			
Citric acid	δ 2.72	-	yes
Formic acid	δ 8.48	-	yes
Phenylpropanoids			
Chlorogenic acid	δ 6.36	-	yes
3,5-Dicaffeoyl quinic acid	δ 6.48	-	yes
Flavonoids			
Luteolin-7-O-glycoside	δ 7.44	-	yes
Apigenin-7-O-glycoside	δ 7.92	-	no
Vitexin analogues	δ 8.00	-	yes
Other			
Choline	δ 3.20	-	no
myo-Inositol	δ 3.64	-	yes
Camphor	δ 0.96	-	yes

*Significant change over time in at least one treatment-genotype combination according to a Kruskal-Wallis test at p=0.01
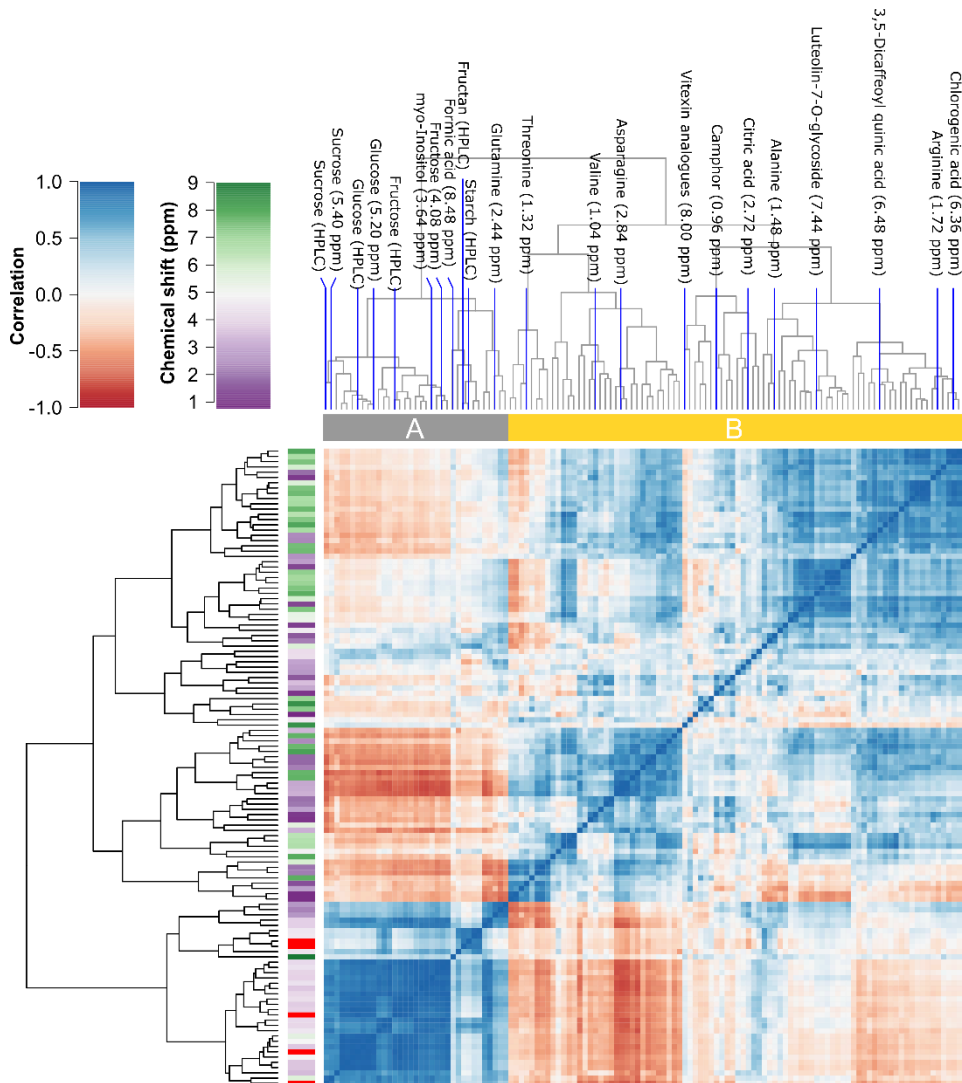†See 'Materials and Methods' for quantitation method

Figure 2.4. Heatmap of Pearson correlation matrix of 128 variables; 122 variables derived from NMR spectroscopy and six variables derived from HPAEC that significantly changed normalized intensity over time in at least one genotype in one of the treatments according to a Kruskal-Wallis test at p=0.01. The order of compounds is based on clustering by Euclidian distance and indicated by the dendrograms. On the vertical colour bar, chemical shift of the bins are depicted as colour, the rows with compounds measured with HPAEC are shown in red. On the horizontal axis, two distinct groups are depicted, and 22 compounds identified with HPAEC and/or NMR are indicated (blue lines).

We conducted a principal component analysis (PCA) in order to visualize maximum variation between samples and identify compounds that explain this variation. The first three components explained 67% of the variation. The score plot (Figure 2.5) of the two components explaining most variation separated carbohydrate starvation from sugar feeding at the first component, with low carbohydrate content on the right, and high carbohydrate content on the left side. This separation was also characterized by high amounts of asparagine and an unidentified compound characterized by a bin at δ 7.32 for sugar-starved samples. The second component separated mainly genotypes. Bins characterizing luteolin-7-*O*-

glycoside, chlorogenic acid and bins at δ 2.12 and δ 2.48, which likely characterize glutamine content, had large influence on this separation. These compounds were particularly high in the MS genotype. They were low in the sensitive genotype and intermediate in the insensitive genotype. Separation between genotypes was explained in the opposite direction by a bin at δ 0.84 and threonine, which were high in the S genotype, low in the MS and intermediate in the I genotype.
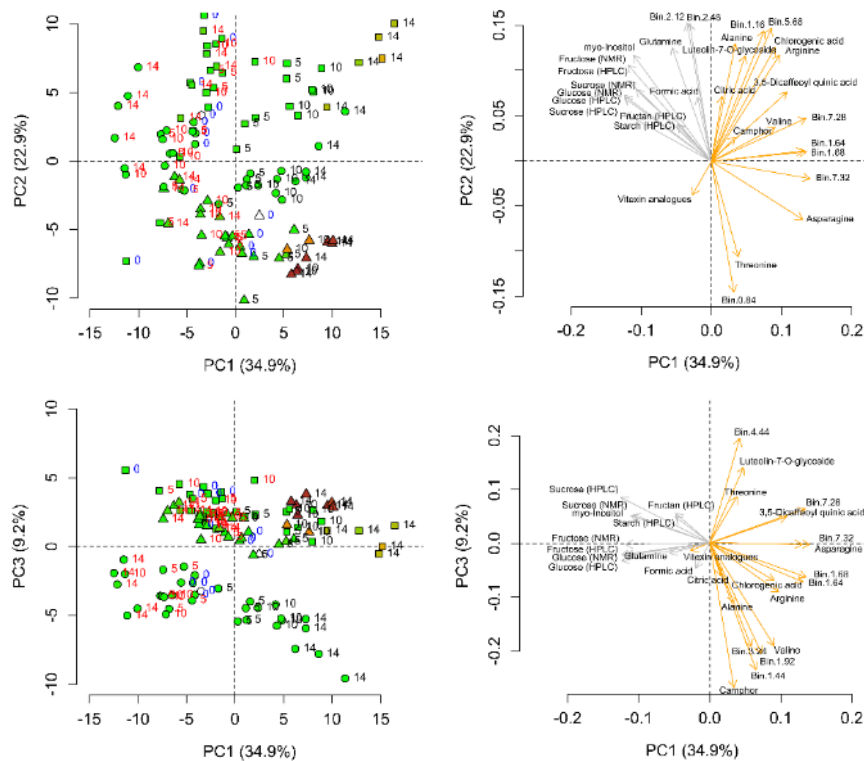

Figure 2.5. PCA score plots (left) and loading plots (right), of principal component 1 and 2 (upper) and component 1 and 3 (lower). Genotypes: Insensitive (○), medium sensitive (□) and sensitive (△). Numbers at symbols in score plots represent day after harvest and their colours represent treatments, which are: directly after harvest (blue), sugar feeding (red) and carbohydrate starvation (black). The colour of the symbols represents the measured colour of the disk florets going from green to brown. In the loading plots, the identified compounds, and five compounds with largest absolute loading along each component were plotted. The colour of the arrows corresponds to the colour of the groups depicted in Figure 2.4, where grey represents group A and orange group B.

In the PCA score plot with the first and third component (Figure 2.5), degreened sugar starved disk florets from both the sensitive and MS genotype clustered together. Degreening only occurred if carbohydrates were low, in combination with presence or absence of compounds explaining variation along the third component. Degreened and brown flower heads were low in camphor. 3,5-dicaffeyol quinic acid and peaks within the bin at δ 4.44, which was too small for identification, were high in degreening tissue.

Four identified amino acids quantified with NMR explaining major variation in the PCA score plots were selected for further inquiry (Figure 2.6). Arginine, asparagine, and valine (Figure 2.6A, B and C respectively) increased in the first days of carbohydrate starvation and their concentration remained constant with sugar feeding. Threonine specifically increased only under carbohydrate starvation in the S genotype, whereas arginine did not change much at carbohydrate starvation in the S genotype (Figure 2.6D).
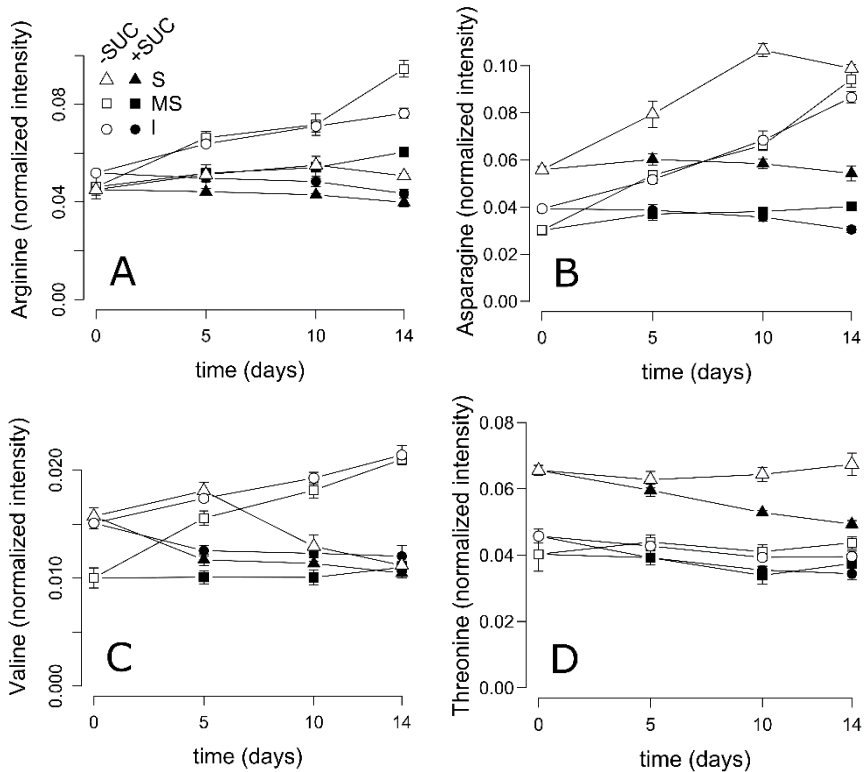
Figure 2.6. The normalized bin intensity of arginine (A), asparagine (B), valine (C) and threonine (D) in disk florets with (closed symbols) or without (open symbols) 50 mM sucrose feeding for three genotypes: sensitive (S; triangles), medium sensitive (MS; squares) and insensitive (I; circles). Error bars represent standard error of the mean, n=5.

The phenolic compounds chlorogenic acid, 3,5-dicaffeoyl quinic acid and luteolin-7-*O*-glycoside generally increased upon carbohydrate starvation (Figure 2.4; Figure 2.7A B and C respectively). They are substrates for polyphenol oxidases and peroxidases, and therefore precursors of brown polymers. The MS genotype showed a stronger increase in chlorogenic acid and 3,5-dicaffeyol quinic acid upon carbohydrate starvation compared to the other two genotypes (Figure 2.7A and B). At harvest, it had much higher levels of luteolin-7-*O*-glycoside (Figure 2.7C), but the compound did not increase upon carbohydrate starvation, whereas it did for other genotypes. The terpenoid camphor was associated with remaining green at carbohydrate starvation (Figure 2.5). It increased in the MS and I genotype, but not in the S genotype. At harvest, it was most abundant in the I and MS genotype but increased more strongly in the I genotype compared to the MS genotype during carbohydrate starvation (Figure 2.7D).
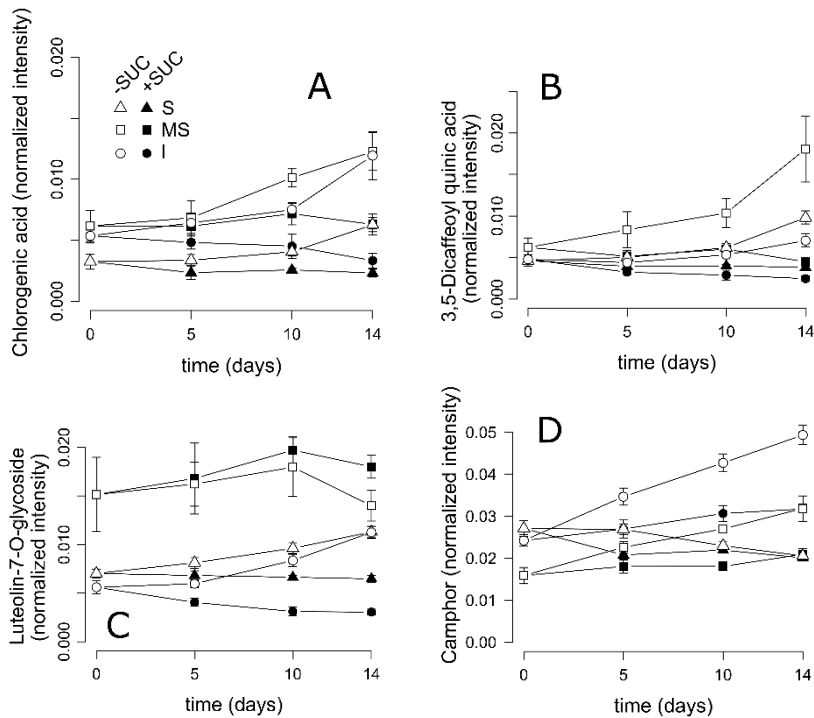
Figure 2.7. Normalized bin intensity of chlorogenic acid (A), 3,5-dicaffeoyl quinic acid (B) luteolin-7-*O*-glycoside (C) and camphor (D) in disk florets with (closed symbols) or without (open symbols) 50 mM sucrose feeding for three genotypes: sensitive (S; triangles), medium sensitive (MS; squares) and insensitive (I; circles). Error bars represent standard error of the mean, n=5.

2.4   Discussion

Although different from commercial vase life trials, our method revealed the same degreening and genotypic differences in susceptibility. By isolating capitula, we could investigate disk floret degreening on a simplified level, without effects of metabolite transport from other parts of the plant. In this section, we bring observations in a broader context, by first defining the type of stress we presume to play a role, than by discussing genotypic differences in carbohydrate metabolism and its effect on susceptibility to degreening, and last by discussing the genotype specific and common responses to carbohydrate starvation.

2.4.1   Sugar feeding effects on appearance

It is very likely that degreening of disk florets is the result of carbohydrate starvation. Application of 50 mM sucrose to isolated flower heads prevented degreening entirely. Other types of soluble sugar like glucose and fructose could also prevent degreening, but not the non-metabolisable sugar analog *myo*-inositol (not shown). This process is similar to broccoli (*Brassica oleracea* var. *italica*) and *Arabidopsis thaliana*, where sucrose feeding strongly reduces sepal yellowing (Irving and Joyce, 1995; Trivellini et al., 2012).

Despite the fact that similar processes occur during carbohydrate starvation, dark-induced deterioration and senescence, there are also substantial differences. These differences are present in crucial pathways related to amongst others amino acid metabolism, lipid catabolism, and flavonoid biosynthesis (Buchanan-Wollaston et al., 2005; Trivellini et al., 2012; van der Graaff et al., 2006). The word 'senescence' is derived from the Latin *senex*, which means 'aged' or 'matured', and is defined as 'the

growth phase in a plant or plant part (as a leaf) from full maturity to death' (Merriam-Webster Online, 2015). Since disk florets are far from reaching full maturation when degreening occurs, it can best be described and treated as a stress-induced deterioration. This concept should also apply to other instances of premature deterioration of plant organs.

2.4.2   Normalization by carbohydrate corrected dry weight

Soluble sugars made up a very large part of the dry weight in chrysanthemum disk florets. At harvest, up to 10.4% of the dry weight consisted of carbohydrates. Under carbohydrate starvation, it is evident that most of these carbohydrates disappear. Normalization by dry weight would result in a substantial artificial increase of other metabolites. Other researchers have solved this issue by expressing a compound per organ, such as per flower (Eason et al., 1997) or per root tip (Brouquisse et al., 1992). This approach does not solve the problem if multiple genotypes are used. As demonstrated in our results, there can be major genotypic differences in carbohydrate content at harvest. In addition, organ size may differ between genotypes, making it impossible to compare genotypes if corrected by organ. In order to enable comparison between genotypes and time points, we attempted to correct for the change in dry weight over time by subtracting the total carbohydrate weight from the dry weight. Other metabolites change and will have effect on the dry weight as well, but their contribution to total dry weight is probably negligible compared to the effect of carbohydrates.

2.4.3   Polysaccharide breakdown is independent of carbohydrate content

In contrast to starch, fructans form by far the most abundant polysaccharides in chrysanthemum disk florets. They make up 24 to 40% of the total amount of carbohydrates. This is different from whole inflorescences, where starch is the most abundant polysaccharide (Trusty and Miller, 1991). Fructans are ubiquitous in the plant kingdom, including the agricultural relevant Liliaceae, Poaceae and Asteraceae (Hendry, 1993). They seem to be an important source of monosaccharides in the form of fructose during the first days of carbohydrate starvation, since during fructan breakdown in the first five days, fructose content remained stable, whereas glucose content decreased strongly.

Breakdown of polysaccharides in disk florets was independent of carbohydrate content. Fructans are synthesized from sucrose by sucrose:sucrose 1-fructosyltransferase and fructan:fructan 1-fructosyltransferase (Ritsema and Smeekens, 2003). Starch is synthesized from glucose through ADP-glucose by starch synthase (Preiss, 2009). Although sucrose and glucose concentrations increased upon sucrose feeding, disk florets were unable to synthesize polysaccharides. Fructans are water-soluble and can be synthesized elsewhere in the plant. Chrysanthemum stems have high fructan concentrations (Adachi et al., 1999), and could therefore be the source of fructans in the disk florets. Developing flowers often show breakdown of polysaccharides (Bieleski, 1993; van Doorn and Van Meeteren, 2003; Vergauwen et al., 2000). Fructan and starch breakdown in disk florets could therefore be part of a developmental process. Since we kept flower heads in the dark, polysaccharide breakdown might also be induced by light deprivation. However, light did not have a positive effect on starch content during the post-harvest phase in lettuce (Zhan et al., 2013).

2.4.4   Genotypic differences in carbohydrate metabolism

Carbohydrate content could only partly explain genotypic differences in degreening susceptibility. Despite this, our results show that occurrence of degreening within genotypes is clearly related to absence of carbohydrates. Although the S genotype showed by far the lowest carbohydrate content at harvest, the total carbohydrate content between the MS and I genotype was not different during carbohydrate starvation. The MS genotype even had a higher carbohydrate content in many cases, despite having a higher respiration rate at harvest. Since carbohydrate starvation likely causes degreening, the MS and I genotype probably have a different response to low carbohydrate content. This difference in response is also indicated by a different response in respiration rate to carbohydrate starvation. Whereas respiration rate is strongly reduced upon carbohydrate starvation in the MS genotype, respiration rate of the I genotype is much less reduced. This more abrupt reduction in

respiration rate in the MS genotype might be an indication of an earlier constraint on the mitochondrial electron transport chain (mtETC). Carbohydrate starvation leads to reduced ADP contents (Brouquisse et al., 1991), which will impair ATP synthase (Gout et al., 2014). Blockage of ATP synthase will lead to over reduction of the mtETC and therefore oxidative stress (Vanlerberghe, 2013). As indicated by the stronger reduction in respiration rate, the MS and S genotypes might have earlier blockage of the mtETC leading to oxidative stress, resulting in cell damage (Rhoads et al., 2006) and earlier degreening. Differences in ADP availability, alternative oxidase activity or other protective mechanisms might cause these differences.

2.4.5   Protein is degraded when degreening starts

As soon as degreening occurred, disk florets lost proteins. Protein degradation often occurs concomitant to loss of quality in stored fresh produce (Collier, 1997; Downs et al., 1997; King et al., 1990) and in sugar starved tissue (Aubert et al., 1996; Brouquisse et al., 1992; Devaux et al., 2003). Proteins, next to lipids, can provide energy as substrate during carbohydrate starvation (reviewed by Araújo et al., 2011). It seems that protein breakdown only occurs if carbohydrate content is low for five days or more, since significant protein breakdown was only observable at day 14 in the MS and I genotype and at day 10 in the S genotype, whereas carbohydrates were already very low at day 10 and 5 respectively. Loss of protein could connect carbohydrate starvation to degreening. Proteins are indispensable for cell survival. If a significant part of viable processes is lost due to breakdown of proteins in essential enzymatic steps, it is inevitable that cells lose membrane integrity and eventually die. However, at formation of the brown polyphenols, proteins can be incorporated in the complex polymeric structure and become insoluble (Stevens and Davelaar, 1996). In addition, a decline of protein content was mainly observed when degreening occurs, and not before. It is therefore more likely that the protein loss is an effect of the reduction of membrane integrity and not a cause.

2.4.6   Metabolic response to carbohydrate starvation is diverse

Treatment duration and sugar feeding affected a large part of the metabolome of chrysanthemum disk florets. In total, 116 out of 253 bins changed significantly over time for at least one of the genotype-treatment combinations, showing that the response to carbohydrate content is very diverse. Most of the metabolites having a change in concentration over time changed only significantly in the sucrose starvation treatment. The response to carbohydrate starvation was therefore more drastic than the response to sucrose feeding.

2.4.6.1   Phenolic compounds and flavonoids

Oxidation of flavonoids and phenylpropanoids most likely caused browning at the last step of the degreening process. They are oxidized if they come in contact with polyphenol oxidases (PPO) and peroxidases upon membrane damage (Ke and Saltveit, 1988; Pourcel et al., 2007). Membrane damage most likely occurs at late stages of carbohydrate starvation. Chlorogenic acid, 3,5-dicaffeoyl quinic acid and luteolin-7-*O*-glycoside increase early during carbohydrate starvation, and therefore most likely before membrane damage occurs. This means that carbohydrate starvation not only results in increased access of PPO to its substrates, but also to an earlier increase in synthesis of those substrates.

Varying classes of phenylpropanoids increase during different kinds of stress (Dixon and Paiva, 1995). Genes involved in flavonoid biosynthesis are strongly upregulated during sugar starvation in *A. thaliana*, and the type of response differs substantially from leaf senescence (Buchanan-Wollaston et al., 2005). Some genes of the flavonoid biosynthesis pathway are specifically regulated by the sugar starvation signalling pathway (Baena-González et al., 2007). Among those is flavanone 3-hydroxylase, which catalyses the synthesis of flavonols from flavanones. This suggests a specific role for flavonoids during carbohydrate starvation. Since flavonoids are potent anti-oxidants (Agati et al., 2012), this role could be in the protection against oxidative stress. Luteolin-7-*O*-glycoside increased in the I and MS genotype during carbohydrate starvation, whereas apigenin-7-*O*-glycoside did not change significantly, resulting in an increased ratio between luteolin and apigenin. This increased ratio is also observed during UV-B

light stress (Markham et al., 1998), and was suggested to result in larger anti-oxidative capacity (Agati et al., 2012). However, the reaction in which luteolin is synthesized from apigenin is catalysed by flavonoid 3'-hydroxylase, which is specifically upregulated during leaf senescence and not carbohydrate starvation (Buchanan-Wollaston et al., 2005). Our observed increase in luteolin might therefore result from an alternative pathway, but it is more likely that flavonoid 3'-hydroxylase activity increases during carbohydrate starvation in chrysanthemum disk florets as opposed to the sugar-starved *Arabidopsis* cell cultures as investigated by Buchanan-Wollaston et al. (2005).

Phenylpropanoid and flavonoid content was genotype dependent, but did not correspond to browning sensitivity. However, since they are important precursors of the brown polymers, their quantity might be important for the intensity of browning that occurs after the loss of green colour. This would suggest the MS genotype would have a higher browning intensity compared to the S and I genotype. We did not measure browning intensity in this study, but visual observations of commercial vase life trials seem to substantiate this statement (results not shown).

2.4.6.2   *Amino acid and nitrogen metabolism*

If changed, the amino acids arginine, asparagine and valine increased over time during carbohydrate starvation. In maize root tips, there is a strong increase of amino acids at 24 to 48 hours of carbohydrate starvation and cell damage became irreversible when amino acid levels declined back to the initial level (Brouquisse et al., 1992). This dramatic transient change was not present in our material, and a decline in amino acid level was only observed in the S genotype for valine and slightly for asparagine. The type of tissue probably causes the difference in timing of these events between chrysanthemum disk florets and maize root tips; root tips have most likely much lower carbohydrate reserves compared to disk florets and are therefore much earlier exhausted of carbohydrates. Since there was no decrease in measured amino acid levels in the MS and I genotype, they probably did not reach the threshold at which cell damage becomes irreversible during the experiment, but the S genotype might.

Arginine and asparagine have a relatively high ratio between nitrogen and carbon. It is suggested they function as scavengers for toxic ammonium (Brouquisse et al., 1992). The use of protein as alternative substrate for respiration is generally suggested as main cause for the ammonium increase during carbohydrate starvation (Brouquisse et al., 1992; Devaux et al., 2003; Gary et al., 2003; Gaufichon et al., 2010). However, higher activity of the key enzyme in phenylpropanoid metabolism, phenylalanine ammonia lyase (PAL), would also result in production of ammonium. Despite a strong increase in asparagine, protein content did not decrease at the first days of carbohydrate starvation. However, compounds synthesized through the phenylpropanoid pathway increased in the first days of carbohydrate starvation. Therefore, upregulation of the phenylpropanoid metabolism through higher activity of PAL might be responsible for release of ammonium and therefore increase in asparagine and arginine at the cost of carbon. This would suggest that upregulation of the phenylpropanoid metabolism is more important for the alteration of nitrogen metabolism during carbohydrate starvation than protein breakdown, especially during early stages of starvation. These hypotheses and their relations are visualized in Figure 2.8.
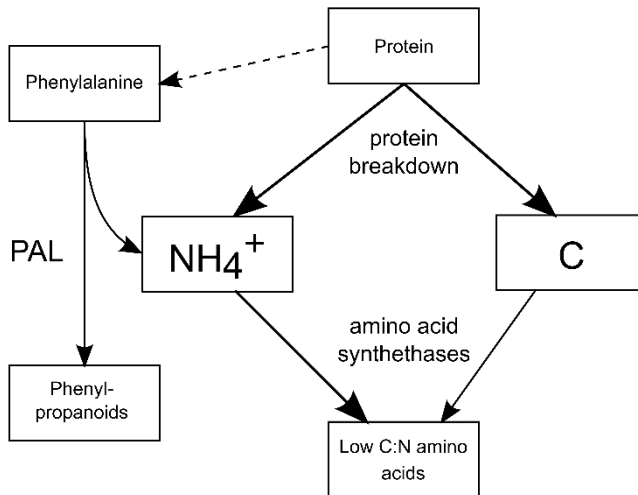
Figure 2.8. Concept map of ammonium metabolism during carbohydrate starvation. Arrow size represents relative flow, the dotted arrow represents a relation of which significance is unknown. PAL: phenylalanine ammonia lyase; C: carbon. Protein can serve as carbon source during carbohydrate starvation, and thereby releases ammonium. Ammonium is scavenged by synthesizing high C:N amino acids like arginine and asparagine, resulting in a net synthesis of free carbon. However, protein breakdown is only significant late during carbohydrate starvation. A second source of ammonium is PAL, which catalyses a rate limiting step for phenylpropanoid biosynthesis. This process leads to net loss of free carbon.

2.4.6.3   Possible roles for camphor

Camphor was the main compound that could explain genotypic differences in the response to carbohydrate starvation in terms of degreening. It is a terpenoid and occurs in several specific plant families. It is the main component in essential oils extracted from flower heads of several species in the chrysanthemum genus, like *Chrysanthemum indicum* (Shunying et al., 2005) and *Chrysanthemum coronarium* (Alvarez-Castellanos et al., 2001; Basta et al., 2007).

The I genotype showed the strongest increase in camphor upon carbohydrate starvation and largest absolute concentration. However, these differences only explained genotypic differences under carbohydrate starvation; the camphor content at harvest did not correspond to susceptibility to degreening. How camphor could play a role in preventing degreening remains unknown. It might be easily oxidized and therefore serve as an anti-oxidant under starvation stress.

Chrysanthemum essential oils have antifungal and antibacterial activity (Alvarez-Castellanos et al., 2001; Kim et al., 2003; Shunying et al., 2005). The antibacterial activity can be specifically addressed to camphor (Lee et al., 2011). The increase of camphor in carbohydrate starved disk florets might be the result of an abiotic stress-induced pathogen defence mechanism, like drought stress-induced increase of defence metabolites in tomato (English-Loeb et al., 1997). Our observed increase in phenylpropanoids likely also adds to this pathogen defence (Dixon et al., 2002). These types of responses could be very relevant for postharvest quality, since large amounts of fresh produce are being lost because of infection of microorganisms.

2.5   Conclusions

Carbohydrate metabolism plays an important role in postharvest performance. Our results show that carbohydrate starvation increases the level of secondary metabolites with biologically and agriculturally relevant functions. Among those are camphor, phenylpropanoids, and flavonoids. Genotypes respond differently in primary and secondary metabolism to carbohydrate starvation. Understanding this differential response between genotypes on a physiological level helps to identify processes causative

for genotypic differences in their response to carbohydrate starvation and susceptibility to degreening. This paper poses hypotheses for causative relations between metabolic events and degreening susceptibility. We are currently testing these hypotheses in a broader set of genotypes. Once causative processes are identified, QTL mapping based on specific phenotypes (at the gene expression, metabolomic, and phenomic level) together with characterization of causative allelic variation based on candidate genes become much more feasible.

2.6   Acknowledgements

2.7   Additional files

Additional file 2.1. Carbohydrate corrected dry weight of all disk florets in one flower. Error bars represent standard error of the mean, n=5. Three genotypes were used: a sensitive (S; triangles), medium sensitive (MS; squares) and insensitive (I; circles). Flower heads were fed with 50 mM sucrose (closed symbols) or not fed (open symbols).

Additional file 2.2. Protein change over time as relative from initial (A). Inlay: protein content of the three different genotypes at day 0. Error bars represent standard error of the mean, n=5. (B) Protein content versus relative hue angle. Flower heads were fed with 50 mM sucrose (closed symbols) or not fed (open symbols). Three genotypes were used: a sensitive (S; triangles), medium sensitive (MS; squares) and insensitive (I; circles).

Additional file 2.3. Respiration rate one day after harvest. Error bars represent standard error of the mean, n=3. S: sensitive; MS: medium sensitive; I: insensitive to degreening.

Additional file 2.4. Respiration rate after harvest over time (A) and plotted against soluble sugar content (B). Flower heads were fed with 50 mM sucrose for 24 hours starting from day 1 and 3 (grey) or not fed (open symbols). Three genotypes were used: a sensitive (S; triangles), medium sensitive (MS; squares) and insensitive (I; circles). Lines in (B) are regression lines as a result of fitting a simple linear model on the three genotypes separately, resulting in an R2 of 53.2, 59.9 and 28.8% for the S, MS and I genotypes respectively. Error bars in (A) represent standard error of the mean (n=3).

Chapter 3

Breeding for postharvest performance in chrysanthemum by selection against storage-induced degreening of disk florets

Geert van Geest, Aike Post, Paul Arens, Richard GF Visser, Uulke van Meeteren

Abstract

Breeding for postharvest performance in ornamentals is challenging, since many different deteriorative processes determine vase life. In order to improve postharvest performance by breeding, selection should take place on these processes separately. To define processes that are important for chrysanthemum postharvest performance, vase life was assessed after two weeks of cold storage in a set of 44 chrysanthemum cultivars. Since disk floret degreening was the most frequent reason for ending vase life, we further investigated this trait in a large bi-parental population (n=381). To quantify disk floret degreening in this large number of genotypes, we developed a high-throughput phenotyping method. The method consists of the quantification of loss of green colour as expressed by an increase of intensity of red divided by the intensity of green (R/G) in dark-held detached capitula. R/G increases when disk florets lose green colour. The increase in R/G correlated significantly with the number of days until disk floret degreening occurred during vase life. This was the case for the 44-cultivar cultivar panel (Pearson's correlation coefficient (ρ) of -0.70; p<0.0001) as well as in a subset of the bi-parental population (n=145; ρ=-0.67; p<0.0001). R/G increase segregated in a quantitative manner in the full bi-parental population, and had a moderately high heritability of 0.73. Carbohydrate content after harvest was measured in a smaller subset of the bi-parental population (n=55). The R/G increase correlated with carbohydrate content (ρ=-0.56; p<0.0001). Since carbohydrate content did not explain all variation in degreening sensitivity, we discuss different possible mechanisms to cope with carbohydrate starvation and avoid degreening. In conclusion, disk floret degreening is an important postharvest trait in chrysanthemum, and it is related to carbohydrate starvation. The quantitative segregation suggests involvement of multiple alleles, probably at multiple loci. The moderately high heritability makes it a suitable trait for QTL mapping, which we will commence in the near future.

3.1   Introduction

In many crop plants, and specifically ornamental crops, breeding for postharvest quality is challenging. A reason for this, is that vase life and shelf life terminating symptoms vary within a crop (Fanourakis et al., 2013; Ferrante et al., 2015; Rico et al., 2007) implying that many different unrelated processes can explain genotypic variation. In addition, a wide range of pre- and postharvest environmental variation can interact with the time to occurrence and severity of these symptoms (Fanourakis et al., 2013). The interaction between genotype and environment is therefore an important factor to take into account while investigating postharvest performance.

Breeding aims to maximize the number of favourable alleles in a certain germplasm or genotype. Timing of occurrence and extent of different postharvest-related deteriorative processes is generally encoded by a broad set of alleles and loci (Carvalho et al., 2015; Moreno et al., 2008; Zhang et al., 2007). This large variation strongly impairs the development of breeding tools that allow estimation of the phenotype from a genotype. The overall postharvest performance of a certain crop should therefore be divided in specific parameters that can be measured reliably and are potentially encoded by a limited set of alleles. Genotypic improvement should take place at the level of these parameters. There are successful examples of breeding for specific mechanisms to increase postharvest performance in cut flowers. An example is described by Onozaki et al. (2001) in which the authors could improve carnation vase life by specifically selecting against ethylene sensitivity. Another is described by Carvalho et al. (2015) in which the authors successfully located QTLs explaining vase life by assessing the stomatal control in rose using detached leaves.

During the postharvest phase, plants and their harvested products are often kept in low light environments or even in the dark. Production of carbohydrates is therefore strongly impaired while respiration continues, eventually leading to a state where carbohydrates are limiting. This results in senescence-like processes, for example loss of proteins (King et al., 1990) and chlorophyll (Buchanan-Wollaston et al., 2005; Trivellini et al., 2012), DNA fragmentation (Devaux et al., 2003), ammonium accumulation (Devaux et al., 2003; King et al., 1990), and eventually cell death. Reducing carbohydrate starvation or its symptoms has large potential to improve postharvest performance, as these processes

lead to unwanted colour changes and eventually increased susceptibility to necrotrophic microorganisms.

Variation in postharvest quality of cut chrysanthemum has been related to several different problems: leaf yellowing (Satoh et al., 2008), leaf wilting (van Doorn and Cruz, 2000; van Meeteren, 1992), flower wilting (Adachi et al., 2000), and disk floret degreening (chapter 2). In order to define specific measurable parameters that explain vase life after storage in chrysanthemum, we assessed vase life of a genotype panel for a broad range of vase life terminating symptoms. Following two weeks of cold storage, disk floret degreening was the most frequently occurring vase life terminating symptom; therefore we developed a method to rapidly measure sensitivity for disk floret degreening derived from the *Arabidopsis* Infloresence Degreening Assay (AIDA; Jibran et al., 2015; Trivellini et al., 2012). In the previous chapter (chapter 2), in which we studied three genotypes, we suggest that carbohydrate content might play a role in explaining genotypic differences in disk floret degreening, but that it is not the only factor. In this study, we analysed a much larger set of genotypes as well as a bi-parental population to investigate the roles of initiation of senescence and carbohydrate metabolism in degreening sensitivity.

3.2   Materials and methods

3.2.1   Plant materials

This study describes two groups of genotypes: a cultivar panel and a bi-parental population. The cultivar/breeding line panel consisted of 44 genotypes with single (daisy-like) white flowers. The bi-parental population consisted of offspring of a bi-parental cross between Deliflor breeding lines DB36451 and DB39287 (Deliflor chrysanten B.V. Maasdijk, the Netherlands). This population consisted of 405 offspring of which for 381 offspring we could get enough plant material for measurements (Figure 3.1). The parents were part of the cultivar panel. Plants were maintained and propagated vegetatively. Cuttings were rooted during two weeks and after that planted in a greenhouse in Maasdijk, the Netherlands (latitude: 51.959311; longitude: 4.214427). Harvest took place when a field consisting of plants of the same genotype reached commercial maturity. Plants growing at the border of a field were not used.

3.2.2   Cultivar panel

From the 44 members of the panel, 19 genotypes were grown in three replicates in a randomized design and the remaining 25 genotypes were grown in one replicate (Table 3.1). A replicate consisted of a field ranging from 20 to 100 plants, depending on the availability of cuttings. Of those, up to six stems were used for experiments (see below). Plants were grown from November 2015 to January 2016 in 16 hour photoperiods for 14 days and after that in 12 hour photoperiods to induce flowering.

Table 3.1. Genotypes used in the cultivar panel.

Name	Breeder	Replicates
Astun	Deliflor Chrysanten B.V.	3
Bacardi	Dümmen Orange	3
Breeding lines*	Deliflor Chrysanten B.V.	1-3
Chic	Royal Van Zanten	3
Delilah white	Deliflor Chrysanten B.V.	3
Gletsjer	United Selections	1
Goethe	Deliflor Chrysanten B.V.	3
Major White	Dekker Chrysanten B.V.	3
Neva	Deliflor Chrysanten B.V.	3
Pinot Blanc	Dümmen Orange	3

*Breeding lines represent 35 genotypes used in the breeding programme of Deliflor chrysanten B.V. They have no name, only a number. Of those, 11 genotypes were grown in three replicates, the rest in one replicate.

3.2.3   Bi-parental population

The offspring and parents were grown in a randomized block design with three replicates in three different seasons: summer (May to July 2015), late summer (August to October 2015), and autumn (September to November 2015). For the parents, three fields per flower trial were planted. Depending on the viability of the mother plants, fields consisted of 10 to 50 plants. Of those, up to six stems were used for experiments (see below). Plants were grown in 12, 12 and 14 days of 18, 21 and 21 hours photoperiods for the summer, late summer and autumn replicates, respectively. After that, they were grown in 11 and 12 hours photoperiods (depending on the season) to induce flowering. Since resources for carbohydrate measurements and vase life tests were limiting, we used a subset of 55 genotypes (including parents) for the carbohydrate measurements (here after referred to as 'subset55') and a subset of 145 offspring genotypes (including parents) for the vase life tests (here after referred to as 'subset145'; Figure 3.1). Subset55 was completely included in the subset145, and vase life of subset55 was assessed in triplicate. The remaining genotypes in subset145 were measured once or in two replicates. All offspring plants in subset55 and subset145 were harvest-ripe on the same day as the parents.
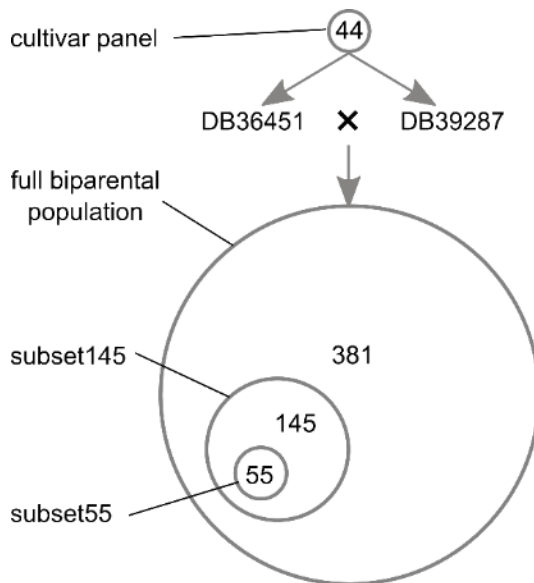
Figure 3.1. A schematic overview of used plant material. The x represents a cross between genotypes DB36451 and DB39287. The numbers in the circles describe the number of genotypes in the population and subsets. Circle size is relative to the number of genotypes. The parents (DB36451 and DB39287) are included in all populations.

3.2.4   Vase life tests

After harvest, five flower stems per replicate field were cut to 75 cm, wrapped in an unperforated plastic sleeve, and stored in a cardboard box at 4°C for 14 days. After this storage period, stems were trimmed to 65 cm. In order to normalize for water balance, the cultivar panel was placed on water at 4°C for three hours. For the bi-parental population our experiment focused on disk floret degreening only. As disk floret degreening is not related to water balance (van Geest et al., 2015), water balance was not normalized for the bi-parental population. For the vase life tests, stems were placed in a vase with 1L of tap water and 25 mg L^{-1} DICA at 20±1 °C, 60±10% air humidity and 12 h photoperiods at 6.5 µmol m^{-2} s^{-1} provided by Philips Coolwhite TL 84 lamps. During vase life testing, flower stems were assessed based on the quality parameters depicted in Table 3.2. Vase life was considered ended if the quality threshold of one of the vase life terminating symptoms was reached. For the cultivar panel, the number of days were recorded until quality dropped below the threshold for each problem separately until at least one for each of the flower-, stem- and leaf quality thresholds were reached. For the bi-parental population, only the number of days until disk floret degreening was reached was recorded.

3.2.5   Colour measurements of detached capitula

If chlorophyll and its precursors are lost, the intensity of green decreases and that of red increases (Schouten et al., 2002). This results in an increase in R/G and this measure was used to quantify loss of green colour in disk florets. For estimating degreening sensitivity, we adapted the AIDA method as described by Trivellini et al. (2012). With the AIDA method, penduncle ends of detached Arabidopsis inflorescences are placed in a medium contained by a microtiter plate, and degreening is assessed by eye or by chlorophyll quantification. We adapted this assay to fit the size and shape of chrysanthemum inflorescences and added a method to quantify degreening using colour measurements. After harvest, the ray florets were removed from the third capitulum as counted from the top and it was cut off leaving approximately 5 cm of the peduncle. As sepals were very large for some genotypes, for all genotypes sepals were cut off to prevent covering the disk florets. Capitula stem ends were placed in 25 mg L^{-1} dichlorisocyanurate (DICA) to prevent microbial growth. The detached capitula were kept in the dark at 20±1 °C and 60±10% air humidity for 14 days. Pictures were taken every working day

(Monday to Friday) in a standardized light environment using a Canon EOS 1000D camera. The red/green (R/G) values were measured as earlier described in chapter 2. Intensity of red and green of the disk florets was analyzed using ImageJ (Schneider et al., 2012) using the "RGB Stack" function, followed by measurement of the greyscale intensity.

3.2.6   Relative electrolyte leakage

Relative electrolyte leakage was measured to quantify the presence of senescence-like processes in the form of ammonium efflux and/or membrane disintegration (Rolny et al., 2011). This was done after the last colour measurement of the capitula at day 14 from the plants of the bi-parental population grown in summer. Disk florets were removed and placed on a sieve in a 175 mL plastic pot, in order to easily separate them from the water for measurements. Disk florets were washed for 20 minutes with 20 mL demineralized water by stirring. After that, the disk florets were rinsed again with 20 mL demineralized water and incubated in 20 mL demineralized water for two hours at 20°C. After this two-hour incubation, EC was measured of the solution (EC_{2h}), after which the solution including disk florets was heated in a microwave at 600 watts for two minutes to break down all membranes. After heating, the water and disk florets were left to cool down at 20°C for three hours and EC was measured again (EC_{total}). Relative electrolyte leakage was expressed as percentage, and calculated as $EC_{2h}/EC_{total}*100\%$.

Table 3.2. Vase life terminating symptoms and their quality thresholds. These quality thresholds are based on the appearance of a bunch of five flower stems placed in one vase.

Symptom	Quality threshold
Disk floret degreening	More than 50% of the capitula of at least three stems show degreening of the disk florets
Leaf browning	More than 50% of the leaves of at least three stems show browning
Leaf wilting	More than 50% of the leaves of at least three stems show wilted leaves
Bud browning	More than 50% of the flower buds of at least three stems show browning
Leaf yellowing	More than 50% of the leaves of at least three stems show yellow leaves
Ray floret browning	More than 50% of the capitula of at least three stems show browning of the ray florets
Ray floret wilting	More than 50% of the capitula of at least three stems show wilting of the ray florets
Stem wilting	More than three stems show wilting of the stem
Stem blackening	More than three stems show blackening of the stem
Multiple reasons	More than three stems shows at least one of the above
Bacterial infection vase water	Vase water is highly turbid

3.2.7 Carbohydrate measurements

Disk florets of the third capitulum (counted from the top) of one flowering stem per field were sampled at harvest. For the plants grown in summer, carbohydrates were measured using an ethanol-based extraction as described in chapter 2, of which the fructan quantification method is based on (Verspreet et al., 2012). Since the ethanol-based method is time and resource consuming, and because solubility of sugars in 80% ethanol might not be optimal (Raessler, 2011), we used a less resource intensive water-based extraction protocol for the late summer- and autumn-grown plants. In this method, enzymes are used to specifically break down polysaccharides like starch and fructans enzymatically. Fructans were hydrolyzed based on the AOAC method 999.03 (Megazyme Fructan Assay Procedure, Megazyme, Bray, Ireland). Disk florets were lyophilized, ground into fine powder and a subsample was weighed accurately between 12 and 20 mg. Two mL of 100 mg L^{-1} thermostable α-amylase (Rohalase; SERVA Electrophoresis GmbH, Heidelberg, Germany) was added and samples were incubated at 90°C for 30 minutes to extract sugars and hydrolyse alpha-linked polysaccharides. After that, 2 mL enzyme mixture was added, containing 2000 units L^{-1} fructanase mixture (purified powder, Megazyme, Bray, Ireland), 1 g L^{-1} bovine serum albumin (BSA), 0.5 g L^{-1} amyloglucosidase from *Aspergillus niger* (Sigma 10115, Sigma, St Louis, MO, USA) in a 200 mM acetate buffer at pH=4.6. The extract was incubated at 60°C for 30 minutes. After extraction and hydrolysis, it was centrifuged at 35,000 x G. The supernatant was diluted 10x and fructose and glucose concentration were measured using HPAEC-PAD as described in chapter 2. Successful enzymatic hydrolysis of fructans and starch was tested using internal standards of inulin and starch in a representative test sample. To check for compatibility between ethanol and water extraction, we measured total carbohydrate content of a set of 24 samples originating from 24 different genotypes with both methods, and found a correlation with an R^2 of 95% between the two methods, showing that results of both methods are comparable with the order among genotypes. However, absolute values could not be compared between methods, since the water-based method had a 15±2% higher yield compared to the ethanol method. This was probably due to the higher solubility of sugars in water compared to 80% ethanol (Raessler, 2011).

3.2.8 Statistical analysis

Data analysis was performed using R version 3.1.0 (R Core Team, 2014). The three planting dates of the bi-parental population were considered as blocks, and phenotypic data were corrected accordingly. Model fitting was conducted using the "lm" function in R. R^2 and p-values calculated using the summary function, and correlation coefficient (ρ) was calculated using the "cor" function. Broad sense heritability (h^2) was calculated using h^2=genotypic variance/total variance (Acquaah, 2007). Variances were estimated by an analysis of variance (ANOVA) of the effect of genotype on the trait.

3.3 Results

3.3.1 Postharvest performance of the cultivar panel

Vase life was defined as the number of days until the quality threshold of one of the vase life terminating symptoms (Table 3.2) was reached. Variation in vase life of the cultivar set ranged from 4 to 17 days (Figure 3.2). The reasons for ending vase life were diverse, but disk floret degreening was the most frequent first vase life problem, with 43 % of the genotypes reaching the quality threshold of disk floret degreening first.
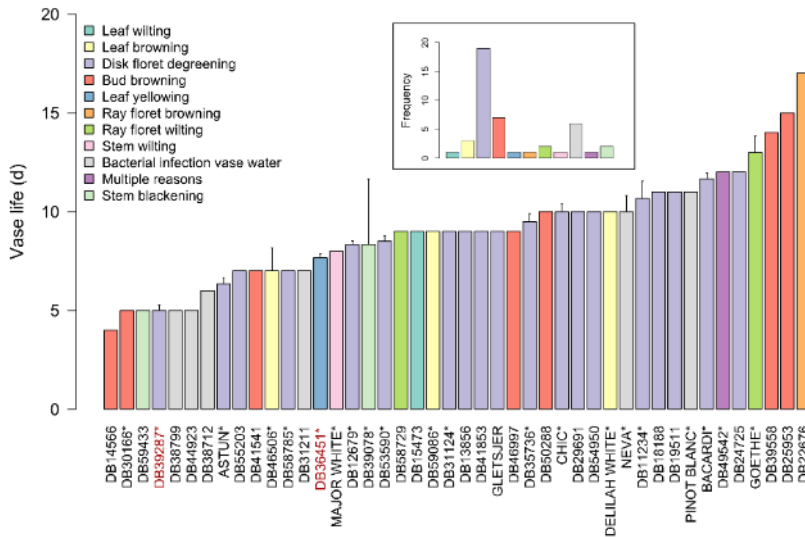
Figure 3.2. Average vase life of the cultivar panel following 14 days storage at 4°C. Vase life was assessed of bunches of five flower stems. Genotypes measured in three replicate bunches are indicated with an asterisk (*). Error bars indicate their standard error. Names of the parents of the bi-parental population are depicted in red. The colour represents the symptom for which average days until occurrence of that symptom was minimal. The inlay shows the frequency of genotypes per first occurring vase life terminating symptom.

For each replicate, disk floret colour was measured of detached capitula after harvest during 14 days. Colour is represented as the intensity of red divided by the intensity of green (R/G) of the disk florets. Genotypes that were most sensitive to degreening started to show loss of green colour (higher R/G ratios) after six to seven days (Figure 3.3). R/G values of insensitive genotypes hardly increased during the measuring period of 14 days. Green colour loss of detached capitula expressed as the R/G increase over 14 days correlated highly significantly with the time until disk floret degreening of flowering stems in the vase life test dropped below the quality threshold as specified in Table 3.2 (Figure 3.4 A). From the tested genotypes, two extremes (Figure 3.3; Figure 3.5A) were chosen as parents to generate a bi-parental population segregating for disk floret degreening.
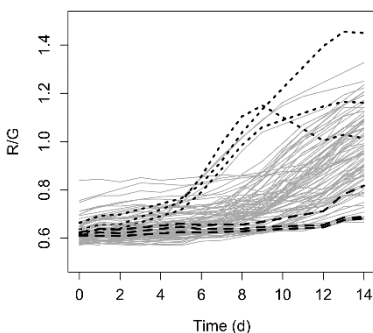

Figure 3.3. Red over green (R/G) values of detached capitula over time of the three replicates of the sensitive parent (DB39287; dotted), three replicates of the insensitive parent (DB36451; dashed) and members of the cultivar panel (grey).
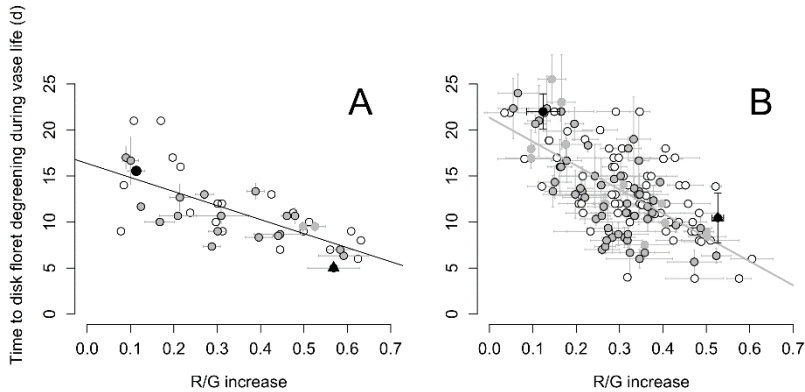
Figure 3.4. Correlation between R/G increase over 14 days of detached capitula and time to disk floret degreening during vase life for the cultivar panel (A) and population subset145 (B). Vase life was measured after 14 d dark storage (4°C) of flowering stems. The black triangles represent the sensitive parent of the bi-parental population, the black circles represent the insensitive parent. The open circles represent genotypes measured with one replicate, closed grey symbols represent genotypes measured with two replicates and closed grey symbols with black outlines represent genotypes measured with three replicates in the vase life tests. The regression line represents the fitted model: $y=-15.24x+16.36$ with an R^2 of 49.4%, a correlation coefficient (ρ) of -0.70 and p-value<0.0001 for (A) and $y=-26.0x+21.3$ with an R^2 of 44.8%, a ρ of -0.67 and p<0.0001 for (B).

3.3.2   Segregation for degreening sensitivity

The population was phenotyped three times for disk floret degreening using the method of detached capitula. Senescence-like processes co-occurred with loss of green colour as indicated by the correlation between R/G increase and relative electrolyte leakage (Figure 3.6; p<0.0001). The growing season had a significant effect on both R/G increase and time to disk floret degreening during vase life (Figure 3.7). Towards the time of year with less light availability, disk florets are more susceptible to degreening. In addition, carbohydrate content was affected by season. Carbohydrate content of the disk florets of summer-grown plants were measured using the ethanol-based method and carbohydrate content of late summer and autumn-grown plants were measured using the water based method (see Materials and Methods). As absolute values between the methods were different, these could not be compared. However the difference between late summer grown and autumn grown plants was highly significant (p<0.0001; not shown), with a higher carbohydrate content for disk florets of plants grown in late summer compared to plants grown in autumn. Season corrected averages of R/G increase segregated in a quantitative manner in the bi-parental population approaching a normal distribution (Figure 3.5B).
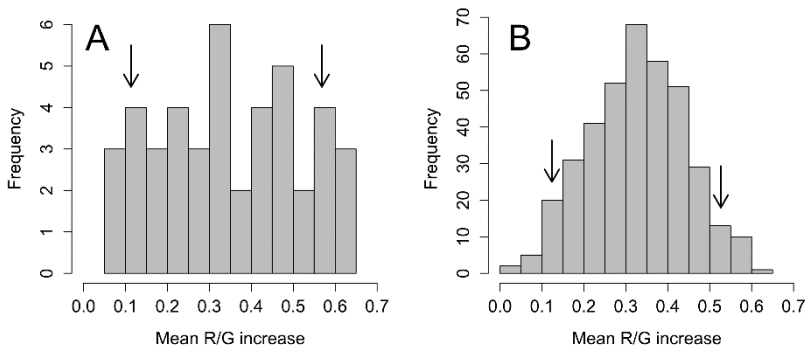
Figure 3.5. Histograms of mean R/G increase of capitula after 14 days in darkness for the cultivar panel (A) and full bi-parental population (B). The R/G values of the bi-parental population were season corrected. The means of the parents are indicated with an arrow (insensitive: left; sensitive: right).
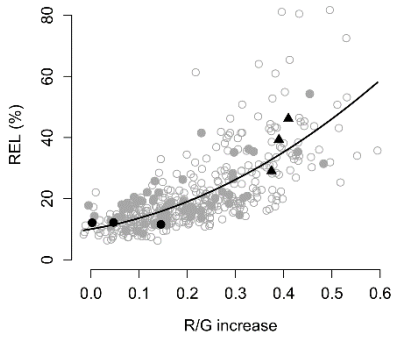
Figure 3.6. Relation between R/G increase of capitula of the full population and relative electrolyte leakage of disk florets (REL) after 14 days of darkness. Open grey circles represent offspring, filled grey circles those offspring for which carbohydrate content was assessed, black triangles the sensitive parent, and black circles the insensitive parent. The regression line represents the fitted model: $y=90.56x^2+26.87x+10.03$ with an R^2 of 53.4% and $p<0.0001$.

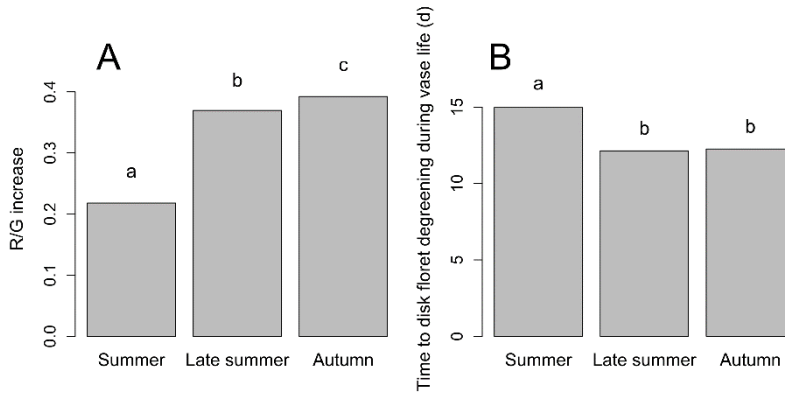Figure 3.7. Average R/G increase (left) and on time to disk floret degreening during vase life (right) of the population per season. Barplots indicate the average of all tested genotypes per season. Letters indicate significant differences at p=0.05 in a paired t-test between all three pairs of seasons.

Also for the progeny of the population, the R/G increase of detached capitula correlated (p <0.0001) with time to disk floret degreening during vase life tests of flowering stems following 14 days of storage (Figure 3.4 B). Carbohydrate content at harvest significantly explained variation in days to disk floret degreening by a simple linear regression model (p=0.002; not shown) and R/G increase (Figure 3.8; p<0.0001). Residuals increase with higher fitted R/G values (inlay Figure 3.8). Therefore, at low carbohydrate content less variation of R/G increase is explained. Genotypes with high carbohydrate content showed low susceptibility to disk floret degreening. However, low carbohydrate genotypes were not susceptible in all cases.
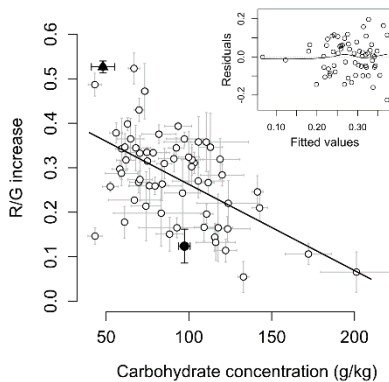

Figure 3.8. The relation between carbohydrate concentration of the disk florets at harvest and R/G increase after 14 days of darkness. Members of subset55 are depicted by open circles, the sensitive parent is represented by the black triangle, the insensitive parent by the black circle. Error bars represent standard errors, with n=3. The regression line represents the fitted model: y=-0.00194x+0.456 with an R^2 of 31.0%, a correlation coefficient (ρ) of -0.56 and p<0.0001. The inlay depicts the residuals versus fitted values of this model, with the line representing the floating average of the residuals.

3.3.3 Heritability of disk floret degreening

In order to estimate selection efficiency of measured traits, we calculated broad-sense heritability (Table 3.3). This is the fraction of variation explained by genotype compared to the total variation. The higher this fraction, the more successful one can select for that trait in breeding. Heritability of days to

disk floret degreening during vase life of flowering stems was clearly lower compared to R/G increase of detached capitula in this subset. When considering the entire population, heritability was even higher. Carbohydrate content was most explained by genotype, with a heritability of 0.755.

Table 3.3. Heritability parameters of the measured variables of subset55 and the full population.

Variable	Unit	n†	Mid-parent value	Grand mean‡	Heritability
Days to disk floret degreening during vase life[a]	days	55	16.2	13.1±5.64	0.59
R/G increase of subset55*	-	55	0.325	0.280 ±0.120	0.688
R/G increase	-	381	0.325	0.326±0.128	0.734
Carbohydrate content of subset55[a]	g kg^{-1}	55	72.7	89.6±34.2	0.755

*subset55: a subset of 55 genotypes out of the entire population that was harvest ripe at the same day as the parents (including parents; see Materials and Methods).
†number of genotypes
‡Mean±standard deviation

3.4   Discussion
Vase life in chrysanthemum can end because of a broad range of different visual symptoms of deterioration. We assessed vase life after a period of dark storage for 14 days at 4°C of the flowering stems of a set of genotypes. For twelve of these symptoms, each occurred at least once in the 44 single white genotypes we tested. However, disk floret degreening occurred most frequently supporting our focus to improve chrysanthemum postharvest performance by investigating this problem.

To investigate genotypic effects on disk floret degreening, we developed a method to screen a large number of plants efficiently in terms of time and costs. The quantification of degreening sensitivity expressed as R/G increase of dark-held detached capitula took much less climate room space. In addition, it also took less time compared to a standard vase life test after two weeks of cool storage, since results are available 14 days after harvest, whereas it takes up to five weeks with a standard vase life test (two weeks of storage included). As seen from the correlation analyses, R/G increase of detached capitula is a good estimator for disk floret degreening during vase life. The presence of stem, leaves, and ray florets apparently do not play a major part in explaining genotypic differences in disk floret degreening of flowering stems after storage. This shows that the status of the capitulum at harvest determines for a large part the postharvest performance. The use of detached capitula to estimate a genotypes' sensitivity to degreening during vase life after storage is therefore justified.

3.4.1   Degreening and carbohydrate metabolism
During postharvest dark-storage, plants and fresh produce often suffer from carbohydrate starvation. Carbohydrate starvation initiates a general catabolic state resulting in degreening concomitantly with senescence-like processes (Buchanan-Wollaston et al., 2005; Trivellini et al., 2012), eventually entirely devaluating crop quality. Postharvest treatments, pre-harvest growth conditions and genotype can reduce carbohydrate starvation during the postharvest phase. Postharvest treatments like sugar feeding often have a positive effect on vase life in general (Ichimura et al., 2005; Yakimova et al., 1996), and can prevent disk floret degreening (chapter 2). Disk floret degreening was more severe and occurred earlier if plants were grown during seasons with lower light intensities. As this co-occurred with lower carbohydrate content of the disk florets at harvest, this is probably the result of lower carbohydrate availability during these growth conditions. The carbohydrate measurements of disk florets at harvest show that large variation exists in carbohydrate content of disk florets between genotypes: up to a factor four difference. As carbohydrate content correlates with R/G increase, the genotype can have a

large effect on the severity of carbohydrate starvation and with that degreening sensitivity. Taken together, we show that variation in carbohydrate content, either varied by growth conditions, treatment or genotype, reduces disk floret degreening, underpinning the major role of carbohydrate starvation in disk floret degreening.

3.4.2   Disk floret degreening and the stay-green concept

Genetic variants with delayed loss of chlorophyll are generally referred to as stay-green phenotypes. Based on the co-occurrence of sensescence-like processes, they can be classified into functional and cosmetic stay-greens (Thomas and Howarth, 2000; Thomas and Ougham, 2014). Cosmetic stay-greens are genetic variants of which only the chlorophyll breakdown is hampered, whereas normal senescence processes continue. As degreening co-occurred with electrolyte leakage, at first sight it seems that the stay-green genotypes are not cosmetic. However, the increase in electrolyte leakage is likely an indication of ammonium efflux (Rolny et al., 2011), caused by chloroplastic protein breakdown and/or production of phenolic compounds out of aromatic amino acids (chapter 2). As retention of chlorophyll also leads to retention of thylakoid membrane associated proteins (Horie et al., 2009), increased electrolyte leakage would co-occur with chlorophyll breakdown and not necessarily with senescence. It is therefore unclear whether the genotypes with delayed degreening have a reduced chlorophyll catabolism or have general delayed senescence.

3.4.3   Possible mechanisms explaining genotypic differences

The correlation between R/G increase and carbohydrate content is significant, but a lot of variation exists around the regression line, especially at low carbohydrate content. This suggests that carbohydrate content at harvest is not the only factor affecting sensitivity to degreening. In addition, in three not directly related (more than third degree) genotypes presented in chapter 2, genotypic differences in carbohydrate content could not explain all differences in susceptibility. Apparently, having low carbohydrate content at harvest enhances susceptibility to degreening, but it is not necessary to be sensitive. Therefore, there seem to be two types of degreening insensitivity in the bi-parental population: carbohydrate dependent and independent. Carbohydrate independent insensitivity seems to be more interesting, since carbohydrates are generally limiting during growth. If these carbohydrates are not necessary for maintaining product quality after harvest, the crop can invest them elsewhere.

3.4.3.1   *Genotypic variation in carbohydrate content*

Genotypic variation in carbohydrate content in the disk florets could be caused by limiting factors in both sink and source strength during growth. Sink strength is defined as the competitive ability of an organ to attract assimilates (Marcelis, 1996). Genotypic variation in carbohydrate content of organs can be linked to variation in sink strength (Murray et al., 2008), and specifically to genes related to regulation of sink strength (Calviño et al., 2008; Nägele et al., 2010). Allelic variation in genes responsible for establishing sink strength could regulate carbohydrate content in disk florets, and breeding for higher sink strength in disk florets is therefore an option to reduce disk floret degreening. However, increasing source strength might be more profitable, since there would be no trade-off for other organs that would receive less carbohydrate if disk florets would receive more. The source strength could vary between genotypes because of genotypic variation in photosynthetic capacity, biomass conversion rate and plant architecture (Flood et al., 2011). A higher source strength usually results in formation of more dry matter partitioning towards the sinks, by the formation of more sink organs (Marcelis, 1996). Assuming the sink strength between the parents are comparable, source strength could indeed explain the differences between the parents, since the insensitive parent carries more capitula (15.2 ± 1.6) compared to the sensitive parent (10.4 ± 1.3) with similar flower size (van Geest, van Meeteren, Post, Visser and Arens, not published). This suggests that the sensitive parents' source is more limiting and therefore produces less capitula with less carbohydrate.

1.1.1   Genotypic variation in degreening independent of carbohydrate content

Variation in degreening sensitivity that is independent of carbohydrate status can be related to a wide

range of processes, ranging from aberrant signaling to variation of timing or extent of specific senescence-like processes. Since these processes are part of complex interactive networks between transcriptional regulation, hormones and reactive oxygen species (ROS) signaling, there are numerous possibilities for genetic variants with improved stay-green traits under stress (Thomas and Ougham, 2014).

Genotypic differences in transcriptional regulation of chlorophyll catabolism or carbohydrate signaling might play an important role in degreening sensitivity. As stated above, it is unclear whether the entire senescence process or only chlorophyll catabolism is delayed. Chrysanthemum phenotypes with delayed chlorophyll catabolism could be caused by variants of genes that result in stay-green phenotypes in Arabidopsis, like chlorophyll b reductases (NYC1 (Horie et al., 2009; Jibran et al., 2015) or NYC1-like (Horie et al., 2009)) or photosystem II associated STAY-GREEN (Sakuraba et al., 2012). Phenotypes with differential transcriptional regulation of overall senescence on the other hand, might be related to gene variants that have roles in carbohydrate signaling. Carbohydrate signaling is proposed to be regulated by the interplay between TOR (Target of rapamycin) and SnRK1 (SNF1-related kinase1) and their subunits (Smeekens et al., 2010). They regulate a wide range of senescence-associated processes including mitochondrial electron transport, glycolysis, amino acid metabolism, autophagy, photosynthesis and protein degradation (Baena-González et al., 2007; Díaz-Troya et al., 2008).

Genetic variants related to protection against oxidative stress could also play a role in delayed degreening at carbohydrate starvation. In chapter 2, we suggest that a more abrupt reduction of respiration rate could explain genotypic differences in sensitivity of degreening independent of carbohydrate content. This reduction in respiration rate could both be a cause or a symptom of oxidative stress. During oxidative stress, the mitochondrial electron transport chain could become overreduced, and this would result in a reduction in respiration rate. However, a too strong reduction in respiration rate as caused by e.g. active regulation of ATP synthase directly or indirectly by carbohydrate signaling could eventually also lead to oxidative stress (Vanlerberghe, 2013). Genotypes with improved protection against oxidative stress or alternative ROS signaling might therefore be less sensitive to carbohydrate starvation.

3.4.4   Heritability and selection for postharvest performance

As genotypic improvement through selection is more efficient with highly heritable traits, we compared heritability values between the measured parameters. Days to disk floret degreening during vase life had lowest heritability, indicating that a large part of the phenotypic variation is caused by uncontrolled environmental variation and measurement errors. Carbohydrate content was most heritable, and is therefore suitable to increase through selection. However, it explains only a small part of the number of days until disk floret degreening during vase life. As heritability values of R/G measurements were only slightly lower, selection at the level of R/G measurements seems to be most efficient to improve postharvest performance in terms of susceptibility to disk floret degreening, since it has higher heritability compared to days to degreening during vase life, and the measurement takes much less resources.

3.5   Conclusions

Although also relevant for crops and plants in nature, genotypic variation in sensitivity to carbohydrate limiting environments is a relatively unexplored research area. Finding alleles that explain this variation can allow us to identify plant strategies to cope with unfavorable light limiting conditions. In order to understand genetics of degreening of chrysanthemum disk florets after carbohydrate starvation, we developed a high-throughput phenotyping method. The moderately high heritability of sensitivity to disk floret degreening makes it suitable for mapping of quantitative trait loci (QTL). Since at least two processes regulate sensitivity to degreening, QTL mapping probably will result in multiple loci. More knowledge on the processes, specifically on sensitivity to carbohydrate starvation, will further aid

selection of phenotypes that can maximize the use of carbohydrates under carbohydrate-limiting conditions.

3.6   Acknowledgements

Chapter 4

Conclusive evidence for hexasomic inheritance in chrysanthemum based on analysis of a 183k SNP array

Geert van Geest, Roeland E Voorrips, Danny Esselink, Aike Post, Richard GF Visser, Paul Arens

Abstract

Cultivated chrysanthemum is an outcrossing hexaploid (2n = 6x = 54) with a disputed mode of inheritance. In this chapter, we present a SNP selection pipeline that was used to design an Affymetrix Axiom® array with 183k SNPs from RNA-seq data (1). With this array, we genotyped four bi-parental populations (with sizes of 405, 53, 76 and 37 offspring plants respectively), and a cultivar panel of 63 genotypes. Further, we present a method for dosage scoring in hexaploids from signal intensities of the array based on mixture models (2) and validation of selection steps in the SNP selection pipeline (3). The resulting genotypic data is used to draw conclusions on the mode of inheritance in chrysanthemum (4), and to make an inference on allelic expression bias (5). With use of the mixture model approach, we successfully called the dosage of 73,936 out of 183,130 SNPs (40.4%) that segregated in any of the bi-parental populations. To investigate the mode of inheritance, we analysed markers that segregated in the large bi-parental population (n = 405). Analysis of segregation of duplex x nulliplex SNPs resulted in evidence for genome-wide hexasomic inheritance. This evidence was substantiated by the absence of strong linkage between markers in repulsion, which indicated absence of full disomic inheritance. We present the success rate of SNP discovery out of RNA-seq data as affected by different selection steps, among which SNP coverage over genotypes and use of different types of sequence read mapping software. Genomic dosage highly correlated with relative allele coverage from the RNA-seq data, indicating that most alleles are expressed according to their genomic dosage. The large population, genotyped with a very large number of markers, is a unique framework for extensive genetic analyses in hexaploid chrysanthemum. As starting point, we show conclusive evidence for genome-wide hexasomic inheritance.

4.1 Introduction

The ability to genotype large numbers of polymorphisms is of major importance for breeding and genetic analysis. Costs for detection and genotyping of a large number of polymorphisms are still decreasing, and therefore become available to an increasing number of agriculturally important plant species, including polyploids. Genetic analysis in polyploids is less straightforward compared to diploids. An example is cultivated chrysanthemum, which is an outcrossing hexaploid (2n = 6x = 54) and has been classified as a segmental allopolyploid (Klie et al., 2014).

Polymorphism detection in species without a reference genome is restricted to methods using a reduced representation of the genome, like restriction enzyme based selection methods (e.g. RADseq or GBS (Baird et al., 2008; Elshire et al., 2011)), bait capture (Saintenac et al., 2011; Uitdewilligen et al., 2013) or RNA sequencing (RNA-seq; e.g. (Koning-Boucoiran et al., 2015; S. Wang et al., 2014)). RNA-seq is particularly useful for polymorphism detection for multiple reasons. First, discovered polymorphisms are in genic regions. Therefore, they have a high chance to represent or to be close to polymorphisms causative for an investigated phenotype. Secondly, lower SNP densities are expected in expressed sequences, which is an advantage in highly heterozygous polyploid species, as polymorphisms in flanking regions interact with marker assays. Thirdly, markers are in regions with transcribed genes, which generally have high recombination rates (Choi and Henderson, 2015), and discovered markers are therefore particularly useful for linkage mapping. Lastly, RNA-seq gives a representation of the transcriptome that helps building resources useful for other analyses.

A disadvantage of the use of RNA-seq is possible discordance between the expression of an allele and the allele's dosage in genomic DNA. Expression of certain alleles might even be completely absent. This feature is also referred to as allelic expression bias, and specifically occurs in allopolyploids (Chen, 2007). In hexaploid wheat for example, most expressed genes present in all subgenomes show expression bias towards one of the subgenomes, and expression is lost in at least one of the subgenomes for most genes (Leach et al., 2014). Another challenge for the use of RNA-seq for polymorphism detection is the *de novo* assembly of raw reads. Multiple splice variants in the transcriptome can represent one gene locus, which makes reconstruction of a locus on the genome challenging. For outcrossing polyploids like chrysanthemum, high heterozygosity (Osborn et al., 2003;

Uitdewilligen et al., 2013) is another challenge. Large variation between alleles makes it difficult to distinguish alleles from homologous genes. This will result in false polymorphism calls if gene homologues are assembled together in one contig or in the inability to detect polymorphisms if alleles are assembled into different contigs (Shahin et al., 2012).

High throughput genotyping of SNP polymorphisms using SNP assays is common in diploids (Chagné et al., 2012; Ganal et al., 2011). Several applications in the allopolyploids wheat and strawberry and in the tetraploids potato and rose have been published (Bassil et al., 2015; Koning-Boucoiran et al., 2015; Vos et al., 2015; S. Wang et al., 2014). SNP assays like Illumina Infinium, Affymetrix Axiom and LGC KASP provide signal intensities for each of the two allelic probes based on fluorescence. Allele dosage can be deduced from clusters, which can be visualised by plotting the two allelic signal intensities against each other. For diploid SNPs, including subgenome specific SNPs in polyploids with disomic inheritance, three clusters are expected: two homozygous and one heterozygous. SNPs in polyploids with polysomic inheritance, and SNPs polymorphic in more than one subgenome in disomic polyploids, show at maximum five clusters for tetraploids and seven for hexaploids. The assignment of a dosage to such clusters can be challenging, as clusters are often shifted (Voorrips et al., 2011), and resolution of the assay might not suffice (Grandke et al., 2016). High heterozygosity aggravates these issues, as undetected adjacent polymorphisms can influence a SNP assay.

The mode of inheritance has large implications for genetic analysis in polyploid organisms. In general, the mode of inheritance relates to the origin of ploidy of the organism: whether it is allopolyploid or autopolyploid. Disomic inheritance is usually a feature of allopolyploids and polysomic inheritance of autopolyploids. Ramsey and Schemske (1998) defined allopolyploids as polyploid organisms that have originated from interspecific hybridization in which genomes of the progenitors are retained, and autopolyploids as organisms originated from within a single species, often as a result of unreduced gametes. Since these definitions address origin, but not the mode of inheritance, allopolyploids not necessarily have disomic inheritance, and autopolyploids do not necessarily have full polysomic inheritance (Doyle and Sherman-Broyles, 2016; Ramsey and Schemske, 2002). Also intermediate modes of inheritance exist in several polyploid organisms (Allendorf and Danzmann, 1997; Leal-Bertioli et al., 2015; Nguepjop et al., 2016; Stift et al., 2008). As the most widely used definitions of allopolyploidy and autopolyploidy are on the origin, and not on the mode of inheritance, in this paper we only aim to separate disomic from polysomic inheritance.

The mode of inheritance in chrysanthemum has been under discussion (De Backer, 2012; Klie et al., 2014). Cultivated chrysanthemum is generally assumed to have originated from multiple species, which would make it an allopolyploid (Anderson, 2007; Dai et al., 2005). However, evidence is scarce. Despite the presumed allopolyploidy, there is evidence for polysomic inheritance in chrysanthemum. Cytological studies of cultivated chrysanthemum report presence of multivalents during meiosis, although most chromosomes pair as bivalents (Chen et al., 2009; Roxas et al., 1995). Multivalents will lead to recombination between all pairing homologous chromosomes and therefore indicate polysomic inheritance. The relatively high number of bivalents is not necessarily an indication of prevalence for disomic inheritance, as bivalents could represent a pairing event between any of the homologues chromosomes (Stebbins, 1940), and bivalent formation is known to be under genetic control in chrysanthemum (Watanabe, 1983).

In addition to cytological observations, hexasomic inheritance is also suggested by the analysis of segregation of molecular markers. Two studies showed that alleles from a single multi-allelic SSR marker have independent assortment, which is only possible with hexasomic inheritance (Klie et al., 2014; Park et al., 2015). Another strong line of evidence for polysomic inheritance is from the earlier work of Langton (Langton, 1989) on the inheritance of a flower colour trait regulated by a single dominant allele. In the study, a self-compatible simplex (dosage of one) individual is selfed, and duplex (dosage of two) progeny is crossed with nulliplex (dosage of zero) genotypes. In the resulting populations, the trait segregates in ratios as expected from hexasomic inheritance, whereas no

segregation would be expected if inheritance was disomic. These results provide strong evidence for hexasomic inheritance. The observations on SSR markers and flower colour are based on a few loci; other locations on the genome might show disomic inheritance. In order to acquire a genome-wide overview of the mode of inheritance, segregation analysis of a large number of markers distributed over the entire genome is required.

Multi-allelic SSR markers are scarce, and self-compatibility is difficult to obtain in chrysanthemum. However, analysis of segregation of high numbers of SNP markers in large outcrossing F1 populations can also provide evidence for the mode of inheritance. One of such analyses involves segregation of markers that are duplex in one parent and nulliplex in the other. If inheritance is disomic and the duplex alleles are on the same subgenome, all progeny will be simplex (one). Existence of these non-segregating duplex x nulliplex (2x0) markers therefore indicates disomic inheritance. If the two alleles are on different subgenomes, disomic inheritance will lead to a 1:2:1 segregation of the dosages 0, 1 and 2. Hexasomic inheritance will lead to 1:3:1 segregation in all cases. Studies that analysed deviations from those types of segregation, in general found duplex markers both fitting hexasomic inheritance as well as disomic inheritance (De Backer, 2012; De Jong and Rademaker, 1986; Klie et al., 2014; Park et al., 2015). Particularly in small populations genotyped with dominant markers, these tests are not powerful, because the segregation distributions (3:1 versus 4:1) are close to each other. Testing for segregation of a large number of markers in a large population with co-dominant markers, probably leads to less ambiguous conclusions.

A third method for estimation of the mode of inheritance is analysis of repulsion linkage (Wu et al., 1992). Estimates of recombination frequencies (r) between markers in repulsion that approach zero point to disomic inheritance. In the case of hexasomic inheritance, pairing should be random with all pairs of homologues chromosomes. In that case, the minimum r of markers in repulsion should be 0.4 (Qu and Hancock, 2001). In chrysanthemum, earlier analysis of repulsion linkage pointed towards hexasomic inheritance (Klie et al., 2014).

In this paper, we present a SNP selection pipeline for chrysanthemum from RNA-seq data (1), a method for dosage scoring in hexaploids from bi-allelic probe fluorescence (2) and validation of selection steps in the SNP selection pipeline (3). The resulting genotypic data is used to draw conclusions on the mode of inheritance in chrysanthemum (4) and allelic expression bias (5).

4.2   Methods

4.2.1   Plant materials

A panel of thirteen genotypes was selected for RNA-seq (Additional file 4.1). We aimed to represent an as broad as possible genetic variation in cut chrysanthemums. Selection was based on flower type, growth habit and absence of less than third degree relationships. The genotypes genotyped with the Axiom array consisted of a bi-parental population of 405 progeny of which the parents were included in the RNA-seq panel (POP1), three bi-parental populations consisting of 53 (POP2), 76 (POP3) and 37 (POP4) progeny respectively (Table 4.1), and a cultivar panel consisting of 63 genotypes.

Table 4.1. Overview of parents used for the bi-parental populations.

	female parent	male parent	size
POP1	DB36451*	DB39287*	405
POP2	DB41234*	DB40360	53
POP3	DB9656*	DB9541*	76
POP4	DB32141	DB39287*	37

* Also in the RNA-seq panel

4.2.2   Sample preparation RNA-seq

Plants were grown according to commercial growing standards in a greenhouse in Maasdijk, the Netherlands. Short photoperiods of 11 hour followed longer photoperiods of 14 hours to induce flowering. To get coverage for different transcriptomes, and therefore more different transcripts, samples were taken from five different combinations of environments, time points and tissues (Additional file 4.7). Samples were ground and approximately 100 mg was used for a 1 mL Trizol extraction according to the manufacturers' protocol (ThermoFisher Scientific, Waltham, MA, USA). After extraction, RNA concentration was estimated using a Nanodrop spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA), and RNA integrity was checked by electrophoresis on a 1.5% agarose gel. Samples were pooled per cultivar in equal ratios. After that, RNA's were cleaned up using a Qiagen RNeasy column according to manufacturers' protocol (Qiagen, Venlo, the Netherlands).

4.2.3   RNA Sequencing

Library preparation and sequencing was carried out by GenomeScan B.V. (Leiden, the Netherlands). Library preparation was done using the strand-specific NEBNext Ultra Directional RNA Library Prep kit for Illumina sequencing (New England Biolabs, Ipswich, MA, USA). In short, messenger RNA was isolated from total RNA using oligo-dT magnetic beads. Then, mRNA was fragmented and reverse transcribed into cDNA. The cDNA was ligated to sequencing adapters and amplified using PCR. Fragment size selection was between 400 and 900 basepairs. Clustering and DNA sequencing was performed using the Illumina cBot and HiSeq 2500 (Illumina, San Diego, CA, USA) according to manufacturers' protocols, resulting in paired-end reads with a length of 126 basepairs at both ends.

4.2.4   Quality filtering, assembly and mapping

Trimmomatic (Bolger et al., 2014) (v0.33) was used for quality trimming and filtering. Adapter sequences were removed using the ILLUMINACLIP option. Because of lower sequence quality in the leading and trailing ends of the reads, three basepairs were trimmed off at both ends using the LEADING and TRAILING options. Low quality regions were trimmed off using the SLIDINGWINDOW option with a window of four basepairs and a minimum quality of 15. Read pairs were discarded if less than 70 bp remained in one or both reads.

The two parents of POP1 were assembled separately. Trinity (Haas et al., 2013) version 2.0.6 was used for assembly using a minimum k-mer coverage of 2. Other settings were set at default values. Based on different quality criteria as described in the results section, the assembly of the female parent was selected as reference transcriptome. To reduce contig redundancy, the transcriptome was clustered using uclust (Edgar, 2010) at 99% similarity. Samples were mapped to the reference transcriptome using Bowtie2 (Langmead and Salzberg, 2012) and bwa-mem (Li and Durbin, 2010). For Bowtie2 (v2.1.0) the --very-sensitive option used and the options -3 and -5 were set to 5 in order to reduce the effect of error-prone read ends on the mapping. For bwa-mem (v 0.7.12), default options were used except for setting the –M option for Picard compatibility. Read duplicates were marked using Picard (http://broadinstitute.github.io/picard/), and duplicates and reads with a map quality smaller than 2 were removed using samtools (Li et al., 2009).

4.2.5   SNP calling and filtering

Our aim was to identify both SNPs that can be used for a wide range of chrysanthemum genotypes and to identify SNPs for linkage mapping in POP1 specifically. To achieve this, we aligned reads of all 13 cultivars (ALL) and reads originating from only the two parents of POP1 (PAR) separately. In addition, we wanted to include SNPs that were called with two different types of mapping software (bowtie2 and bwa-mem). Therefore, four alignment files were created: the reads of the ALL set aligned with bowtie2 and bwa-mem and the reads of the PAR set aligned with bowtie2 and bwa-mem (Additional file 4.8). SNPs were called using QualitySNP (Nijveen et al., 2013) from these four files separately. To reduce the number of false positives and rare SNPs, the option minimalNumberOfReadsPerAlleleP was set to 0.08 for the ALL call and 0.04 for the PAR call. The flanks were set at 35 bp and maxNumberOfSNPsInFlanks

was set to 1. A list with SNP flanking sequences of 35 bp at both sides was exported using the QualitySNP GUI.

We continued with SNP filtering by use of custom made R (R Core Team, 2014) (v3.1) scripts. All SNPs called from one type of mapping software were combined. From the "variations" output file of QualitySNP, the number of reads for each SNP allele was extracted. SNP-cultivar combinations with a total coverage greater than twelve were used to estimate the zygosity of the cultivars; whether it was homozygous or heterozygous. For each SNP in each genotype the relative coverage of the minor allele was calculated as the fraction of the coverage of the minor allele compared to the total coverage. Genotypes with a relative allele coverage smaller than 0.005 or greater than 0.995 were assigned homozygous and heterozygous otherwise. To select against ambiguous SNP calls, the assigned zygosity (heterozygous or homozygous) were used to filter out groups of SNPs that had the same flanking sequences, originated from different contigs, and showed different zygosities in any of the cultivars.

We selected against SNPs that were detected as heterozygous in all genotypes, as they have a large chance of being false positives, or subgenome defining. False positive SNPs can originate from mapping reads originating from two or more recently duplicated genes to one locus, and would appear as a heterozygous SNP in all genotypes (Figure 4.1A). Subgenome specific SNPs that are homozygous within the subgenome would also appear heterozygous in all genotypes (Figure 4.1B). In both cases, these SNPs are not interesting for linkage mapping, QTL analysis and association studies, as they will not segregate in a population. In order to select against them, only markers were kept that gave at least one heterozygous and one homozygous called genotype in the panel. We did not use this selection for SNPs originating from the alignments of PAR set since SNPs heterozygous in both parents are very informative for linkage mapping, and we did not want to deplete our dataset for these SNP types.
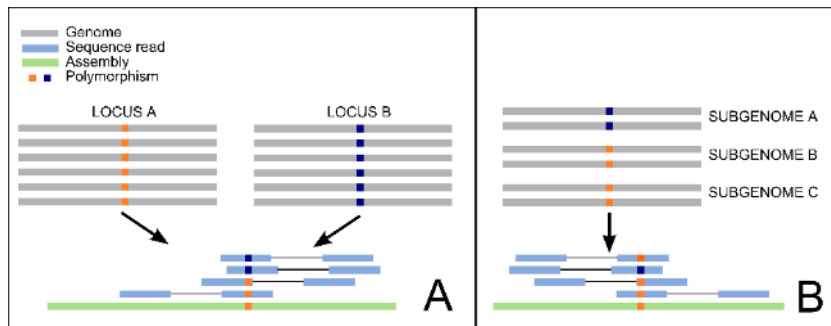

Figure 4.1. Examples of false positive SNPs against which was selected. In (A) two recently duplicated loci have SNPs between each other but not within. In (B) a SNP that is homozygous and specific for a subgenome. In both situations, all genotypes appear to be heterozygous and SNPs will not segregate in a population.

After selection against SNPs heterozygous in all genotypes, flanking sequences including the reference SNP allele were aligned to the reference transcriptome using BLAST with an e-value cut-off of 1e-5. Contigs assembled by Trinity are classified into different hierarchical categories (Haas et al., 2013). Markers with a hit in different groups of contigs as separated by the highest hierarchical category (component) or without a 100% hit including the reference allele were discarded. After that, SNPs selected from the two types of alignment software were taken together, and duplicates were removed. Further filtering by the Affymetrix bioinformatics team resulted in a set of 183,130 SNPs of which 34,068 could be tiled from both directions.

4.2.6   Genotyping with Axiom array
An Axiom genotyping array was designed by Affymetrix. We expected resolution to be an issue for high confidence dosage estimation (Grandke et al., 2016), since for a hexaploid a maximum of seven

genotype clusters can be expected instead of the regular three in a diploid and five in a tetraploid. Therefore, each probe was tiled four times instead of the regular two times. Genotyping was performed by Cigene (Ås, Norway). The Affymetrix bioinformatics team postprocessed raw signals of each of the four probe replicates into normalized signal intensities per probe.

4.2.7   Dosage scoring and quality filtering

Marker dosage was called from ratios of signal intensities of the genotyping array using a modified version of fitTetra (Voorrips et al., 2011), that allowed dosage scoring in other ploidy levels than tetraploid. We refer to it as fitPoly. The option p.threshold was set to 0.97. After running fitPoly, a composite quality score was calculated based on conflicts between segregation and parental dosage (allowing for both hexasomic and disomic inheritance), conflicts between scores assigned to replicate samples of the parents, missing parental scores and number of missing values. A threshold for this composite score was determined upon visual inspection of the signal intensities and scored dosages per marker. All markers below this threshold were filtered out. Probes of markers that were tiled from both sides were compared. Probes from one SNP locus with less than 4% different dosage scores were merged into one marker. Others were kept in the dataset as separate markers.

Two individuals of POP1 with more than 5% unexpected dosages based on parental dosages were removed. Then, marker dosages were converted to their most fundamental form, as described by Bourke et al. (Bourke et al., 2016). After that, markers and individuals with more than 10% missing values were removed. Markers were considered duplicates if all their non-missing dosages were equal. Duplicate markers were grouped and the marker with the least missing values was kept in the dataset as representative. Skewedness of simplex x nulliplex (1x0) markers was quantified using the probability of a X^2 test assuming a 1:1 segregation. Probabilities were corrected for multiple testing (referred to as q-values) and markers with q < 0.01 were removed. After filtering, the dataset consisted of two parents and 398 F1 progeny genotyped with 28,485 markers.

After analysis of segregation of markers in POP1, it became clear that preferential pairing was absent. We re-ran the pipeline without filtering for duplicate markers on all bi-parental populations assuming hexasomic inheritance. In this case we also supplied fitPoly with information on the population each sample belonged to (four F1 populations, their parents, or the cultivar panel), which allowed fitPoly to use the expected segregation ratios to improve model fitting (Zych et al, submitted). To investigate repulsion linkage in POP3, and to investigate the marker distribution of POP1 compared to the first run, we repeated the post fitPoly processing of markers to those populations as described for the first run.

4.2.8   Pedigree simulation

In order to estimate the expected distribution of repulsion linkages with known modes of inheritance, we simulated F1 populations of 400 individuals each. We used PedigreeSim V2.0 (Voorrips and Maliepaard, 2012) with 900 simplex x nulliplex (1x0) markers randomly placed on each of the 9 chromosomes. All chromosomes had a length of 100 cM, the centromeres were positioned at 50 cM. Hexasomic and disomic inheritance were simulated by setting the prefPairing parameter at 0 and 1 respectively for each chromosome. For each of these situations one F1 population was simulated.

4.2.9   Linkage analysis and statistics

Recombination frequency and LOD were calculated as described in Van Ooijen and Jansen (2013). Statistical analysis, other calculations and plotting were performed with R (R Core Team, 2014). Venn diagrams were plotted with the R package VennDiagram (Chen and Boutros, 2011). Multiple testing correction was performed with the R package qvalue (Dabney and Storey, 2014).

4.3   Results

4.3.1   RNA sequencing, assembly, and alignment

RNA-seq resulted in an average of 100.4 M reads for the deep-sequenced parents of the large population (405 individuals, POP1) and on average 70.4 M reads for the 11 other sequenced cultivars

(Additional file 4.1). Sequence assembly resulted in 270,186 contigs for the female parent and 275,397 contigs for the male parent (Additional file 4.2). Clustering with uclust (Edgar, 2010) at 99% similarity reduced the number of contigs to 227,213 and 231,634 respectively. As the average contig length in the female parent was longer and total number of contigs was lower, the assembly of the female parent was considered as higher quality and therefore used as reference transcriptome. Mapping reads of all cultivars to this assembly using bwa-mem resulted in an average alignment rate of 88.6±0.9%, for bowtie2 this was 81.6±0.7%.

4.3.2   SNP filtering
In total 183,130 SNPs were included in the array. Of these, 106,844 originated from discovery in the full panel (ALL call). The other 76,286 SNPs were identified using data from only the parents of POP1 (PAR call), which were selected using less stringent filtering. Most SNPs (65.8%) could be identified from the alignment files of both mappers (Additional file 4.3).

4.3.3   Mode of inheritance in POP1
The first run of the SNP dosage scoring pipeline was performed to investigate segregation in POP1. Signal intensities of other genotypes were used to correctly estimate dosage. In the first run of the pipeline, we estimated dosage of 28,485 markers that were unique and segregated in POP1 (Figure 4.2). In the second run of the pipeline, we assumed hexasomic inheritance. The number of scored non-duplicate markers segregating in POP1 was very similar in this run (Additional file 4.4). Most of the markers had a dosage of 0 (nulliplex) in one parent and 1 (simplex) in the other. The paternal parent seemed to be more heterozygous compared to the maternal parent considering the higher number of nulliplex x simplex (0x1) markers and nulliplex x duplex (0x2) markers in this parent.
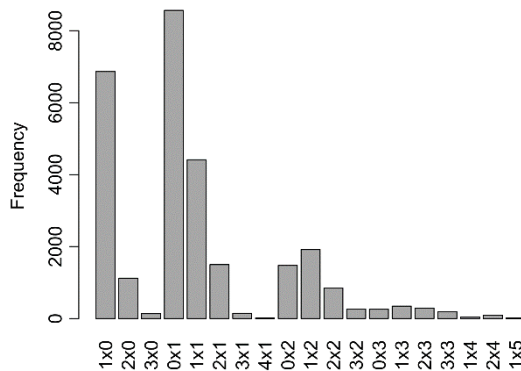

Figure 4.2. Distribution of marker types in POP1 after the first run of the pipeline (see Methods). Marker types are depicted as "dosage maternal parent" x "dosage paternal parent". Total number of markers: 28,485.

In order to investigate the mode of inheritance in POP1, we tested the segregation of all 2,597 2x0 markers. None of the 2x0 markers showed only simplex scores in the offspring. The markers were subsequently tested for goodness of fit to a 1:2:1 segregation as expected from disomic inheritance, or 1:3:1 segregation as expected from hexasomic inheritance. We used multiple testing corrected p-values, q-values, which resulted from a X^2 test of deviations from the two expected segregations. In general, X^2 tests having hexasomic segregation as null hypothesis had higher q-values compared to disomic segregation (Figure 4.3A), suggesting better fits to hexasomic segregation. For 1,938 out of 2,597 SNPs (74.6%) hexasomic inheritance was not rejected at q=0.01. For 323 SNPs (12.4%) disomic inheritance was not rejected, of which 153 were also not rejected for hexasomic inheretance. For 489 SNPs (18.8%) both segregation types were rejected, indicating skewed segregation or SNP scoring errors. On average, the frequencies in each genotypic class of all 2x0 markers, were more similar to hexasomic inheritance than to disomic inheritance (Figure 4.3B).
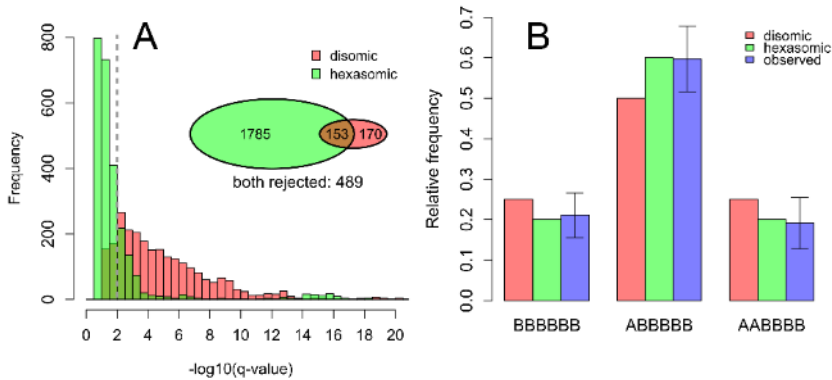
Figure 4.3. Histogram of q-values of all 2x0 markers tested for deviations of disomic and hexasomic segregation using a X^2 test (A). Inlay: venn diagram of 2x0 markers that did not reject disomic inheritance (red) and hexasomic inheritance (green) at q = 0.01 (dotted grey line in histogram). Total number of markers: 2,597; markers that were rejected for both types: 489. Barplot of expected genotype frequencies assuming disomic and hexasomic inheritance, and observed average frequency per class (B). Error bars indicate standard deviations.

To compare linkages in coupling and repulsion, we calculated r between all 1x0 markers of POP1 and POP3. For both POP1 and POP3, very large numbers of marker combinations were linked within a Haldane's distance of 8 cM ($r<0.074$), whereas none were linked in repulsion at that distance (Table 4.2). The minimum recombination frequency from all linkages in repulsion was 0.15 and 0.08 for POP1 and POP3 respectively. We compared the distribution of r for repulsion and coupling linkages from POP1 to two simulated datasets (Figure 4.4). The simulated datasets differed in preferential pairing; in the first dataset, we imposed full hexasomic inheritance, so random pairing of homologues and in the second full preferential pairing, so disomic inheritance (Figure 4.4B and C). In POP1, the distribution of r for linkages in repulsion tended towards higher values of r which is comparable to the simulated dataset in which we imposed hexasomic inheritance (Figure 4.4A).
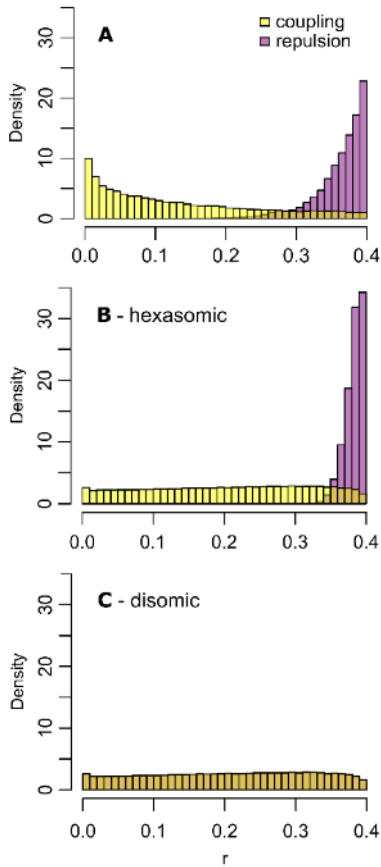
Figure 4.4. Distribution of recombination frequency between simplex x nulliplex markers in repulsion (purple) or in coupling phase (yellow) in POP1 (A), and in a simulated dataset where inheritance was completely hexasomic (B) and completely disomic (C).

Table 4.2. Statistics of comparison of repulsion and coupling linkages of markers segregating in POP1 and POP3

Population	Size*	1x0 markers	Minimum recombination frequency (r)		Linkages within 8 cM† ($r < 0.074$)	
			Coupling‡	Repulsion	Coupling	Repulsion
POP1	398	15,433	0	0.15	606,566	0
POP3	72	6,461	0	0.08	46,822	0

*After quality filtering of individuals (see methods section)
†Haldane's distance
‡Estimated phase: coupling or repulsion. Recombination frequency in repulsion was calculated as for disomic inheritance.

4.3.4   Genotyping array validation

We re-ran the SNP dosage scoring pipeline to estimate dosage for all genotyped individuals, which was different from the first run, in which we only aimed to estimate dosage for POP1. For this second run, hexasomic inheritance was assumed and the information on the distinction of the F1 populations and cultivar panel was used. In total, 73,936 markers (40.4%) could be called by fitPoly and had a missing value rate lower than 20%; those SNPs were considered as successfully discovered. Of those, 62,679 segregated as expected assuming hexasomic inheritance from the parental genotypes in at least one of the mapping populations and had a missing value rate lower than 10% (Additional file 4.5). These markers are suitable for genetic analyses that need high quality marker data, like linkage mapping.

In total 34,068 SNPs were tiled from both sides and therefore represented by two markers. Of those, 17,170 could be scored and segregated as expected in POP1 from at least one side. The two sides can be treated as two different marker assays, and we compared the results per SNP. For 55% (9,438) of those SNPs only one of both sides showed clear clustering (Additional file 4.6). Of the SNPs for which both probes showed clear clustering (7,549; 45%), 1.1% (183; 0.6% of total) did not correspond to each other.

Markers that were called from both the bowtie2 and bwa-mem alignment had a higher success rate than markers that were called with either one of the two types of mapping software alone (Figure 4.5). Markers called with bowtie2 had a slightly higher success rate than those called with bwa-mem.
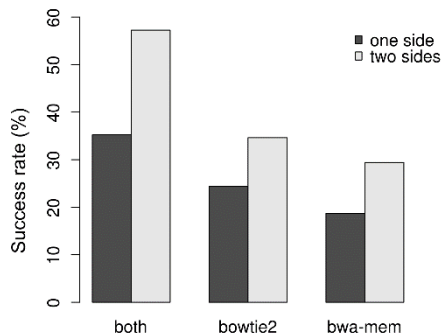

Figure 4.5. Percentage of segregating SNPs per class in which a SNP was discovered using alignment files of either type of mapping software, or one of the two specifically.

SNPs with a coverage higher than 100 as averaged over all thirteen sequenced genotypes had a substantially higher success rate compared to SNPs with lower average coverage (Figure 4.6A). In the ALL call, we selected only SNPs that were homozygous in at least one genotype. We assumed this would have a positive effect on the success rate. In the PAR call however, also SNPs were allowed that were heterozygous in both parental genotypes assessed. The comparison of the three groups (both heterozygous, only mother heterozygous or only father heterozygous) within the PAR call showed a clear positive effect on the success rate if one of the parents was heterozygous and the other homozygous (Figure 4.6B).
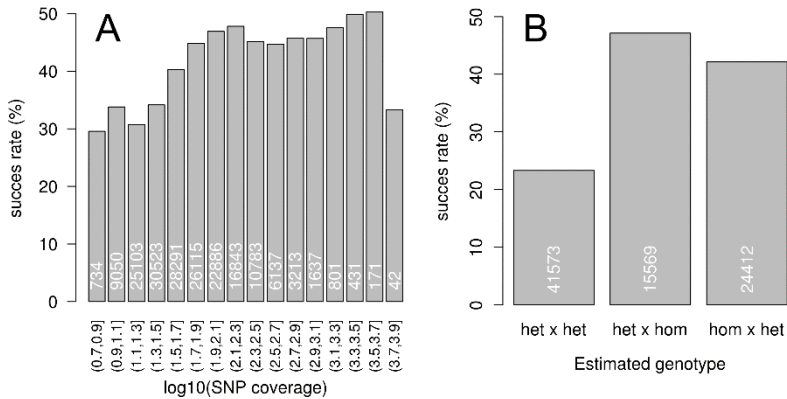
Figure 4.6. Success rate versus the average SNP coverage in the RNA-seq data per coverage interval (A). Success rate of SNPs from the PAR call versus the estimated genotype (heterozygous: het, or homozygous: hom) of the two parents of POP1 (B). The number of markers per class are written in white in the bars.

4.3.5   Allelic expression

For the SNPs that segregated as would be expected from the parental dosages in POP1, we compared the genomic dosage of the parents of POP1 with the relative allele coverage in the parents from the RNA-seq data (Figure 4.7). The average relative coverage per SNP allele per sequenced genotype matched the expected dosages. Distribution of relative coverage was more dispersed than expected from a binomial distribution, while the difference in dispersion between the binomial distribution and observed distribution was similar over dosages.
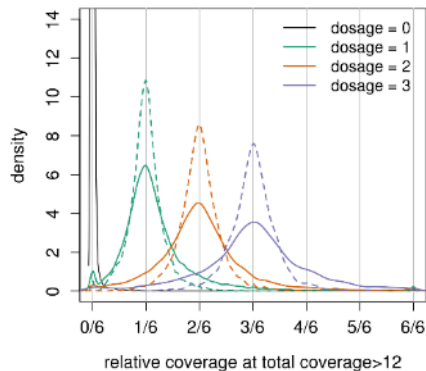

Figure 4.7. The density of relative coverage of RNA-seq data in each SNP class as genotyped by the Axiom array based on genomic DNA. Total number of analysed SNPs: 52,052. Solid lines represent the observed density and dashed lines represent expected density based on a binomial distribution. The figure represents data that originates from the parents of POP1. SNPs were filtered based on correspondence between parental dosages and observed segregation ratios and had a relative dosage greater than twelve.

4.4   Discussion

4.4.1   SNP filtering from RNA-seq data

Transcriptome assembly from short-read RNA-seq data of a heterozygous polyploid organism comes with challenges. One of those arises when trying to separate alleles from gene homologues (Shahin et

al., 2012). In sequence data from genomic DNA, unexpected variation in coverage and unexpected numbers of alleles per locus can be used to identify wrongly assembled contigs (McKinney et al., 2017). However, variation in coverage cannot be used with RNA-seq data, since expression varies strongly between genes. Detection of an unexpected number of alleles is difficult in a hexaploid, as the number of alleles per locus can vary between two and six. By selection against SNPs that are heterozygous in all genotyped cultivars, we tried to select against SNPs that were detected on an assembly of transcripts originating from two homologues loci. This seemed to have been successful, as SNPs from the PAR call that were heterozygous in both genotypes had a much lower success rate compared to SNPs from the same call for which one of the two parental genotypes was homozygous. The lower success rate of SNPs that are heterozygous in both parents of the PAR call could also be due to the higher chance of complex segregation ratios. For example, a duplex x duplex (2x2) SNP, would have an expected segregation ratio of 1:6:11:6:1 resulting in five clusters for the dosages 0, 1, 2, 3 and 4 respectively. Markers with more than two clusters have a higher chance of missing values or wrongly assigned dosages, as more clusters usually have a higher chance to overlap each other.

The use of two different types of mapping software resulted in three sets of SNPs: SNPs only identified with either bowtie2 or bwa-mem, and SNPs identified with both mappers. As there is some level of independence between the two SNP calls from both alignments, it would be expected that SNPs discovered with both mappers have a higher success rate (Clevenger et al., 2015). This was indeed the case, SNPs identified with both mappers had a much higher success rate (58% versus <35%).

4.4.2   Allelic expression

From the segregating SNP markers in our dataset, SNP allele read counts in the transcriptome conformed to their dosage estimated from the array data. Chrysanthemum therefore deviates from the allopolyploids wheat, cotton and camelina that show subgenome-specific expression patterns in a large number of expressed genes (Adams et al., 2003; Kagale et al., 2014; Leach et al., 2014; Zhang et al., 2015). In autopolyploids on the other hand, expression patterns are generally conform genomic dosage (Albertin et al., 2005; Church and Spaulding, 2009), like we observed in chrysanthemum.

Variation of relative allele coverage in the chrysanthemum read data was larger than would be expected from a binomial distribution. The source of this extra variation could be both biological and technical. The biological reason would be allelic expression imbalance of some loci, which is also common in diploids (Bell et al., 2013; Heap et al., 2009; Vogelstein et al., 2002). A technical reason could be allelic bias, caused by higher chance of alignment of reads that exactly match the reference allele compared to the alternative allele or scoring errors, but this should have been visible as deviations from expected relative dosages in our distributions.

4.4.3   The mode of inheritance

Our dataset gives evidence of complete or near-complete hexasomic inheritance in chrysanthemum. A first indication is the absence of non-segregating 2x0 markers. Presence of those type of markers would indicate disomic inheritance. Analysis of the segregation ratios of the 2,597 segregating 2x0 markers pointed towards hexasomic segregation. Only 6.5% of 2x0 markers were rejected for hexasomic segregation and not rejected for disomic segregation. It is likely that a large number of the markers fitting only disomic segregation had genotyping errors or skewed segregation, as the number of markers not fitting any of the two types (18.8%) was much higher.

Conforming to the analysis of 2x0 markers, comparison of linkages between 1x0 markers in coupling and repulsion phase also pointed towards absence of disomic inheritance in two populations (POP1 and POP3). There were no linkages in repulsion with a distance smaller than 8 cM, while a very large number of marker combinations in coupling were linked within this distance (606,566 and 46,822 for POP1 and POP3 respectively). This is contrary from what would be expected from disomic inheritance, since with disomic inheritance, the ratio between the number of linkages in coupling and in repulsion would be 1:1 (Wu et al., 1992), irrespective of the threshold of r. We did find linkages between markers

in repulsion in which r was lower than 0.4, the minimum expected recombination frequency in repulsion for full hexasomic inheritance (Qu and Hancock, 2001). In the simulated dataset in which we imposed full hexasomic inheritance, we also found linkages with r below the threshold. However, there were fewer, and minimum r was higher (0.29 compared to 0.15 and 0.08 for POP1 and POP3 respectively). The lower minimum recombination frequency in the real datasets could be caused by incomplete disomic inheritance (Qu and Hancock, 2001). However, genotyping errors or commonly occurring lethal allelic combinations leading to segregation distortions (Liu et al., 2010; Manrique-Carpintero et al., 2016) are also plausible reasons, and were not taken into account in the simulation.

Our conclusions deviate from the most recent and elaborate paper on the mode of inheritance in chrysanthemum authored by Klie et al (Klie et al., 2014). We conclude that inheritance in chrysanthemum is completely or nearly completely hexasomic, whereas Klie and colleagues conclude chrysanthemum to be a segmental allopolyploid (in which allopolyploidy is used in the context of mode of inheritance, so disomic). However, the analysis of segregation and repulsion linkage as reported by Klie et al. do not contradict our results. These authors also conclude that their results point to hexasomic inheritance. Despite this, they base their final conclusion on earlier suggestions of disomic inheritance that were wrongly based on observations of prevalence of bivalents during meiosis (Chen et al., 2009; Li et al., 2011). As reviewed in the background section, the observation of a predominance of bivalents is not an indication of disomic inheritance, as homologues can still pair with any other homologous chromosomes in a bivalent (Stebbins, 1940), and prevalence of formation of bivalents seems to be under genetic control in chrysanthemum (Watanabe, 1983). Based on earlier work done on the mode of inheritance of chrysanthemum that resulted in conclusive evidence (Klie et al., 2014; Langton, 1989; Park et al., 2015), and our results on two bi-parental populations, we suggest to classify cultivated chrysanthemum as a polyploid with hexasomic inheritance.

4.5   Conclusions

We present an application of the use of next generation sequencing and high-throughput genotyping in hexaploid chrysanthemum. Development of these resources opens up many possibilities for plant improvement at the level of the genome. As a first step, we were able to find conclusive evidence for near-complete hexasomic inheritance on a genome-wide scale. With these resources and the information on the mode of inheritance we will now be able to progress in the development of genomic tools for plant improvement in chrysanthemum, like linkage mapping and mapping of traits.

4.6   Acknowledgements

4.7   Additional files

Additional file 4.1. Overview of genotypes used for RNA-seq

Additional file 4.2. Statistics of assemblies of the two parents.

Additional file 4.3. Venn diagram of the number of SNP markers called with bowtie2 and bwa-mem.

Additional file 4.4. Distribution of marker types in POP1 after the second run of the pipeline (see Methods). Marker types are depicted as "dosage maternal parent" x "dosage paternal parent". Total number of markers: 27,902.

Additional file 4.5. Venn diagram of number of markers segregating in different bi-parental populations.

Additional file 4.6. Pie chart of comparisons of both SNP assays of markers tiled from both sides segregating in POP1. Corresponding: both assays gave less than 4% conflicting dosages. One missing: one of the assays could not be called by fitPoly. Not corresponding: two assays gave different results. Total number of SNPs: 17,170.

Additional file 4.7. Environment and tissue type of different samples.

Additional file 4.8. Overview of SNP calling steps. For more information, see materials and methods. QSNP: QualitySNP.

Chapter 5

polymapR – R package for linkage analysis and genetic map construction from F1 populations of outcrossing polyploids

Peter M. Bourke, Geert van Geest, Roeland E. Voorrips, Twan Kranenburg, Arwa Shahin, Richard G. F. Visser, Paul Arens, René Smulders, Chris Maliepaard

Abstract

New software tools are needed if we are to realise all the opportunities offered by modern genotyping platforms for genetic research and breeding in polyploid crops. Up to now, genetic analyses have often been performed using tools designed primarily for diploid species, which ignores the greater complexity of polyploid genetics. In this article we present polymapR, an R package for performing linkage analysis and integrated map construction from segregating bi-parental mapping populations in outcrossing autopolyploid species. This package can currently handle tetraploid and hexaploid marker datasets, and can thus be applied to a wide range of crops such as potato (*Solanum tuberosum*), rose (*Rosa hybrida*), leek (*Allium ampeloprasum*), alfalfa (*Medicago sativa*), blueberry (*Vaccinium corymbosum*), chrysanthemum (*Chrysanthemum x morifolium*), sweet potato (*Ipomoea batatas*) or kiwifruit (*Actinidia deliciosa*) for breeding and research purposes. It can detect and estimate the strength of preferential chromosome pairing, which can lead to differences in the expected segregation patterns of markers and ultimately introduce bias in the maps produced. This bias currently can be corrected for at the tetraploid level only. The polymapR package has been tested and validated on a number of F1 populations of potato, rose and chrysanthemum, using dosage-scored single nucleotide polymorphism (SNP) markers. With mapped marker numbers ranging from just under 7000 SNPs in potato to over 30,000 SNPs in chrysanthemum, the polymapR package can handle the sorts of high-density marker datasets that are currently being produced.

5.1   Introduction

In recent years there has been an acceleration of progress in the understanding of the genetics underlying important traits in autopolyploid plant species. This has been to a large extent due to developments in high-density genotyping platforms using single nucleotide polymorphism (SNP) markers which have found increasing application in polyploids. For example, high-density SNP arrays have been developed in potato (Felcher et al., 2012; Vos et al., 2015), rose (Koning-Boucoiran et al., 2015), alfalfa (Li et al., 2014) and chrysanthemum (this thesis, chapter 4), enhancing the scope for genetic studies in these species.

In polyploid species, as opposed to diploids, co-dominantly scored markers can possess multiple classes in the heterozygous condition, usually termed marker "dosage". For example, in a tetraploid there are five possible dosage classes of a bi-allelic SNP marker, namely nulliplex (AAAA) with a dosage 0, simplex (AAAB) with a dosage 1, duplex (AABB) with a dosage 2, triplex (ABBB) with a dosage 3 and quadruplex (BBBB) with a dosage 4. In a hexaploid, the number of dosage classes at a bi-allelic locus rises to seven. Various software have been developed to convert the signal from *e.g.* SNP arrays into these discrete dosage calls for polyploids, such as fitTetra (Voorrips et al., 2011), SuperMASSA (Serang et al., 2012) or ClusterCall (Schmitz Carley et al., 2017). Marker dosages have been shown to be useful for a variety of purposes, including linkage map construction (Bourke et al., 2017, 2016; Hackett et al., 2013; Vukosavljev et al., 2016), genome-wide association studies (Lindqvist et al., 2014; Rosyara et al., 2016), determination of identity-by-descent (IBD) probabilities (Hackett et al., 2013; Zheng et al., 2016), pedigree analysis (Endelman et al., 2017), population genetics (Vos et al., 2015), estimation of linkage disequilibrium decay (Vos et al., 2017) and quantitative trait locus (QTL) analysis in bi-parental populations (Hackett et al., 2014; Massa et al., 2015; Rak et al., 2017).

Genetic linkage maps have been traditionally used for both exploratory trait mapping (often termed QTL analysis) and the subsequent fine mapping of traits, as well as assisting genome assembly efforts by guiding the integration and orientation of contigs. High-density linkage maps may also improve our understanding of the chromosomal composition and genetics of polyploid species, uncovering such phenomena as double reduction (Bourke et al., 2015) or partially-preferential chromosome pairing (Bourke et al., 2017). In many polyploid species which lack reference genome sequences, linkage maps are also a (vital) first genomic description of that species.

Despite the importance of both linkage maps and polyploid crops, there are still relatively few software tools available for polyploid linkage map construction. Allopolyploid species showing disomic inheritance can be treated (genetically-speaking) as diploids, although it remains challenging to incorporate markers which are not sub-genome specific to bridge between these maps, as described for the case of allo-octoploid strawberry (van Dijk et al., 2014, 2012). With the remaining sub-genome specific markers, the full range of diploid mapping software is available. However, in the case of polyploid species which do not exhibit disomic inheritance (autopolyploids and segmental allopolyploids), the options available to the research community are limited.

Probably the most well-known autopolyploid mapping software is TetraploidMap (Hackett et al., 2007; Hackett and Luo, 2003), which has been used in numerous linkage analyses and QTL studies in autotetraploid species such as potato, alfalfa, rose and blueberry (e.g. Bradshaw et al., 2008; Gar et al., 2011; McCallum et al., 2016; Robins et al., 2008) since its release. Recently, its successor TetraploidSNPMap (TSNPM; Hackett et al., 2017) has been released to accommodate high-density marker data from SNP arrays, performing linkage mapping, IBD probability estimation and QTL analysis all within a single software module. However, despite its broad range of applications, it can only handle autotetraploid datasets and works exclusively on the Windows platform. Linkage studies in species exhibiting strong preferential chromosomal pairing or at different ploidy levels are currently not possible using this software.

An alternative polyploid mapping software is the PERGOLA package in R (Grandke et al., 2017). However, this software has been developed for use with F2 or backcross populations from homozygous parents only. In many cases, either due to inbreeding depression or the difficulties imposed by polysomic inheritance, developing inbred lines in autopolyploids is not possible or practical, and F1 populations from two heterozygous parents are used instead (a development based upon the two-way pseudo-testcross strategy (Grattapaglia and Sederoff, 1994), where only 1:1 segregating markers of such a cross were used).

In short, there is currently no software which can perform linkage mapping at various ploidy levels under a variety of inheritance models for outcrossing species using dosage-scored SNP marker data. Here we present polymapR, an R package for linkage mapping in outcrossing polyploid species which can generate linkage maps for polysomic tetraploids and hexaploids, accommodating either fully tetrasomic or mixed meiotic pairing behaviour (segmental allopolyploidy) at the tetraploid level. Its modularity will facilitate extensions to other ploidy levels (*e.g.* triploid populations from tetraploid and diploid parents) as well as its adaptation to other marker genotyping technologies in the future.

5.2   Description
The polymapR package consists of four parts – data inspection, linkage analysis, linkage group assignment and marker ordering, which are detailed below. A description of the functions within polymapR is described in the vignette which accompanies the package, going through all the steps in a typical mapping project. For consistency and simplicity, all examples mentioned here describe a tetraploid cross.

5.2.1   Data inspection, filtering and preparation for linkage analysis
The input data for polymapR is dosage-scored SNP marker data, which can be generated from a number of packages such as fitTetra (Voorrips et al., 2011), fitPoly (*in preparation*, Voorrips *et al.*) or ClusterCall (Schmitz Carley et al., 2017). Both fitTetra and ClusterCall are limited to tetraploid data whereas fitPoly can work over multiple ploidy levels. Regardless of how it is generated, the input dosage-scored marker data should consist of a column of marker dosages for the mother, one for the father followed by a column for each of the offspring of the F1 cross. Checks for marker skewness and shifted markers (when dosage scores are shifted by a fixed amount, which may occur in polyploid datasets but can be identified by discordance between parental and offspring scores) have been

incorporated in polymapR from a suite of tools developed for the fitTetra package (Voorrips et al., 2011).

After skewness checking, the first step in data preparation is the conversion of marker dosages to their simplest form, such that the sum of the parental dosage scores is minimised (Bourke et al., 2016). There are two possible conversions – a relabelling of the reference and alternative allele in both parents, or a single-parent relabelling if the other parent is homozygous. For example in tetraploids, a triplex x simplex (3 x 1, *ie.* having a maternal dosage of three and a paternal dosage of 1) marker may be converted to a 1 x 3 marker but not a 1 x 1 marker. Conversely, 3 x 0, 3 x 4 and 1 x 4 markers may all be converted to 1 x 0 due to the homozygosity of parent 2. Marker conversions are performed to reduce the number of marker segregation classes for the linkage analysis (which is directed according to the parental dosages), but have no effect on the pairwise results. In a tetraploid there are nine fundamental segregation types (Bourke et al., 2016), rising to nineteen for a hexaploid (chapter 6). Identifiable double reduction scores are preserved during conversion (*e.g.* a dosage of 0 from a 3 x 0 marker becomes a dosage of 2 in its converted form as a 1 x 0 marker), allowing an investigation of double reduction post-mapping. Any impossible scores (like a dosage of 3 or 4 from a 1 x 0 marker) are made missing. After mapping has been completed, the user has the option to return to the original marker coding in the phased map output files, to minimise mix-ups when tracking specific alleles (*e.g.* in a subsequent QTL analysis).

High-quality data facilitates the generation of high-quality maps. One indication of poor data quality is a high proportion of missing values. The user may choose to screen out markers or individuals with more than a desired rate of missing values (by default up to 10% is tolerated) or duplicate individuals (which may indicate possible mix-ups in genotype labelling) which can be identified through a pairwise correlation plot. Identical markers, which often occur in high-density marker datasets with limited population sizes, can be identified and reduced to one representative marker for the mapping steps, and reintegrated later. A principal component analysis (PCA) can also be performed and visualised, which may highlight some unwanted structure in the population (for example due to pollination from an unknown external pollen parent or from self-pollination) or outlying individuals. In such cases, PCA plots using markers which are homozygous for one of the parents only may provide additional information (see for example Vukosavljev *et al.* (2016)).

5.2.2 Linkage analysis

5.2.2.1 Linkage analysis under a polysomic model

In autopolyploid species with polysomic inheritance, it is possible to model meiotic pairing structures as random bivalents or multivalents. In practice, both pairing structures tend to occur, with a relatively low frequency of multivalents in stable autopolyploids (Bomblies et al., 2016; Santos et al., 2003). The main consequence of multivalent formation from a genetic perspective is the phenomenon of double reduction, where two segments of a particular homologue can end up in the same gamete and become transmitted together to F1 offspring. In previous studies it has been demonstrated that multivalents and double reduction introduce a small bias in the estimation of recombination frequency under a random bivalent model, but that this bias can safely be ignored if the rate of quadrivalent pairing is low (Bourke et al., 2016, 2015). Estimating the rate of multivalents can be achieved by a study of double reduction products (Bourke et al., 2015) or by predicting the meiotic pairing structures (Zheng et al., 2016), but is not necessary in practice.

Under a random bivalent model, there are three possible bivalent pairing conformations in a tetraploid. This rises to fifteen in a hexaploid. In general, for any even ploidy $p = 2n$ there are $c = \frac{(2n)!}{(2^n).n!}$ possible bivalent pairing conformations to be considered. Given any pair of marker loci, we model the unknown recombination frequency between them as r and consider the contribution of recombinant homologues with a within-bivalent probability of $\frac{1}{2}r$ and non-recombinant homologues with a within-bivalent

probability of $\frac{1}{2}(1-r)$. In cases where recombinants and non-recombinations cannot be distinguished, both are assigned an equal probability of $\frac{1}{2}$. Assuming the pairing is random, the probability of any particular pairing configuration is $\frac{1}{c}$ (in the case of preferential pairing, we introduce a preferential pairing factor to model deviations from randomness here).

The expected frequency of each offspring class n_{ij} ($0 \le i, j \le 2n$) is first summed over all c bivalent conformations:

$$E(n_{ij}) = \sum_{k=1}^{c} \frac{1}{c} f_k(r, 1-r)$$

where $f_k(r, 1-r)$ denotes a function of r and $1-r$, dependant on the marker combination considered. Given these expected frequencies, we relate them to the observed counts of individuals in each class $O(n_{ij})$ to yield the likelihood function $\mathcal{L}(r)$:

$$\mathcal{L}(r) \propto \prod_{i,j=0}^{2n} E(n_{ij})^{O(n_{ij})}$$

The likelihood equation results from equating the first derivative of the log of the likelihood function with zero:

$$\sum_{i,j=0}^{2n} \left(O(n_{ij}) . \sum_{k=1}^{c} \frac{1}{c} \frac{d}{dr} \ln(f(r, 1-r)) \right) = 0$$

In cases where no analytical solution exists, we use Brent's algorithm (Brent, 1973) to numerically maximise the log likelihood function in the bounded interval $0 \le r < 0.5$. For any pair of markers there are a number of possible phases between these markers to consider, which describe the physical linkage between marker alleles. In the case of a pair of 2 x 0 markers, these phases are termed "coupling", "mixed" and "repulsion" (Figure 5.1A). As the phase between markers is initially unknown, we must compute expressions for each of the possible phases, and select the most likely one as the phase for which $0 \le r < 0.5$ which maximises the log of the likelihood (Hackett et al., 2013). In the case of pairs of 1 x 1 markers in a tetraploid it was found that the phase which minimises r performs only marginally better (Bourke et al., 2016), but this is the only exception we have so far discovered.
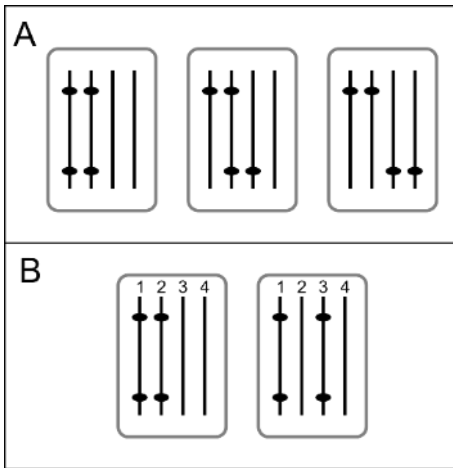
Figure 5.1. Example of phase considerations in a tetraploid. The three phases considered for a pair of 2 x 0 markers (A), from left to right, "coupling", "mixed" and "repulsion". In the case of preferential pairing between homologues 1-2 and 3-4 (B), we must consider two separate types of coupling phase, either coupling within preferential bivalents (left) or coupling between preferential bivalents (right). In the extreme case of an allotetraploid, this distinction could also be termed "subgenome-specific" versus "subgenome-straddling". In both scenarios, we follow the procedure outlined in the text to derive the likelihood function with parameters r and ρ.

Finally, we also compute the logarithm of odds (LOD) score, which is approximately related to the inverse of the standard deviation of the estimate. The LOD score provides a useful measure of the confidence in the estimate and is used for both marker clustering and marker ordering, where they (or their square) provide the weighting structure for the weighted metric multidimensional scaling approach we have adopted (Preedy and Hackett, 2016).

$$LOD = log_{10}\left(\frac{\mathcal{L}(r = \hat{r})}{\mathcal{L}(r = 0.5)}\right)$$

where $\hat{r}$ is the maximum likelihood estimate of r.

5.2.2.2   Linkage analysis in the presence of preferential chromosomal pairing

In certain polyploid species it is possible that the meiotic pairing is neither fully random nor fully partitioned into exclusively-pairing subgenomes, a situation described as segmental allopolyploidy (Stebbins, 1947). Regardless of the underlying mechanism, the result of preferential pairing is that both the segregation ratios and the co-inheritance of marker alleles are affected. In the example of a 2 x 0 marker introduced earlier, the expected segregation ratio in a polysomic autotetraploid is 1 : 4 : 1 between offspring dosage classes 0, 1 and 2. With increasing preferential pairing, this ratio will approach 1:2:1 in the case of subgenome-straddling markers (*i.e.* with segregating alleles on both sub-genomes, exampled in Figure 5.1B right) and will approach non-segregation in the case of subgenome-specific markers (*i.e.* with segregating alleles on a single sub-genome, Figure 5.1B left).

In order to model this behaviour, we introduce a preferential pairing parameter ρ, such that (in the case of a tetraploid) the probability of the pairing configuration 1-2 / 3-4 is $\frac{1}{3} + \rho$ and the probability of pairing configurations 1-3 / 2-4 and 1-4 / 2-4 is $\frac{1}{3} - \frac{\rho}{2}$. Although this appears to be non-symmetric, the labelling of homologues is arbitrary and merely a convenience for the derivation of the likelihood formulae. Attempting to model preferential pairing at higher ploidy levels introduces further complications; Zhu et al. (2016) have proposed a solution for hexaploids by introducing three preferential pairing parameters θ_1, θ_2 and θ_3 to model deviations in bivalent configurations 1-2, 3-4 and 5-6 respectively, with all other configurations having a probability of $\frac{1}{15} - \frac{1}{12}(\theta_1 + \theta_2 + \theta_3)$. A general

pairing model for hexaploids was also proposed by Sybenga (1995) who parametrised the number of bivalents, quadrivalents and hexavalents as well as chiasma frequencies of chromosome arms. In our software, we have not (yet) attempted to model segmental allohexaploidy, and confine our attention to the tetraploid level for now.

In contrast to alternative approaches, we do not simultaneously estimate ρ and r, which can lead to an over-estimation of the preferential pairing parameter (Wu et al., 2002) – although this may offer the advantage of modelling variable preferential pairing rates along the chromosome rather than a single fixed estimate. Instead, we estimate the chromosome-wide strength of preferential pairing after map construction and thereafter correct the pairwise recombination frequency estimates to revise the maps. The details of this approach have been described elsewhere (Bourke et al., 2017). A robust method of preferential pairing detection and estimation is to use inheritance probability estimates such as those provided by TetraOrigin (Zheng et al., 2016); in polymapR we offer a simpler approach which uses closely-linked repulsion marker pairs to test for deviations from random pairing and simultaneously estimate the strength of this deviation (Bourke et al., 2017):

$$\rho = \frac{2(n_{00} + n_{11}) - 4(n_{01} + n_{01})}{3(n_{00} + n_{01} + n_{10} + n_{11})}$$

where n_{01} is the number of offspring with a dosage "0" at marker A and "1" at marker B *etc.*

Given a parent- and chromosome-specific estimate for the preferential pairing factor ρ, we modify the expression for the expected frequency of individuals in marker class n_{ij} of a tetraploid as follows:

$$E(n_{ij}) = \left(\frac{1}{3} + \rho\right) f_1(r, 1-r) + \left(\frac{1}{3} - \frac{\rho}{2}\right) f_2(r, 1-r) + \left(\frac{1}{3} - \frac{\rho}{2}\right) f_3(r, 1-r)$$

Due to the lack of symmetry in this equation, we must carefully consider all possible conformations *within* each phase, an example of which is shown in Figure 5.1B. The key consideration is to distinguish between subgenome-specific and subgenome-straddling markers – once all possibilities have been identified the rest of the analysis is straightforward. In some phases combinations in a tetraploid there are up to four distinct conformations to be considered. The procedure for estimating r and LOD remain otherwise the same as described previously for the case of random bivalents. The inclusion of preferential pairing imposes an extra computational burden, as each phase can lead to up to four sub-phase conformations, all of which are calculated prior to selection of the most likely phase and its associated r and LOD score.

Finally, in both the case of random and preferential pairing, linkage calculations can be run in parallel (using the packages doParallel and doSNOW (Revolution Analytics and Weston, 2014a, 2014b)) on any Windows or Unix-like multi-core desktop computer (by setting the *ncores* argument greater than 1), resulting in significant time-savings. The procedure only calculates as many pairwise comparisons as needed, by first splitting the marker data into chromosomal clusters before performing a complete pairwise linkage analysis. In this way, even high-density marker datasets with tens of thousands of markers can be processed in a few hours.

5.2.3   Linkage group assignment

In diploid studies, the term linkage group is generally understood to be loosely synonymous with the term chromosome, albeit without explicitly stating that a collection of markers which show mutual linkage are in fact on the same chromosome. In an autopolyploid there are two, possibly three levels of linkage groups to be recognised – homologue groups, parent-specific groups and integrated chromosomal groups. We generally ignore parent-specific groups, as the other two grouping levels are sufficient to describe the precise conformation of a marker in relation to others, allowing us to define chromosome-scale haplotypes of parental markers. Thus for the assignment of markers we focus on chromosomal and homologue groups.

The first step in linkage group assignment is to cluster the 1x0 linkage data, for which we currently use the R package igraph (Csardi and Nepusz, 2006). Clustering is performed using the pairwise linkage LOD scores as thresholds for edges in an undirected graph with markers as nodes; the LOD for independence can alternatively be used if desired, which is reported to be more robust against the effects of skewed marker data (Van Ooijen and Jansen, 2013). In practice with well-filtered datasets both should deliver similar results. Markers (nodes) can be connected to a cluster by a single significant linkage over the LOD threshold used.

A number of visual aids are provided to assist in clustering (Figure 5.2). In general, clustering should be performed over a suitable range of LOD thresholds (*e.g.* from LOD 3 to 10) in order to inform the choice of LOD score to partition the data into both homologues and chromosomes (Figure 5.2A, B). If chromosome and homologue clusters cannot be readily identified using 1x0 markers alone (due to the confounding of true repulsion-phase linkages between homologues and false linkages to other chromosomes), the strategy adopted is to separate out coupling-phase homologue clusters and re-connect these into chromosomal clusters using another marker type (Figure 5.2C). 2x0 or 1x1 markers are preferred because of their relatively high information content and because they are generally expected to be present in sufficient numbers for this task. Occasionally, this approach may cause the fragmentation of certain homologues, leading to a greater number of clusters than expected. Various possibilities to merge these fragments are provided (Figure 5.2D, E, F).
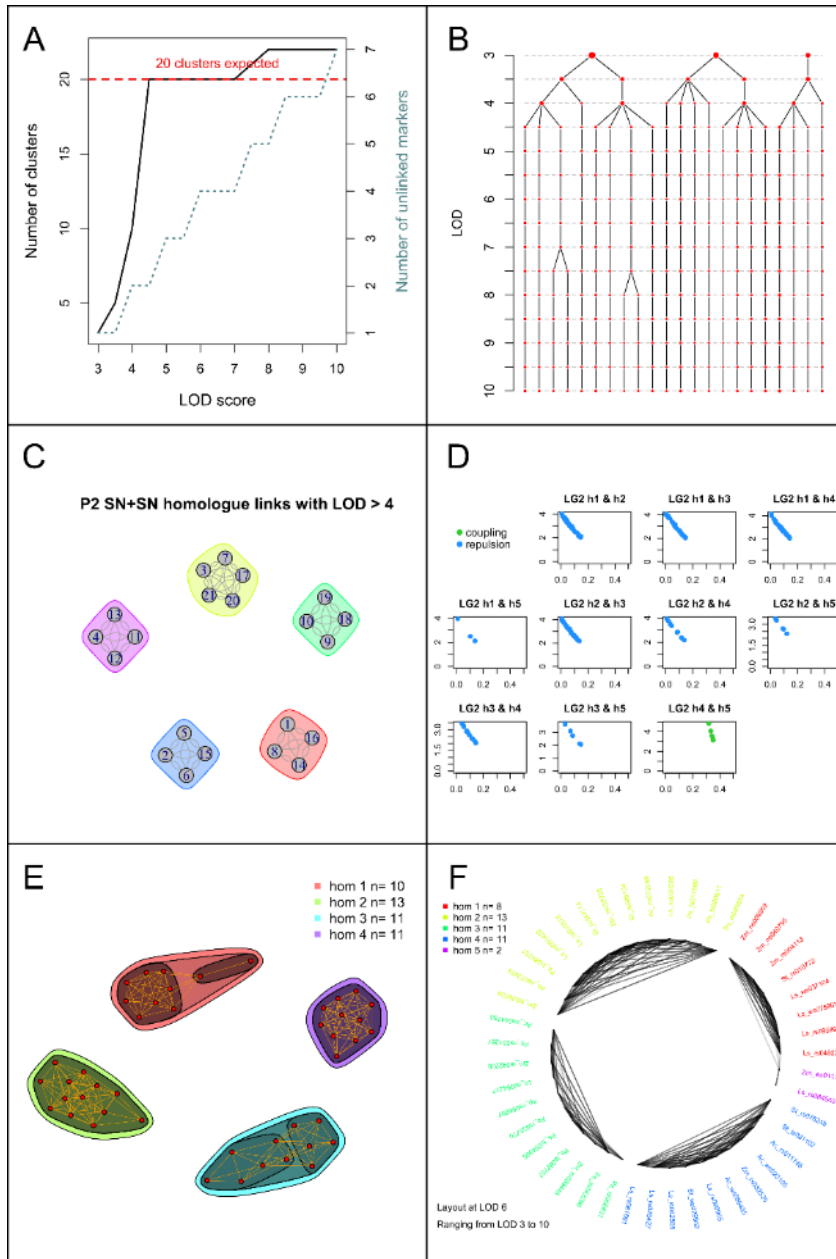
Figure 5.2 Example visualisations produced by polymapR to facilitate linkage group identification and marker clustering. (A) As LOD score is increased, the number of 1x0 clusters increases, as does the number of single-marker clusters (unlinked markers). For a given ploidy and chromosome base number, the expected number of (homologue) clusters is also shown. (B) Alternative representation of 5.2A, akin to JoinMap's grouping tree (but not an interactive plot), which shows the splitting of each cluster as the LOD score is increased. In this example, five chromosomal clusters are identified at LOD 3.5, which split into four homologue clusters between LOD 4.5 and 7. In some situations, at the higher LOD values needed to split clusters into homologue clusters, other homologue clusters may fragment and thus may need to be merged. (C) Using linkage to other marker segregation types such as 2x0 markers, homologue clusters can be associated into chromosomal clusters, if this was not achieved using 1x0 data alone. Here, five chromosomes are represented. (D) If homologue fragments occur, cross-homologue phase information can help

determine which fragments to merge (Bourke et al., 2017). Here, homologues 4 and 5 show only coupling-phase linkage and should therefore be joined as a single homologue. (E) Alternative approach to merge fragments showing network of linkages over a range of LOD scores. Here, four homologues were successfully identified and merged directly. (F) Alternative representation of 5.2E, showing these connections in a circular format instead.

Following the definition of the chromosome and homologue structure using the 1x0 markers, all other marker classes are assigned to both homologues and chromosomes using their linkage to these markers. This constitutes the final phase assignment of all marker types (as distinct from the pairwise phases that are also determined in the linkage analysis but which, for the most part, are not directly used). For example, a 1x1 marker may show significant linkage (user-determined, but with a default value LOD > 3) to a large number of 1x0 markers on homologue 1 of chromosome 1 in parent 1, as well as a large number of 0x1 markers on homologue 2 of chromosome 2 in parent 2. This marker is then phased as 1000x0100, and the chromosome numbering in parent 2 is relabelled to be consistent with that of parent 1.

5.2.4 Marker ordering

One of the challenges of marker ordering and map construction in autopolyploid species using marker dosages is the variable accuracy of recombination frequency estimates which must be integrated somehow. Ordering algorithms which only use unweighted recombination frequency estimates are unlikely to find an optimal map order, as there is no distinction between equal estimates of r from situations with vastly different information contents and variances. A thorough description of this issue is provided in Preedy and Hackett (2016). Within the polymapR package, marker ordering can be achieved in two ways – either using the weighted regression algorithm as originally proposed by Piet Stam (Stam, 1993) or to use the multi-dimensional scaling algorithm as implemented in the MDSmap package (Preedy and Hackett, 2016). The weighted regression algorithm has been programmed in R and is available as an option within the polymapR package. Given the computation efficiency of the MDS algorithm, in almost all circumstances this will be the preferred choice. Identical markers that were originally set aside can be added back to the final maps after marker ordering is complete.

5.3 Results and Discussion

5.3.1 Software output – final linkage maps

The final output of the polymapR package is phased integrated map information. Maps can either be generated per homologue or per chromosome (or indeed per parent), facilitating the definition of haplotypes within a population. A record is kept in a log file of any markers that were removed during the procedure, either at the early data pre-processing stage or later in the marker clustering or ordering steps, as well as logging the exact function calls that generated each step, improving project reproducibility and later reporting. Visualisations are provided throughout the mapping procedure, facilitating the diagnosis of issues as well as summarising the results. An example of an integrated map with five chromosomes, generated using the sample data provided with the package, is shown in Figure 5.3. The polymapR package also generates input files for TetraOrigin (Zheng et al., 2016) which can calculate IBD probabilities for tetraploid datasets, useful for QTL analysis.
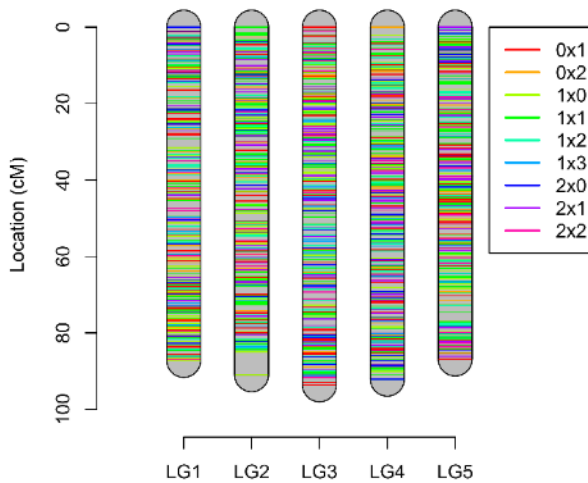
Figure 5.3. Example map output visualisation showing five integrated chromosomes generated using the sample tetraploid dataset provided with polymapR. The colours in the legend depicts the dosage of the parents for the mapped markers ('dosage parent 1' x 'dosage parent 2').

5.3.2   Application of polymapR to real datasets

Developmental versions of the polymapR package have already been used for linkage map construction in various polyploid crops such as potato (Bourke et al., 2016), rose (Bourke et al., 2017) and chrysanthemum (chapter 6). In all of these studies, high-density marker datasets were generated using SNP arrays. Marker dosage genotypes for the tetraploid species were scored with fitTetra (Voorrips et al., 2011), while the chrysanthemum genotypes were scored using fitPoly. The high-density integrated linkage maps from these studies have been used in subsequent applications such as QTL analysis. In the case of potato, the linkage map was shown to be consistent with the physical sequence as well as with previously-published genetic maps (Bourke et al., 2016). In rose, a high degree of co-linearity was demonstrated between the linkage map positions and their syntenic positions on the *Fragaria vesca* genome assembly (Bourke et al., 2017). For chrysanthemum, the maps showed local concordance with the high density linkage map of *Lactuca sativa* (this thesis chapter 6). The maps produced covered almost all expected homologues and identified all expected chromosomes (12 in the case of potato, seven in rose and nine in chrysanthemum) and represent the highest-density polyploid linkage maps published in these species to date.

5.3.3   Comparison with other polyploid mapping software

As described in the introduction, there is currently a limited number of options for linkage mapping in autopolyploid species. We compared the performance and applicability of polymapR with two currently-available software options, namely TetraploidSNPMap (TSNPM) and PERGOLA.

5.3.3.1   TetraploidSNPMap (TSNPM)

TSNPM is a highly-developed piece of software for Windows with a graphical user interface which uses optimised routines for initial marker clustering and offers interactive cluster plots for assignment of linkage groups. It goes beyond linkage map construction to compute IBD probabilities and perform QTL interval mapping as well. Given that polymapR uses the same assumption of random bivalent pairing as is used in TSNPM and the same ordering algorithm (MDSmap (Preedy and Hackett, 2016)), we did not expect much difference in the map output or computational time between polymapR and TSNPM for autotetraploid datasets.

The only clear differences are in the maximum number of markers that can be mapped, the initial clustering of 1x0 markers and the procedure for assigning parental marker phase. In TSNPM the current marker limit is 8000 SNPs, in line with the size of the SolCAP 8303 potato SNP array (Felcher et al., 2012). The polymapR package sets no limit on the number of markers in a dataset, but does bin identical markers to prevent needless calculations. Regarding the initial clustering, polymapR uses the LOD for linkage (the LOD for independence can also be used), whereas TSNPM initially uses a simple matching coefficient as a proxy for the recombination frequency, which is likely to be computationally more efficient.

Marker phasing in polymapR is automatic, by selecting phase based on the counts of significant linkages to 1x0 homologue clusters and ignoring any spurious linkages that go against the general trend. On the other hand, phase assignment often requires manual intervention in the TSNPM pipeline, since a fully-phased map file is required as input for IBD probability determination and subsequent QTL analysis. In the absence of coupling linkages (or either coupling or repulsion linkage in the case of duplex markers) polymapR cannot automatically assign a marker to a homologue; when a marker cannot be fully phased in correspondence to its parental dosage (which determines the expected number of homologues it should be linked to) the marker is omitted from the output phased map file, although retaining its position on the integrated map file. The user can, if desired, investigate the phase assignment further and attempt to determine its most likely phase manually. However, this manual intervention is not required on a per-marker basis (which would be a serious hindrance with high-density marker datasets). Robust automatic tetraploid phasing from marker dosage data and integrated map positions can also be achieved post-mapping using external software such as TetraOrigin (Zheng et al., 2016). It is also worth pointing out that high-density marker datasets contain many more markers than are needed for the accurate phasing of F1 populations (as opposed to parental phasing), and therefore small omissions of fully-phased markers is in practice not likely to be a serious issue.

Furthermore, the generation of homologue maps is straightforward in polymapR, which can be useful in identifying problematic markers or regions when homologue and integrated maps are compared. The main difference between TSNPM and polymapR is that polymapR can handle autotetraploid and autohexaploid as well as segmental allotetraploid data, whereas TSNPM is currently confined to autotetraploid data. polymapR is also cross-platform given that it is written in R.

5.3.3.2 PERGOLA

The PERGOLA package in R has been developed for F2 or backcross populations from an initial cross between homozygous parents. Such a situation is highly unusual for most polysomic polyploids; to our knowledge, none of the polysomic polyploid crops (potato, rose, leek, alfalfa, chrysanthemum, sugar cane *etc.*) are bred through inbreds. This has two important reasons. First, many polysomic polyploids suffer from severe inbreeding depression. Second, inbreeding requires many more generations before homozygosity is reached compared to a diploid or disomic polyploid. In a polysomic hexaploid for example, it would take 25 generations of selfing an F1 individual before 90% homozygosity is reached. The applicability of the PERGOLA software to real populations in polysomic polyploids is therefore limited.

Despite the highly-unusual type of population, we simulated a small F2 dataset of a selfed F1 individuals randomly chosen from a cross of two inbred parental lines using PedigreeSim (Voorrips and Maliepaard, 2012), leading to a marker dataset of 500 duplex x duplex markers over 5 chromosomes. The calculation of recombination frequencies took a mere 3.54 seconds in PERGOLA, in comparison to 28 minutes using polymapR (on a single core; using 6 cores this step took 8 minutes). However, for polymapR this particular marker combination is the most complex, with nine possible phase combinations in the parents to be separately calculated per marker pair, and with extremely complicated likelihood functions for each phase (all 25 dosage combinations are possible in a tetraploid, from n_{00} to n_{44}). It is therefore a somewhat unfair comparison, as PERGOLA labours under no such "generalist" difficulties. Phase considerations are trivial and therefore ignored by PERGOLA because of

their simplistic population assumptions. If such populations could be generated, PERGOLA would produce excellent map orders (in our simulation, all five chromosomes were identified, with near perfect marker order in each, although the map lengths were inflated – from 200 cM using the Kosambi mapping function to 400 cM using Haldane's (when 100 cM was expected). polymapR also produced near-perfect maps with map-lengths of approximately 90 cM using Haldane's mapping function). polymapR can handle data from both cross-pollinating *and* inbred populations whereas PERGOLA cannot, but given the performance difference, PERGOLA would appear to be the software of choice for inbred polyploid populations, should they be developed.

5.3.4   Concluding remarks

The development and release of polymapR comes at a time when there is increasing need for tools to perform genetic analysis in polyploid crops. Understanding the genetic control of important biological traits in polyploid species will have a large impact on plant breeding (or in the case of certain polyploid fish such as rainbow trout or Atlantic salmon, animal breeding as well), facilitating the adoption of genomics-driven breeding decisions such as marker-assisted selection or genomic prediction into breeding programs. For these advances to take place, high-density and accurate maps showing the relative position of markers on chromosomal groups are needed – which is precisely what polymapR delivers. Over the coming years it is predicted that polyploid genotyping will move towards multi-SNP sequence-based haplotypes, which polymapR will in time be adapted for.

5.4   Availability of software

The software, a vignette (manual) and test data set are provided as additional file with this thesis (Additional file 5.1).

Chapter 6

An ultra-dense integrated linkage map for hexaploid chrysanthemum enables multi-allelic QTL analysis

Geert van Geest, Peter M. Bourke, Roeland E. Voorrips, Agnieszka Marasek-Ciolakowska, Yanlin Liao, Aike Post, Uulke van Meeteren, Richard G.F. Visser, Chris Maliepaard, Paul Arens

Abstract

Construction and use of linkage maps is challenging in hexaploids with polysomic inheritance. Full map integration requires calculations of recombination frequency between markers with complex segregation types. In addition, detection of QTL in hexaploids requires information on all six alleles at one locus for each individual. We describe a method that we used to construct a fully integrated linkage map for chrysanthemum (*Chrysanthemum x morifolium*, 2n = 6x = 54). A bi-parental F1 population of 406 individuals was genotyped with an 183,000 SNP genotyping array. The resulting linkage map consisted of 30,312 segregating SNP markers of all possible marker dosage types, representing nine chromosomal linkage groups and 107 out of 108 expected homologues. Synteny with lettuce (*Lactuca sativa*) showed local colinearity. Overall, it was high enough to number the chrysanthemum chromosomal linkage groups according to those in lettuce. We used the integrated and phased linkage map to reconstruct inheritance of parental haplotypes in the F1 population. Estimated probabilities for the parental haplotypes were used for multi-allelic QTL analyses on four traits with different underlying genetic architectures. This resulted in the identification of major QTL that were affected by multiple alleles having a differential effect on the phenotype. The presented linkage map sets a standard for future genetic mapping analyses in chrysanthemum and closely related species. Moreover, the described methods are a major step forward for linkage mapping and QTL analysis in hexaploids with polysomic inheritance.

6.1   Introduction

A linkage map is a starting point for localization of genomic regions that are associated with agriculturally important traits. This makes it an important tool for DNA-informed breeding (Peace, 2017). For polyploids, DNA-informed breeding has lagged behind compared to diploids, because genotyping co-dominant markers and linkage map construction in polyploids requires specialized methods. Such methods need to be able to handle multi-dose markers. As opposed to diploids, polyploids have multiple conformations of heterozygous genotypes; a hexaploid can have five different heterozygous genotypes ranging from a dosage of one to a dosage of five. Together with the two homozygous conformations, this adds up to seven different dosage scores.

Many linkage maps of polyploids are constructed with single-dose (present/absent) markers using methods developed for diploids. Those kinds of maps are limited to representing only individual homologues. Integration of separate maps of homologous chromosomes is needed for transferability of results between mapping studies and mapping of traits with a complex genetic architecture. In an integrated map, all markers are located relative to each other, resulting in one representation of the positions of all mapped loci, irrespective of the phase of their alleles. This enables comparisons of linkage maps based on different populations. Map integration requires estimation of linkage between single-dose markers in repulsion or linkage between higher dose markers. Estimation of linkage of markers in repulsion is different from diploids and can only be done with very low confidence, especially in a hexaploid (Wu et al., 1992). Segregation ratios of higher dose markers are fairly complex, and calculation of recombination frequency needs specific statistical methods (Hackett et al., 1998). Because of the complicated nature of recombination frequency estimation between higher dose markers, dedicated software is required.

In an outcrossing species, the number of alleles that can affect a trait in a single individual is the same as the ploidy level (Figure 1.5A). For QTL detection in an outcrossing full-sib population without any prior knowledge on the involved alleles, all twelve possible alleles that can be inherited from the parents should therefore be taken into account (Figure 1.5B). With use of the positions of markers on a non-integrated linkage map of homologues, only information on the presence or absence of one out of twelve parental alleles is available (Figure 1.5C). If the other eleven alleles are ignored, any QTL that does not have underlying alleles with major effect on the trait will be missed. Multi-allelic QTL mapping needs therefore information of all alleles per locus.

Most progress in linkage mapping in an outcrossing hexaploid with polysomic inheritance has been reported in sweet potato (*Ipomoea batatas*). In this species, several non-integrated maps have been published (Cervantes-Flores et al., 2008; Chang et al., 2009; Kriegner et al., 2003; Monden et al., 2015; Shirasawa et al., 2017; Ukoskit and Thompson, 1997), with two publications reporting information on homologous chromosomes without actually integrating the maps. This information is based on duplex x nulliplex (2x0) and triplex x nulliplex (3x0) markers (Cervantes-Flores et al., 2008; Ukoskit and Thompson, 1997) or based on alignment to a reference genome (Shirasawa et al., 2017). Similar to sweet potato, chrysanthemum is an outcrossing hexaploid with polysomic inheritance. Reported methods for linkage map construction (Zhang et al., 2011a, 2010) and QTL analysis (Zhang et al., 2013, 2012a, 2012b) for chrysanthemum have been limited to methods developed for diploids, and constructed maps are therefore not integrated.

In a hexaploid, an integrated linkage map is most preferably constructed by estimation of linkage with higher dose markers. Those multi-dose markers can connect homologous chromosomes within parents and between parents and can therefore be used to integrate them. For tetraploids, methods to estimate linkage between higher dose dominant markers have been developed (Hackett et al., 1998), and applied to construct integrated linkage maps (Luo et al., 2001; McCallum et al., 2016; Meyer et al., 1998). Later, these methods have been extended and applied to bi-allelic SNP markers (Bourke et al., 2017, 2016; Hackett et al., 2013). Such methods would need to be extended to hexaploids in order to generate integrated linkage maps with use of higher dose markers.

An integrated linkage map can be used to reconstruct inheritance of parental haplotypes to approach a representation as in the example in Figure 1.5B. The two alleles of bi-allelic SNPs can be in linkage disequilibrium with multiple haplotypes, each having a different effect on the phenotype. Such haplotypes can be identified based on the configuration of neighbouring alleles. Methods for reconstruction of haplotype inheritance by estimating probabilities of identity-by-descent (IBD) in tetraploid bi-parental populations have been developed (Bourke, 2014; Hackett et al., 2013; Zheng et al., 2016). Although all methods are theoretically extendible to hexaploids, the method developed by Bourke (2014) is ploidy-level independent and is therefore directly applicable to hexaploids.

In this paper, we describe the construction of an integrated linkage map from all possible marker dosage types in hexaploid chrysanthemum. We are setting the standard for transferability of results by chromosomal linkage group numbering based on synteny with lettuce (*Lactuca sativa*) and by generating a core set of SNP markers that can be used to anchor future maps. With the integrated linkage map, we reconstruct haplotypes based on parental origins using a relatively simple procedure. We demonstrate the usefulness for QTL mapping for four traits for which information of all twelve segregating alleles was taken into account.

6.2   Materials and methods

6.2.1   Plant material and phenotyping

We analysed the segregation of SNP markers in a bi-parental population that consisted of 406 individuals originating from a cross between DB36451 (P1) and DB39287 (P2), two daisy-type, white chrysanthemum cultivars. Phenotyping took place in the same experiment as described in chapter 3. In short, the offspring and parents were grown in three randomized blocks in each of three seasons: summer (May to July 2015), late summer (August to October 2015), and autumn (September to November 2015). A replicate consisted of a field containing 10 to 50 plants. Plants were grown in 12, 12 and 14 days of 18, 21 and 21 h photoperiods for the summer, late summer and autumn, respectively. To induce flowering, they were subsequently grown in 12 h photoperiods for the plants grown in summer and late summer and in 11 h photoperiods for the plants grown in autumn. Flower colour was recorded based on visual observation. If flowers were completely white, they were scored as 0, if they were slightly pink as a 1, and pink flowers were scored as 2. Flowering time was recorded as the number of days (at short photoperiod) needed to reach commercial maturity for at least 50% of the

plants grown in a single field. The number of ray florets was counted from the third flower head from the top of one flower stem for each replicate. The phenotypic scores obtained for disk floret degreening are described in chapter 3. Heritability was calculated by dividing the estimated genotypic variance by phenotypic variance. Variances were estimated using an analysis of variance (ANOVA) with trial and genotype as fixed effects.

6.2.2   Mitotic chromosome counting

For mitotic metaphase chromosome analysis, ± 1 cm long roots were collected from DB36451 (P1) and DB39287 (P2) and incubated in eppendorf tubes in ice water for 24 hr and then fixed in ethanol–acetic acid (3:1) solution for 12–24 hr. Roots were stored in fixative at -20°C until use. For chromosome preparations, the root tips were washed 4 times 5 min in enzyme buffer (0.01 M citric acid-sodium citrate, pH 4.8) and incubated in an enzyme mixture containing 1% (w/v) pectolyase Y23, 1 % (w/v) cellulase RS at 37°C for about 1.5 h. Squash preparations were made in a drop of 45% acetic acid and frozen in liquid nitrogen. The cover slips were removed by using a razor blade. The slides were then dehydrated in absolute ethanol, air dried and stained with 1 µg/ml 4,6-diamidino-2-phenylindole (DAPI, Sigma) in Vectashield (Vector Laboratories). Images of fluorescently stained chromosomes were acquired using a Canon digital camera attached to an Axiophot microscope with an appropriate filter and then processed using software (Axio Vision 4.2). For each genotype, the total number of chromosomes was determined for 5-10 metaphases.

6.2.3   Genotyping and marker quality filtering

Genotyping was performed with a 183k Affymetrix SNP array, as described in chapter 4. In short, the array was designed based on RNA-seq data of 13 cultivars, including both parents of the population. A reference transcriptome was assembled based on the reads originating from DB36451 (the female parent of the population), and reads of all 13 other cultivars were aligned against this assembly. From these alignment files SNPs were called, while retaining information from which transcript contig they originated.

Dosage scoring from array output was largely performed as described in chapter 4. Because genomic dosage was highly correlated with the number of reads per allele from our sequence data (chapter 4), we estimated dosage per SNP of the parents a-priori based on the sequence data, and used this information for SNP calling (Zych et al., in preparation). This resulted in 67,916 SNP markers with expected segregation in the population based on parental dosages. Similar to the description provided in chapter 4, we removed non-segregating markers, markers with >5% missing values, individuals with >10% missing values and skewed markers (p<0.001 based on a X^2-test assuming polysomic inheritance). We defined non-unique markers as markers for which all non-missing dosage scores were equal. Groups of non-unique markers were represented by a single marker with the least missing values. This representative marker was used in further mapping steps with all other unique markers. After ordering, the other markers in the represented group were assigned to the same position as the representative marker.

6.2.4   Linkage map construction

The linkage map was constructed with the functions developed for the polymapR R package (chapter 5). Below, the steps are elaborated in more detail. To calculate recombination frequency (r) and LOD scores of marker pairs, the method as described by Bourke et al. (2016) was modified for hexaploids, *i.e.* using the assumption of completely random bivalent pairing. Initially, marker dosages were converted to their most fundamental form as previously described (Bourke et al., 2016), resulting in nineteen separate marker segregation types (Additional file 6.1). For all possible marker combinations (Additional file 6.2), functions for pairwise estimation of r were then derived. In a hexaploid species, fifteen bivalent pairing scenarios are possible, in comparison to three for a tetraploid. For each combination of marker types, there are multiple phases possible depending on the conformation of the markers within one or both parents. All possible phase combinations were calculated for each marker

pair (*i.e.* all phases having a distinct likelihood function), since the phasing of marker pairs is unknown before mapping. The recombination frequency (and associated LOD) was selected among those estimates of r in the range $0 \leq r < 0.5$ which maximised the log of the likelihood function (Hackett et al., 2013). The accuracy of recombination frequency estimation and phase assignment was checked using a small simulated hexaploid dataset generated in PedigreeSim (Voorrips and Maliepaard, 2012), which showed a high degree of concordance between the true and expected results for most marker combinations. In cases where the accuracy of the estimate was lower, the LOD score reflected this (being loosely related to the inverse of the variance of the estimate). Overall, for each marker type combination (Additional file 6.2) a total of 104 linkage functions were derived in Mathematica 10.0 (Wolfram Research, 2015) and converted to R language (R Core Team, 2014) for the linkage analysis.

To construct backbone clusters that would represent homologues, simplex x nulliplex (1x0) markers were clustered at a LOD score of 10. To identify chromosomal linkage groups (CLG), multi-dose markers can be used to provide bridge linkages between pairs of 1x0 markers, therefore associating clusters into CLG. Abundant multi-dose markers provide the most information, among those are uniparental duplex x nulliplex (2x0) markers (Bourke et al., 2017) or bi-parental markers, like simplex x simplex (1x1) markers. In our case, the use of 1x1 markers showed the clearest associations between 1x0 clusters, and these marker types were therefore used to indentify CLG. Unlinked markers in clusters smaller than five markers were not used in further mapping steps. Linkage information of the bi-parental 1x1 markers were used to assign consensus numbering to the linkage groups between parents. After construction of this backbone clustering, all other marker types were assigned and phased to a CLG and homologue based on linkages with 1x0 markers with a LOD score greater than five. To complete information on all pairwise linkages, for each marker combination within a linkage group, recombination frequency and LOD were calculated with the derived functions. The markers were ordered using MDSmap (Preedy and Hackett, 2016), with parameter settings as suggested by the authors: we used Haldane's mapping function, two dimensions for the principal curves, and LOD^2 as weights. We did not observe any notable change on the map ordering between two and three principal curve dimensions, and we therefore chose to use the most simple setting of two dimensions. After the first round, problematic markers were removed based on visual inspection of the principal curves and the difference in distance between nearest neighbouring markers as estimated from recombination frequency and the distance on the map. This difference is represented by the nearest neighbour fit (Preedy and Hackett, 2016) and markers exceeding a value of four were considered problematic and thus removed. This was repeated if the next round resulted in a reduction of the total nearest neighbour fit. From the integrated map, all marker alleles were assigned to a homologue. This assignment was based on coupling linkages with 1x0 markers that formed the backbone clustering. If there were at least five coupling linkages with 1x0 markers at LOD greater than 5, alleles were assigned to a homologue. If the number of marker alleles was not equal to the number of assigned homologues, the marker was not included in the phased map.

As SNP markers were discovered from an RNA-seq derived transcriptome assembly, each marker is associated with a transcript contig sequence. This information was used to investigate the quality of the map. Markers of the type 1x0 that originated from the same transcript contig should have a distance of 0 cM on the integrated map (assuming the contig was assembled correctly). For each homologue combination and for each linkage group, an overall deviation was quantified by calculating the root mean square error (RMSE) of these differences on the integrated map for all mapped 1x0 markers originating from the same transcript contig.

To enable alignment of any future linkage maps in chrysanthemum by gene sequences, markers were identified that originated from contigs representing characterized genes. For this, mapped markers were aligned to all proteins from the UniProt database from *Chrysanthemum x morifolium* (taxonomy ID 41568) using BLASTX (Altschul et al., 1997). Hits were filtered for alignment lengths greater than 100

and more than 95% identity. A subset of markers originating from these filtered transcript contigs spread over all linkage groups was selected to form a reference linkage map.

6.2.5   Synteny with lettuce

To investigate the synteny of the integrated linkage map with lettuce (*Lactuca sativa*), mapped transcript contigs were aligned to the mapped unigenes of lettuce as available from the Lettuce SFP Chip Project website (http://chiplett.ucdavis.edu, accessed February 20th 2017; Truco et al., 2013) using BLAST (Altschul et al., 1997). Unique hits with an e-value smaller than 1E-100 were used to assess synteny. Chrysanthemum CLG were renumbered based on the number of alignment hits with the lettuce linkage groups.

6.2.6   IBD probabilities

In order to estimate presence of parental haplotypes in the offspring, we calculated IBD probabilities as described in Bourke (2014). Based on the phased map and linkage information, IBD probabilities per marker locus were calculated for each member of the F1 population in two steps. The information was stored in a three-dimensional array for each linkage group, with marker, offspring individuals and homologue on the x, y, and z dimensions. In the first step, only fully informative dosage scores were used to fill the IBD probability array. This means that any scores in the progeny that were larger than zero and smaller than the sum of the parental dosage scores were considered as non-informative (*e.g.* for a 1x1 marker, progeny with a dosage of 0 or a dosage of 2 were considered informative, and with a dosage of 1 non-informative). Progeny dosage scores of zero indicate absence of all homologues to which the marker alleles were linked. Progeny scores that equal the sum of the parental dosages indicate presence of all homologues. Genotypes of which the marker alleles were absent were given an IBD probability of zero for those homologues, and genotypes of which all marker alleles were present were assigned a probability of one. Probabilities which could not be estimated were given a starting probability of 0.5. In the second step, inter-marker distance was used to estimate the IBD probabilities of adjacent markers.

If $P_j = 1$:

$$P_i = P_j - r_{ij}$$

If $P_j = 0$:

$$P_i = P_j + r_{ij}$$

where i and j indicate two markers, P the IBD probability and r recombination frequency as calculated using the reverse Haldane's function from estimated distance on the integrated map. Because in most cases P_i can be estimated based on linkage with multiple informative markers, P_i was updated if the estimated P was lower than previously estimated (if $P_j = 0$), or if P was higher than before the current estimate (if $P_j = 1$) – *i.e.* more extreme probabilities were favoured. After assignment of IBD probabilities, the sum of IBD probabilities per parent was normalized to three (as three alleles are inherited from each parent). For each homologue in each F1 individual, a cubic spline was fitted over IBD probabilities versus position to calculate IBD probability interpolations over 1 cM intervals. Genotype information content (GIC) for interval k at homologue h for n individuals was calculated with the following formula:

$$GIC_{hk} = 1 - \frac{2}{n} \sum_{i=1}^{n} | P_i - \lfloor P_i \rfloor |$$

where

$$\lfloor P_i \rfloor = 0, \quad 0 \leq P_i \leq 0.5,$$

$$\lfloor P_i \rfloor = 1, \quad 0.5 < P_i \leq 1$$

This results in a score for GIC ranging from 0 to 1, where 0 represents a locus with little information, and 1 with complete information.

6.2.7   QTL mapping

QTL analysis was performed on block-corrected mean phenotypic values using an IBD probability model, as described before for tetraploids (Bourke, 2014). An additive model modified from Kempthorne (1957) as suggested by Hackett et al. (2013; 2014) was modified to the hexaploid level:

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

where a_i and X_i are the main effects and indicator variables for allele i, respectively. The parameters representing homologue 1 and homologue 7 were taken as the reference classes and were therefore omitted from the model as in all cases three alleles are inherited per parent. To calculate the significance threshold for detecting significant QTL, a thousand permutations were run with randomly permuted phenotypes (Churchill and Doerge, 1994), taking the 5th percentile of the (ordered) minimum p-values from each genome-scan analysis as an approximate significance threshold. The used significance threshold was the 5th percentile of the minimum p-values of each permutation. To identify homologues affecting the trait, the IBD model was re-run for each allele separately.

6.3   Results

6.3.1   Linkage map

After removal of markers that were non-segregating, had distorted segregation or had more than 5% missing values, 30,532 markers remained in the dataset. Of those, 21,345 had unique dosage scores acroos the progeny (Figure 6.1). Because markers with identical dosage scores in each individual (non-unique markers) will map to the exact same position, they were reduced to a single, unique, marker for calculation of linkage and ordering. The others were added to the linkage map after map construction with only unique markers.
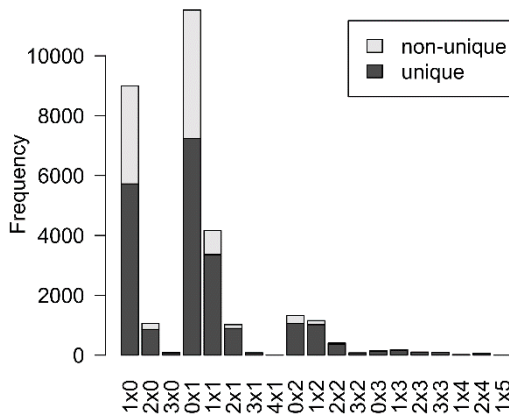

Figure 6.1. Distribution of 19 different marker types segregating in the bi-parental population. Total number of markers: 30,532, of which 21,345 were unique. The non-unique markers had duplicate dosage scores across the population. The labels on the x-axis represent marker segregation types such as simples x nulliplex (1x0) etc. ("dosage parent 1" x "dosage parent 2").

Simplex x nulliplex and nulliplex x simplex (1x0 and 0x1) markers were used to construct backbone clusters that represent homologues. This resulted in 54 clusters for P1 and 53 clusters for P2 each containing five or more markers. Identification of CLG (chromosomal linkage groups) with simplex x simplex (1x1) resulted in a network of nine CLG representing all homologue clusters of both parents. Chromosome counting showed that both parents had 2n=54 chromosomes (Additional file 6.3), the expected euploid chromosome number. Out of the 108 expected homologues (54 per parent), 107 were

identified (Additional file 6.4). All other marker types were subsequently assigned to a CLG based on linkage with 1x0 and 0x1 markers. In total, 21,159 unique markers (99.1%) could be assigned. Markers were ordered per CLG based on recombination frequency with LOD^2 as weights, resulting in CLG map lengths ranging from 64.5 to 95.0 cM. After ordering, the groups of non-unique markers were added to the linkage map based on the position of their unique representing marker, resulting in a linkage map containing 30,312 markers (99.3% of initial; Table 6.1). Of the ordered markers, the alleles of 28,638 (93.8% of initial) could be phased to an expected number of homologues based on parental dosages with at least five significant linkages to 1x0 markers (Table 6.1), resulting in a fully phased linkage map.

Table 6.1. Summary statistics of integrated linkage map

CLG	Length (cM)	Total markers	Phased markers	Contigs*	Rounds†
1	82	2,595	2,528	1,199	2
2	77.3	3,184	3,110	1,411	3
3	64.5	2,970	2,786	1,269	3
4	84.2	3,601	3,215	1,557	3
5	90.3	3,498	3,427	1,508	3
6	91.1	3,619	3,533	1,585	2
7	81.6	3,936	3,464	1,621	2
8	95	3,805	3,499	1,604	2
9	86.1	3,104	3,076	1,338	3
sum	752.1	30,312	28,638	13,092	-
mean	83.6	3,368	3,182	1,454.7	-

*Number of transcript contigs associated with mapped markers.
†Number of rounds of problematic marker removal and re-ordering after the first ordering.

6.3.2 Synteny with lettuce and reference map

We aligned mapped transcript contigs of chrysanthemum with mapped lettuce unigenes (Figure 6.2). We aligned the 13,092 mapped chrysanthemum transcript contigs to 12,841 mapped lettuce unigenes, and obtained 4,757 unique hits with an e-value smaller than 1E-100. This resulted in the identification of syntenic linkage groups between lettuce and chrysanthemum. All combinations of linkage groups of chrysanthemum and lettuce with maximum number of hits were unique, except for two (Additional file 6.4; Additional file 6.5). These two had both most hits with one lettuce linkage group. Of the two ambiguous CLG, the CLG with least hits was renumbered with the number of the non-assigned lettuce LG. This combination still had 126 hits, indicating partial similarity. Syntenic analysis per LG resulted in identification of large regions with linear correspondence between locations of genes, so the genomes appear to be partly co-linear at local scale. This was not clear at a larger scale, as syntenic regions were scattered across the linkage map of lettuce, which can be interpreted as that each chromosome carries major inversions and translocations. With use of this data, we based the numbering of chrysanthemum CLG on the number of significant alignments of mapped transcripts. To mark these nine chrysanthemum linkage groups, we present 92 CLG-defining SNP markers. These are evenly spread over all nine CLG and originate from 85 contigs representing genes coding for protein entries of the UniProt database (Figure 6.3; Additional file 6.6). This should be a useful tool for future studies in chrysanthemum.
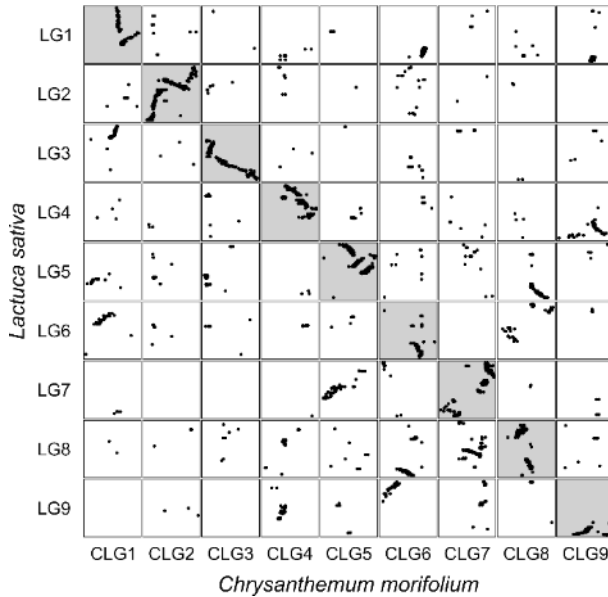
Figure 6.2. Synteny between the lettuce ultra-high density map (Truco et al., 2013) and chrysanthemum. Each dot represents a significant alignment between lettuce unigenes and chrysanthemum transcript contigs.
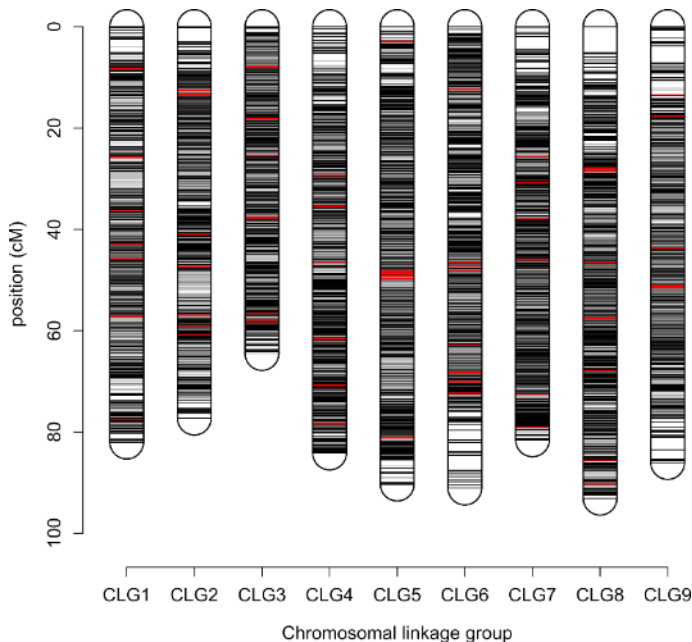

Figure 6.3. Integrated linkage map of phased markers with 1x0 markers (black), other marker types (grey) and CLG defining markers (red).

6.3.3 Linkage map quality

We used two analyses to evaluate the quality of the linkage map. First, to investigate the concordance between estimated pairwise r ($r_{pairwise}$) and r based on map distance (r_{map}), these two estimators of r

were plotted against each other (Additional file 6.7). With high LOD scores, these two estimators were in concordance with each other over a wide range of *r* (from 0 to 0.3). Second, to evaluate the position of nearby 1x0 markers in coupling and repulsion, we aligned the position on the integrated map of 1x0 markers that originated from the same transcript contig from the RNA-seq assembly. The positions of markers that originated from the same contig and had the same phase (6,937 markers in total) aligned nearly perfectly (Figure 6.4), indicating low error-rates. The position of markers phased on different homologues (8,352 markers in total) was more spread. The residual mean squared error (RMSE) was calculated for each linkage group (Figure 6.4) and each combination of homologues from the same linkage group (Additional file 6.8). RMSE was generally below 5 cM, with some outliers, on CLG 2, 5 and 8. These outliers were caused by one, two, and two markers respectively.
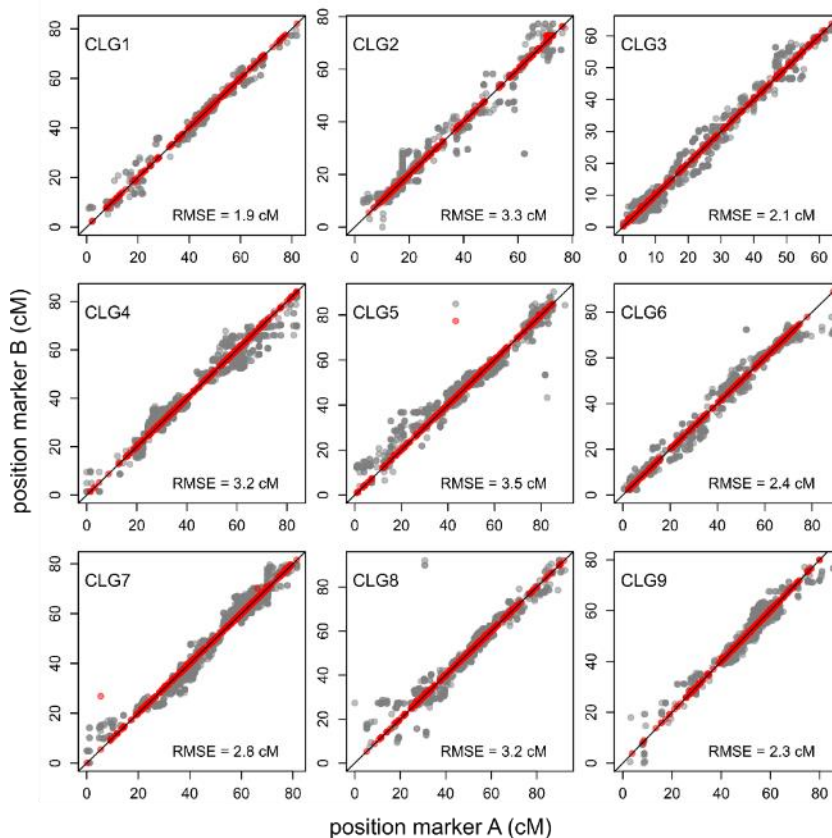
Figure 6.4. Scatterplot of marker positions of 1x0 markers on the integrated map that originated from the same transcript contig. Each dot represents a combination of markers that originated from the same transcript contig. The red dots indicate markers phased on the same homologue, grey dots on different homologues. The black line represents y=x.

6.3.4   IBD probabilities
The presence of each of the twelve segregating haplotypes per locus was estimated in all progeny individuals at 1 cM map intervals, which was expressed in IBD probabilities. In the middle of the CLG, the IBD probabilities could be estimated with high confidence (Additional file 6.4). If there were no markers in large parts of one homologue, the IBD probabilities could still be close to 0 or 1, because information from the five other homologues can complement the missing information. Even if no markers were mapped on the entire homologue, e.g. homologue 12 from P2 on linkage group 4, IBD

probabilities were complemented with information from the other five homologues. Genotype information content was lower towards telomeres, because in those regions markers were often missing in a large range in at least two homologues and informative markers were present on only one side.

6.3.5 QTL mapping

The population was phenotyped for four different traits: flower colour, flowering time, disk floret degreening and number of ray florets. All four traits had a moderately high heritability ranging from 0.68 to 0.72 (Table 6.2). The phenotypes were fitted against the IBD probabilities at 1 cM intervals with an main effects model.

Table 6.2. Summary of phenotypes

Trait	Mean phenotypic value		Broad-sense heritability	Standard deviation
	P1*	P2*		
Flower colour (score)	0	0	0.71	0.613
Flowering time (days)	51.7	49.1	0.70	3.03
Disk floret degreening	0.124	0.527	0.68	0.115
Number of ray florets	34.3	26.3	0.72	5.15

*P1: DB36451; P2: DB39287

Two regions were highly-significantly associated with flower colour, at CLG5 and 7, and one region at CLG9 was slightly associated (Figure 6.5). The highly significant loci were both simplex QTL (Table 6.3; Figure 6.6A; Additional file 6.9). Analysis of variance of the interaction between the associated alleles showed a highly significant ($p < 1E-16$) interaction, indicating that both alleles need to be present to get a pink flower colour. Together, the two 1x0 markers that were most closely linked to each of the QTL explained 48.2% of the variation, indicating that the trait is mainly inherited by two major alleles segregating from two loci from each of the two parents. There is a minor QTL on CLG9, but some genotypic variation is still to be explained by undetected QTL.

For flowering time, we found three clear QTLs on CLG2, 3 and 4 and one minor QTL at CLG8 (Figure 6.5; Table 6.3). For the simplex QTL on CLG4, presence or absence of the allele at homologue 11 had a major effect on the trait (Table 6.3; Additional file 6.9D). In both loci on CLG2 and 4 presence of different alleles could have a positive effect, a negative effect or no significant effect on the phenotype (Table 6.3; Figure 6.6B; Additional file 6.9C). Therefore, at least three alleles underlie the QTLs.

For disk floret degreening, three QTL located on CLG4, 5 and 8 were detected (Figure 6.5). The QTL on CLG4 was simplex and the phenotype was mainly affected by the presence of homologue 11, whereas on CLG5 and 8 multiple alleles underlie the QTL (Additional file 6.9F & G; Figure 6.6C). The QTL on CLG8 explained most phenotypic variation, and presence of the allele on homologue 5 had the strongest effect on the mean value of disk floret degreening.

For number of ray florets, two minor QTL were found on CLG3 and 4. The QTL on CLG3 was affected by alleles from two homologues from the paternal parent with opposite contributions to the trait (Table 6.3; Additional file 6.9H). In addition, the QTL on CLG4 was affected by alleles that originated from only the maternal parent with opposite effects from different homologues (Table 6.3; Figure 6.6D).
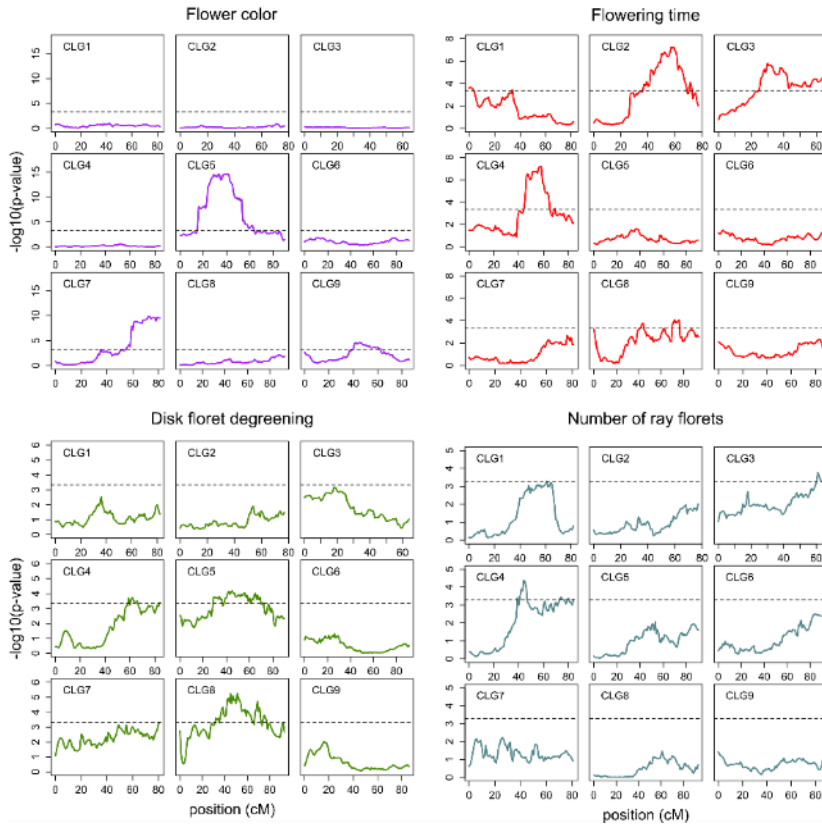
Figure 6.5. QTL analysis of flower colour (purple), flowering time (red), disk floret degreening (green) and number of ray florets (blue). Significance thresholds based on 1000 permutations are indicated with the dashed line.

Table 6.3. Summary statistics of the different QTL.

CLG	QTL peak position (cM)	p-value*	adjusted R²	Associated homologues†
Flower colour				
5	41	2.2E-15	0.208	7+
7	73	1.46E-10	0.151	6+
9	45	2.26E-05	0.080	1-, 2-, 7-
Flowering time				
2	59	6.20E-08	0.116	4+, 5-, 9-, 10+
3	30	1.70E-06	0.096	2+,3-,7+,8-
4	58	6.58E-08	0.115	11-
8	72	8.10E-05	0.071	6-
Disk floret degreening				
4	62	1.90E-04	0.066	4+
5	45	6.66E-05	0.073	4+, 11+, 12-
8	51	5.56E-06	0.089	3+, 5-, 9+, 12-
Ray floret number				
3	61	1.67E-04	0.066	7-, 11+
4	44	4.12E-05	0.076	2+, 4-

*At the peak interval.
†Homologues which presence has a significant effect on the phenotype ($p < 0.01$). A "+" behind the homologue number indicates a positive effect on the phenotypic value, a "-" denotes a negative effect.
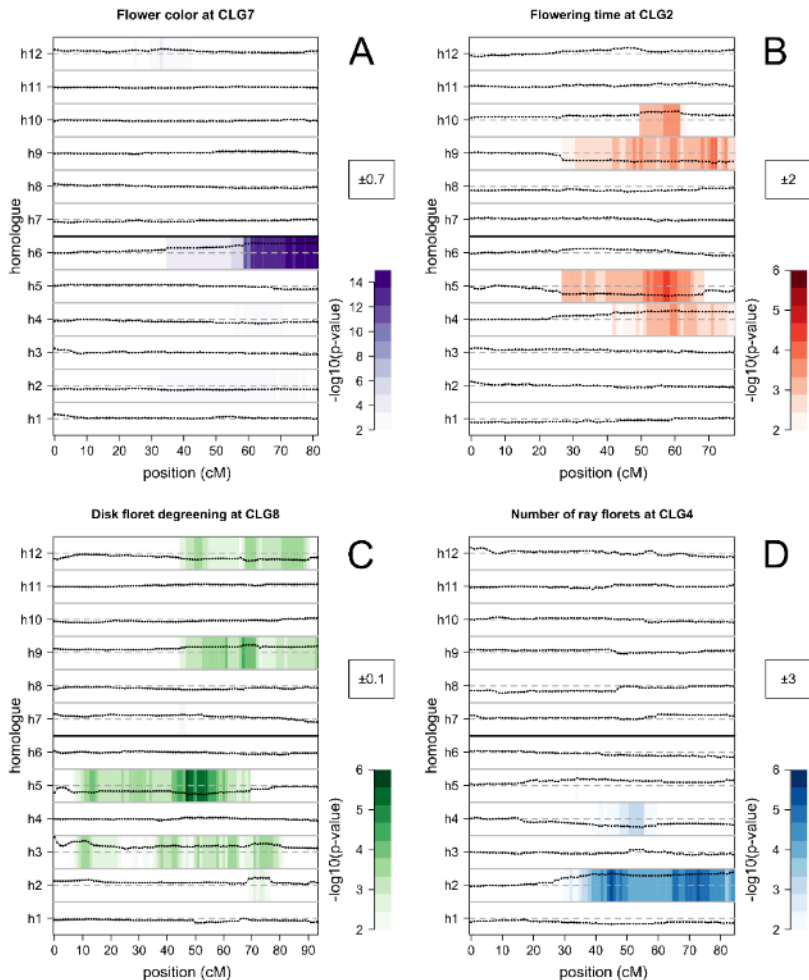
Figure 6.6. Analysis per homologue for four QTL: Flower colour at CLG7 (A), flowering time at CLG2 (B), disk floret degreening at CLG8 (C), and number of ray florets at CLG4 (D). The p-value for testing the significance of the explained variation of IBD probabilities of a single allele versus phenotype ($Y_G = \mu_C + \alpha_i X_i$) is shown as a heatmap. The estimated effect of full absence or presence of an allele on the phenotypic value is shown in the black points. The plot limits of the effect is shown in the black box, meaning that the black points can range within the negative and positive value between the boundaries between homologous indicated by grey lines. The dotted grey lines represent an effect of zero.

6.4   Discussion

In this paper, we report the first integrated linkage map in a hexaploid species with polysomic inheritance. We were able to assign multi-dose markers to their parental homologues. With this phasing information, we could reconstruct inheritance of haplotype alleles in the bi-parental population and perform QTL analyses. We provide major steps to overcome a number of limitations to linkage map construction based on SNP markers in hexaploids, including full map integration and phasing.

6.4.1   An integrated and phased linkage map

The ultra-dense integrated map contained markers of all 19 possible types in a hexaploid (Additional file 6.1 & Additional file 6.2). With our approach, we first defined backbone marker clusters that

represented homologues based on linkages between simplex x nulliplex (1x0) markers. With 1x1 markers that contained information about homologous chromosomes, we created networks of 1x0 linkage groups that represented CLG (chromosomal linkage groups). Subsequently the other marker types were assigned to these backbone clusters. Because we first defined backbone marker clusters based on 1x0 markers, definition of homologues relied on presence of 1x0 markers. Lack of 1x0 markers on a homologue is possible if there is a high degree of inbreeding, the population is a selfing, or if there is selection for a phenotype for which alleles have an additive effect. However, our experience in polyploid mapping to date has shown that genotyping platform can be enriched for low-dosage markers and they therefore tend to be the most abundant marker type (Bourke et al., 2017, 2016; chapter 4). Nevertheless, we were not able to define one homologue on linkage group 4, even though all others each had 175 SNP markers or more. Too few markers therefore seems not very likely. A reason could be that some combinations with alleles on this homologue might have been lethal, and 1x0 markers on this homologue might therefore have had highly distorted segregation. These markers would have been filtered out prior to linkage mapping.

The MDSmap algorithm (Preedy and Hackett, 2016) has proved particularly useful for the weighted ordering of our large number of diverse marker types. As the LOD for linkage varies for the same values of r for different marker type combinations and phases, a weighted ordering algorithm was required. The most frequently used algorithm for weighted ordering is based on a weighted linear regression (WLR) algorithm as deployed in JoinMap (Van Ooijen, 2006). However, this algorithm is limited to approximately a hundred markers per linkage group. Because of this restriction, Bourke et al. (2016) used the WLR algorithm to construct homologue maps separately, which were later integrated. In a subsequent mapping study in tetraploid rose, the MDSmap algorithm was used to construct an integrated map containing over 25k SNPs, without the need for binning or the separation of homologue maps before integration (Bourke et al., 2017). The MDSmap algorithm also forms the core of the map-ordering module within TetraploidSNPMap software (Hackett et al., 2017), although its release as a separate R package opens up the possibility of high-density mapping at any conceivable ploidy level given pairwise recombination frequency information. With the MDSmap algorithm, we were able to order all markers from a CLG in one run (ranging from 1,721 to 2,404 markers per CLG), resulting directly in an integrated map. This number of markers would previously have been completely intractable using a WLR. With the new algorithm, such maps can be produced on an average desktop computer within hours. As the ordering step is time and resource efficient, running multiple rounds of mapping with removal of problematic markers is much more feasible.

Linkage map quality quality was assessed based on two analyses: on the concordance between $r_{pairwise}$ and r_{map}, and on the relative position between 1x0 markers originating from the same contig from our transcriptome assembly. According to the comparison of $r_{pairwise}$ and r_{map}, the two estimators were in concordance if $r_{pairwise}$ could be estimated with high confidence (high LOD). Therefore, there is little discrepancy between the distance and ordering of different combinations of markers on the linkage map and their initial estimation of r. The second analysis resulted in information on the quality of local integration of homologous chromosomes. This is based on the assumption that recombinations are essentially absent within a transcript contig. The maximum contig length was 13 kb (not shown). Because of this small physical distance, we would expect that markers originating from the same contig have a distance very close to 0 cM. From the position of markers originating from the same contigs, a difference from zero can be calculated, and with that a measure for error; we used the RMSE. The RMSE of 1x0 markers mapped on the same homologue was generally very low indicating both a high-quality assembly of transcripts containing mapped markers, and high-quality local ordering at the level of the homologue. The RMSE of 1x0 markers phased on different homologues was higher. Of all marker-type combinations, the estimation of genetic distance between 1x0 markers in coupling is most accurate. Estimation of distance between 1x0 markers in repulsion relies on higher dose markers, because high-confidence estimation of recombination frequency in repulsion of 1x0 markers was not possible with our population size of 406. The positions of 1x0 markers that are on different homologues

relative to each other are therefore estimated with lower certainty than if they were in coupling phase. However, errors in estimating distance between these 1x0 markers in repulsion were in general lower than 5 cM. If serious ordering issues occurred, a much higher value would be anticipated. Nevertheless, there were four homologue combinations with RMSE values higher than 10 cM. Only one or two markers per CLG caused these high values. As these markers were not associated with any notable stress on the map, it is likely that these markers were actually from different loci in the genome, and the contigs they originated may be the result of a chimeric contig assembly from two very similar transcripts originating from the same chromosome.

Earlier linkage maps of chrysanthemum are based on RAPD, ISSR, AFLP (Zhang et al., 2010) and SRAP markers (Zhang et al., 2011a). A disadvantage of these types of molecular markers is that they are difficult to transfer, and different linkage maps therefore cannot be integrated. SNP markers are sequence based, and executing single SNP assays like KASP™ or TaqMan™ are commonly applied laboratory procedures. They can therefore be flawlessly transferred between laboratories. To set a standard for chrysanthemum, we present the sequences of a set of 92 well-distributed SNP markers originating from conserved coding sequences that can be used as a core set to align future linkage maps to each of the chromosomal linkage groups presented here.

6.4.2   Estimating IBD probabilities

We used a relatively simple approach to estimate IBD probabilities for absence or presence of parental haplotypes in our segregating population (Bourke, 2014). The method only uses information of dosage scores if they are fully informative. This means that in case of a 1x1 marker for example, a dosage of 0 and a dosage of 2 in the progeny is fully informative (while assuming absence of double reduction), because it represents inheritance of respectively none of the associated homologues or both. A dosage of 1 is not fully informative as it is not known from which parental homologue the allele originated. Therefore, higher dose markers carry relatively few informative dosage scores. A consequence of our method is that it is only accurate if markers with a large fraction of informative dosage scores are equally distributed over the homologues. In our data, parts of homologues were sometimes poorly endowed with informative markers. This did not turn out to be problematic if at that position all other five homologues for that parent carried enough information. More sophisticated methods have shown that higher dose markers add more information to the estimation of IBD probabilities (Hackett et al., 2013; Zheng et al., 2016). Such methods could result in more accurate IBD estimates, but an adequate marker distribution over all homologues is key to all methods.

The accuracy of genetic analysis based on IBD probabilities relies on the quality of the integrated map. If the estimation of distance between markers with alleles on different homologues is poor, estimation of IBD probabilities of alleles on the presumed same locus will be wrong, and will therefore provide a poor representation. However, the RMSE of the marker positions on the integrated map was generally well below 5 cM. This would not have a large effect on the estimation of IBD probabilities, because according to Haldane's mapping function a distance of 5 cM corresponds to a recombination frequency of 0.047, resulting in a relatively low error of 4.7% on the estimation of IBD probabilities.

6.4.3   QTL mapping

With the integrated map and IBD probabilities, we were able to perform a multi-allelic QTL analysis. In a polyploid, this type of analysis has large advantages over the use of methods that are developed for diploids, because QTL that are regulated by multiple different alleles can be detected and their genetic architecture investigated (Hackett et al., 2014). In a polyploid, more than two alleles can underlie a QTL. This means that the QTL genotype does not only have a dosage, but can also be multi-allelic (*i.e.* not only different conformations of the alleles A and B, but also combinations of e.g. A,B,C,D,E and F are possible within a locus). To investigate the genetic architecture and with that the occurrence of multi-allelic QTL, we performed a QTL analysis that makes use of using IBD probabilities for four traits with different underlying genetic architecture.

The major loci associated with flower colour were bi-allelic. Together, they explained a large part (48.2%) of the phenotypic variation and were affected by one allele for each of the two loci. The two loci clearly showed an interaction, suggesting that presence of both alleles is needed for pink colouration. In chrysanthemum, pink colouration is caused by anthocyanin accumulation (Stickland, 1972). The interaction between alleles could be caused by the requirement of two enzyme variants needed for the production or regulation of production of anthocyanin, or two gene copies that are required for the same limiting step, needing the additive effect of both to become visible.

Several QTLs associated with flowering time, disk floret degreening and number of ray florets were multi-allelic. These QTLs had underlying alleles with a positive effect, a negative effect and no significant effect on the phenotype, indicating presence of at least three alleles. The exact number of unique alleles that affect the phenotype is difficult to determine. Two haplotypes that have the same effect on the phenotype could have the same underlying polymorphism affecting the phenotype, which would make them the same alleles. On the other hand, they could contain different causative polymorphisms that have a similar effect on the phenotype. Based on our data, it is not possible to uniquely identify such alleles, because our analysis is based on genetic linkage, and the causative alleles cannot be identified.

Compared to flower colour, the genetic architecture for flowering time was more complex. The QTL at CLG4 was bi-allelic, meaning that presence of one allele affected the trait, whereas the other eleven alleles did not significantly affect the phenotype. However, in two other major QTL on CLG2 and 3 multiple alleles were involved. Other studies on the inheritance of flowering time in chrysanthemum also suggested involvement of multiple loci (Zhang et al., 2013, 2011b). Flowering time in short day plants is mainly the result of an interaction between growth rate and signal transduction of environmental cues like day-length and temperature. As these cues are strictly controlled in a greenhouse, the role of the environment would be expected to be relatively small. This is supported by the relatively high heritability (0.70), which was also found earlier (De Jong, 1984). However, genetic regulation of signal transduction and growth rate is likely complex and it is therefore not surprising that multiple loci are involved.

Disk floret degreening is an important determinant of postharvest performance of chrysanthemum after long storage (chapter 3). Two multi-allelic QTL were identified, and one simplex QTL. These QTL explained only a small fraction of the phenotypic variation. Disk floret degreening is a physiologically complex trait; in the investigated population it is related to carbohydrate content of the disk floret at harvest (chapter 3). Many sub-traits could affect carbohydrate content, including genotypic variation related to photosynthetic rate and source-sink relationships. Furthermore, it was shown that carbohydrate content is not the only factor affecting degreening (chapter 2). It is therefore not surprising that we did not find major QTL for disk floret degreening. Dissecting the trait further by phenotyping for sub-traits such as carbohydrate content, or by backcrossing progeny harbouring specific trait characteristics might help to further identify specific loci underlying this complex trait.

The number of ray florets had the highest heritability of the investigated traits, but least variation could be explained by detected QTL. Asteraceae plants carry composite flower heads that are comprised of multiple florets. Those florets can be categorized into disk florets and ray florets. The number of ray florets is affected by the number of florets on a capitulum and organ identity of those florets. Regulation of floret identity is generally inherited through one or two major loci in Asteraceae (Gillies et al., 2002). It is therefore quite unexpected we did not find any major QTL associated with the trait. As both parents were of the single flower type, it is possible that both lacked allelic variation in the major genes, and we only found variation in more complexly regulated minor allelic effects.

In the QTL analyses, possible interactions between alleles were not taken into account. An alternative model as described by Hackett et al. (2014) that uses all possible genotype classes as parameters would enable detection of interactions. However, the method we used to estimate IBD probabilities is

not able to estimate probabilities for these genotype classes directly. More importantly, in a tetraploid, there are 36 possible genotype classes ($\binom{4}{2} \times \binom{4}{2}$), leading to a model with 36 parameters that is already prone to over-fitting. In a hexaploid this would be 400 genotype classes ($\binom{6}{3} \times \binom{6}{3}$), leading to 400 parameters; over-fitting would definitely become an issue.

Our results show that hexaploidy in chrysanthemum complicates QTL analysis because multiple alleles with a differential effect can underlie an associated locus. With the integrated map and IBD probabilities we were able to identify inheritance of parental haplotypes in the progeny, enabling us to identify effects of specific alleles that affected the phenotype. We indeed found clear examples in which different alleles from the same locus and parent affected the trait negatively or positively. With these findings we show that polyploids with polysomic inheritance can harbour much more diversity on a single locus compared to a diploid, and this is very important to take into account during QTL detection and breeding.

6.5   Conclusions

The methods described in this paper enable construction of integrated linkage maps in hexaploids with polysomic inheritance. With such integrated linkage maps, it is possible to perform QTL analysis that takes all possible alleles into account at the same locus. This has major impact on the possibilities for localization of genomic loci and their genetic architecture associated with traits in chrysanthemum, but also for other agriculturally-important hexaploid species such as sweet potato, kiwi and persimmon.

6.6   Acknowledgements

6.7   Additional files

Additional file 6.1. Dosage conversions for each marker type.

Additional file 6.2. Network representing all linkage functions. The dots represent a marker type as "dosage parent 1" x "dosage parent 2". The edges represent each possible function to calculate linkage between the two marker types. Within each function, multiple phase combinations are possible. Black lines represent unique functions, grey lines represent that were interchangeable with a representing unique function.

Additional file 6.3. DAPI-stained metaphase chromosomes of the parents of the F1, DB36451 (A) and DB39287 (B).

Additional file 6.4. GIC and marker distribution for each linkage group. GIC is depicted in the bars running from yellow (GIC = 0.2 to blue (GIC = 1). Vertical lines represent markers, in which 1x0 markers are depicted in black, and other marker types in grey.

Additional file 6.5. Number of significant alignment hits (E-value < 1E-100) for each combination of chrysanthemum CLG with lettuce LG

Additional file 6.6. Position, gene and sequence of linkage group defining markers.

Additional file 6.7. Scatterplots between pairwise estimation of r ($r_{pairwise}$) and r based on distance on the ordered linkage map (r_{map}). Colour of the dots is based on LOD score of $r_{pairwise}$. LOD scores greater than 50 are depicted in yellow.

Additional file 6.8. Heatmap of RMSE of all combinations of 1x0 markers originating from the same contig, for each homologue combination.

Additional file 6.9. QTL analysis per homologue. Legend as in Figure 6.6.

Chapter 7

General discussion

From the 17th century onwards, the trade in ornamentals has been a major economic activity. World-wide, the economic importance of ornamental trade keeps increasing. To ensure a sustainable and growing chain of propagation, growth and distribution, challenges lay ahead for growers, traders and breeders. Because the genetic material is key for many improvements to enable sustainability, breeders will play an important part to improve the entire chain. Technological and scientific advances that enable DNA-informed breeding and high-throughput phenotyping have large potential to make major steps forward in genetic improvement. However, many major traits are heterogeneous, and many ornamentals, like chrysanthemum, are polyploid and outcrossing, which makes the study of inheritance a complicated task. In this chapter, I discuss the progress in knowledge and technology that has been described in this thesis in a framework of genetics and breeding of ornamentals and outcrossing polyploids.

7.1 Breeding for heterogeneous traits

Breeding aims to combine the most favourable traits into one genotype. Many important traits have multiple underlying sub-traits; those traits can be called heterogeneous. Postharvest performance in chrysanthemum falls into this category since a whole range of symptoms can make the appearance of a flower stem unacceptable. A proven strategy to cope with heterogeneous traits in breeding, is to subdivide the trait of interest into its sub-traits and improve the higher order trait by selection on the sub-trait level. However, to define those sub-traits, underlying symptoms and their importance need to be investigated.

7.1.1 Disk floret degreening is related to carbohydrate starvation

Knowledge on the physiological background of genotypic differences helps breeding by development of phenotyping methods, applying candidate gene approaches and discovery of interactions between traits. To use physiological knowledge for breeding, knowledge on the involvement of a physiological process in a phenotype is one thing, connecting this to genotypic differences is another. To aid breeding for postharvest performance, we aimed to identify such processes underlying disk floret degreening (chapter 2 and 3). Disk floret degreening is an important component of postharvest performance for chrysanthemum flowers that have been stored for more than two weeks. In chapter 2, we showed that feeding with sucrose can prevent disk floret degreening, which indicated that carbohydrate starvation plays a role. However, genotypic differences in carbohydrate content did not explain differences in disk floret degreening in all cases. It therefore seemed that genotype-dependent differences existed in the processes that are related to carbohydrate metabolism that underlie phenotypic differences.

To draw conclusions on carbohydrate starvation as underlying physiological process of genotypic variation in disk floret degreening, we needed to study a large set of genotypes. Because of the trait complexity, a large cultivar panel of distantly related cultivars would have likely diluted the effects of different underlying sub-traits that reside in the panel. Therefore, identification of the processes that cause phenotypic variation required a set of genotypes with limited genetic variation affecting disk floret degreening. In chapter 3, we used a bi-parental population to reduce the genetic variation to find evidence for the involvement of carbohydrate content of the disk florets on disk floret degreening. The parents had a large difference in sensitivity to disk floret degreening and a corresponding difference in carbohydrate content of the disk florets. We show that carbohydrate content of disk florets co-segregated with degreening and therefore explains the difference in disk floret degreening in the parents. However, the relatively weak correlation between carbohydrate content and disk floret degreening also indicated that other processes must be involved. This is substantiated by the absence of a correlation of carbohydrate content and susceptibility to degreening in the three investigated cultivars in chapter 2. The conclusions on the underlying processes of disk floret degreening drawn in chapter 3 therefore only apply to the investigated population.

7.1.2   Breeding for sub-traits

The relations between sub-traits of heterogeneous traits often have a hierarchic structure. In our case, disk floret degreening is one of the sub-traits of postharvest performance, carbohydrate content is a sub-trait of disk floret degreening, and carbohydrate content likely has multiple underlying sub-traits. This raises the question how far one should go into disentangling heterogeneous traits into sub-traits. At every step going further down into sub-traits of sub-traits, genetic complexity is reduced, heritability might be increased, but explained variation in the highest-level trait is inevitably lost. In addition, like the differences in carbohydrate content in the bi-parental population, the variation of sub-traits might become only relevant for a smaller part of the germplasm. The level at which to collect phenotypic information is therefore a balance between accuracy and resource-efficiency of the phenotyping method, the link with the main trait and the genetic complexity. From a breeding point of view, the ideal level to phenotype is where traits are heritable and measuring them is resource-efficient while maintaining the link with the high-level trait.

In our case of disk floret degreening, we used colour measurements of detached flowers as a phenotype. The developed method was cheap, simple, correlated with and had a much higher heritability than vase-life tests. Because disk floret degreening is quite high up in the cone of hierarchical trait subdivisions, it is not unexpected that it segregated quantitatively. The QTL analysis of sensitivity to disk floret degreening resulted in three QTL, which did not explain much phenotypic variation: between 7 and 9%. The power of those molecular markers to predict postharvest performance is much smaller compared to the developed phenotyping method. From a practical point of view, the phenotyping method is therefore a much more feasible selection criterion.

In conclusion, breeding for heterogeneous traits requires understanding of the underlying sub-traits. Because it is often not known *a priori* what kind of underlying sub-traits are important, research on heterogeneous traits can be risky in terms of outcomes. The success of finding and phenotyping the sub-traits depends on heritability, the ease at which parameters are measurable and the available knowledge on the trait and its underlying sub-traits.

7.2   The mode of inheritance: implications for breeding

In chapter 4, we provide conclusive evidence for hexasomic inheritance in chrysanthemum. This means that each of the six homologous chromosomes can pair at meiosis, and there is recombination over generations between all homologous chromosomes. The mode of inheritance is not only relevant for methods used for genetic mapping (Bourke et al., 2017), it has also large implications for trait inheritance, heritable fixation and genetic load.

The six alleles residing in a hexaploid can harbour three times more variation than two in a diploid. The resulting multitude of allelic combinations add even more to the possible variation in germplasm. However, in disomic polyploids this large variation is likely not present in reality. A first reason is that alleles on different sub-genomes in disomic polyploids are often subjected to silencing, loss and diversification (Jackson and Chen, 2010; Levy and Feldman, 2004). Secondly, heterozygosity can be reduced within a few generations of inbreeding because only two chromosomes can pair during meiosis. In polysomic polyploids however, all homologous chromosomes can recombine. Therefore, heterozygosity within the recombining genome is more easily maintained. In addition, loss or silencing of alleles require many generations to become fixed, and heritable specialization of sub-genomes does therefore not take place. Each of the six alleles in a polysomic polyploid can have a different effect on phenotypic variation. Therefore, different combinations of alleles can lead to a very broad range of phenotypes that are regulated by a single locus. In a polysomic hexaploid, there are 20 ($\binom{6}{3}$) different possible gametes for each locus instead of 8 (2^3) in the case of disomic inheritance (Figure 1.4). The range of possible allelic combinations in a bi-parental population is therefore very large, and factors greater compared to disomic inheritance ($20 \times 20 = 400$ vs. $8 \times 8 = 64$ for a single locus). Therefore, hexasomic inheritance can strongly increase phenotypic variation compared to disomic inheritance. This

is amongst others illustrated by the large range of flower colours and shapes in F1 populations in chrysanthemum (Figure 7.1).

Figure 7.1. Examples of F1 populations in hexaploid chrysanthemum. Each picture represents the progeny of a single cross.

The diverse possibilities of allelic combinations and dosages makes successful heritable fixation of traits more difficult in polysomic polyploids compared to disomic. Even in the simplest case of a dominantly inherited trait, the phenotype will be difficult to keep from segregating in future generations. If the donor parent is simplex for the trait, it will segregate in a 1:1 ratio in the progeny, and dosage would need to increase to quadruplex (four) before all progeny inherits at least one of the dominant alleles. Conversely, if the dominant allele is duplex on one subgenome in a disomic polyploid, all progeny will inherit one dominant allele, and the phenotype will therefore not segregate. Trait fixation becomes even more challenging for a polysomic polyploid in case of involvement of multiple different alleles. Only a small fraction of the progeny will contain the most favourable allele combination. Especially if this combination is favourable because of interaction, and not additivity, optimal selection procedures based on the genotype are complex.

In a hexaploid, unfavourable alleles have a high chance to be masked by other alleles. If one of the six alleles in a polysomic hexaploid gets an unfavourable mutation, there will probably be no effect on the phenotype, because most of such mutations are recessive. Even if the dosage of such an allele increases, it will not be detected in the phenotype. Therefore, unfavourable alleles will easily remain in

the breeding germplasm as genetic load, even after multiple generations of selection (Otto and Whitton, 2000). All six alleles would need to be replaced by favourable ones in all parents, before effects of unfavourable alleles are absent in the offspring. This has also important implications for introgression breeding. If wild or distant material is used as a donor for a trait of interest, alleles of most or all other loci will have to be eliminated in recurrent (pseudo) back crosses. In the first backcross only 5% of the offspring will be homozygous for the recurrent alleles for each locus, whereas this would be 50% in a diploid (Figure 7.2). This means that reducing the allele frequency of the unfavourable background of the donor parent needs many generations and stringent selection.

Breeders of polysomic polyploids have to work with complex trait inheritance, limited heritable fixation and genetic load. The complex trait inheritance is characterized by many possible combinations of alleles and differential dosage effects. This has a profound positive effect on the effective population size and trait diversity. The frequency of alleles can be successfully changed based on selection on the phenotype if they have an additive or dominant effect on the trait. The allele frequency of alleles that interact or are recessive are much more difficult to alter, and this should be taken into account during breeding and selection.
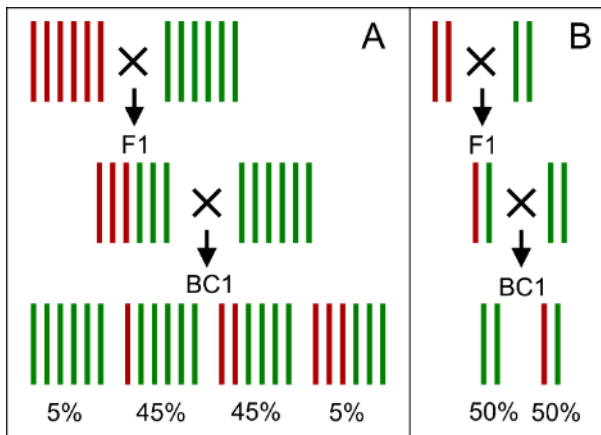
Figure 7.2. Example of reducing unwanted background in a (pseudo) backcross in a hexaploid with polysomic inheritance (A) and a diploid or a locus within a subgenome in a polyploid with disomic inheritance (B).

7.3   Allo- or autopolyploid?

In recent literature, chrysanthemum has been postulated as a (segmental) allohexaploid (e.g. Klie et al., 2014; Li et al., 2011), meaning that polyploidy would have originated from a hybridization between multiple species. Most arguments given in the literature for allopolyploidy is based on prevalence of bivalents at meiosis (Chen et al., 2009; Zhang et al., 2010), but bivalents are not necessarily a feature of disomic inheritance (Stebbins, 1940; Watanabe, 1983). Cultivated chrysanthemum could have originated from a single hexaploid ancestor or be the product of multiple hybridizations with closely related autohexaploid species as postulated by Dowrick (1953). Multiple hybridizations with several related species are common in many crop species, but such hybridizations do often not make a polyploid allopolyploid; *e.g.* an introgression segment is not evidence for allopolyploidy. The question which can be raised therefore: are there reasons remaining to categorize cultivated chrysanthemum as an allohexaploid?

A common feature of allopolyploids is that they usually have disomic inheritance, which is in contradiction with the strong evidence for full hexasomic inheritance in cultivated chrysanthemum (chapter 4). In addition, the non-biased gene expression in cultivated chrysanthemum (chapter 4) is more a characteristic of autopolyploids (Albertin et al., 2005; Church and Spaulding, 2009) than of

allopolyploids (Adams et al., 2003; Albertin et al., 2006; Leach et al., 2014). However, these findings do not exclude allopolyploidy; there could have been an interspecies hybridization that gave rise to higher ploidy levels, and at some point, the sub-genomes of all chromosomes started recombining resulting in full hexasomic inheritance. Recombination between sub-genomes has been reported in several allopolyploids (Allendorf and Danzmann, 1997; Leal-Bertioli et al., 2015; Nguepjop et al., 2016; Stift et al., 2008), and it could therefore be possible that some allopolyploids move towards full polysomic inheritance. However, this is difficult to prove, and the opposite, sub-genome specialization towards full diploid inheritance, is a commonly observed phenomenon (Levy and Feldman, 2004; Song et al., 1995). The move towards diploïdization on the one hand or towards polysomic inheritance on the other, probably also depends on the genetic distance between the hybridized species. This makes the classification of allo- or autopolyploidy problematic, because involvement of different closely related species will be difficult to reconstruct if the sub-genomes recombine.

As opposed to the presumed allopolyploidy, autopolyploidy should be considered as another and even more likely origin of cultivated chrysanthemum. As there is crossing over between homologous chromosomes, it might never be possible to distinct allohexaploidy from autohexaploidy. If one would have to make a choice, chrysanthemum has more genetically relevant features of an autohexaploid than an allohexaploid, and is therefore best described as an autohexaploid.

7.4   Polyploidy: how necessary is the evil?
DNA-informed genetic improvement in outcrossing polyploids needs more resources, knowledge and time compared to inbreeding diploids. New traits that interact or are recessive are difficult to fix, and segregation analysis is more complex. To overcome such disadvantages, efforts are ongoing to move from a vegetatively propagated tetraploid potato towards diploid F1 hybrids propagated through seeds (Lindhout et al., 2011). Polyploids are well represented in the most important ornamental species; six out of the top 10 cut-flowers at the Dutch auctions are polyploid. This is most likely not a coincidence. The question which can be raised therefore is whether polyploidy is really *needed* for polyploid crops.

In general, there can be two reasons why polyploids have a higher yield or fitness compared to diploids. The first is genetic load; the chance of masking of deleterious alleles is higher in polyploids (Otto and Whitton, 2000). The second reason is progressive heterosis, caused by specific allelic interactions between multiple alleles resulting in multi-allelic over-dominance (reviewed by Birchler et al., 2010). These two reasons are not exclusive, because progressive heterosis could be explained by the genetic load theory, through the concept of pseudo-overdominance (Figure 7.3). If progressive heterosis based on single genetic loci is absent, the genetic load theory would assume that after removal of all genetic load, polyploids would not outperform diploids in terms of yield or fitness. Because selection against unfavourable alleles in diploids is much more efficient compared to polyploids, breeding at the diploid level would quickly result in outperforming genotypes. The progressive heterosis theory would imply existence of specific interactions between multiple alleles that cause an improved phenotype over the main effect of each individual allele. If these interactions exist, diploids would not outperform polyploids if the same allelic variation is used in the breeding program.

Similar to adaptation in nature, selection for improved traits during breeding is not only about selection against unfavourable alleles, but also about using variation. The effective population size that carries variation is potentially much larger in polyploids compared to diploids. Alleles that are unfavourable at first, might become favourable if selection criteria change. Unfavourable or "unused" alleles have higher chance to remain in polyploid than in diploid breeding germplasm. Therefore, polyploids can have much higher adaptation speed than diploids (Otto and Whitton, 2000). Polyploidy might therefore be favourable for the process of domestication itself (Dubcovsky and Dvorak, 2007), because fast adaptation to new circumstances, like cultivation, increases the chances of success of domestication.

Whether polysomic polyploidy is advantageous in a domesticated crop or whether it is just an inconvenient heirloom of their suitability for domestication, probably depends on the need for change

and adaptation of the breeding germplasm, but also on the possibility for clonal propagation. For crops that need to adapt quickly to new environments or new consumer needs and have a clonal mode of propagation, polysomic polyploidy might be favoured. For ornamentals, consumer demand for specific flower types or colour can strongly vary over years. In addition, development of completely new flower types or colours can open up entirely new market niches. Increasing variation is therefore perhaps favoured over true-to-type inheritance for ornamental breeding. Contrary, progress in food crops is usually reached by approaching a continuous ideotype, and less extreme phenotypic changes might be required. More importantly, economically feasible multiplication of food crops often needs a massive multiplication system through F1 seeds, which would become severely more difficult in the case of polysomic polyploidy. Therefore, maintaining diversity through polyploidy is probably more advantageous in clonally propagated ornamentals than in many food crops.
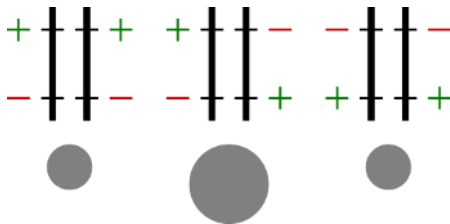
Figure 7.3. Pseudo-overdominance. The size of the grey circles represents a phenotypic value. The green "+" represent a dominant allele with positive effect, the red "-" an allele with a recessive negative effect.

7.5 Disentangling hexaploid genetics: impact and prospects

In chapter 6, we describe the construction of an integrated linkage map for hexaploid chrysanthemum. A linkage map is a representation of the recombining genome. It therefore provides detailed information on the level of independence of allelic segregation. Consequently, it enabled us to develop methods to execute multi-allelic QTL analyses. The integrated linkage map and methods for QTL analysis provide essential information for the design of DNA-informed plant breeding programmes, because the genetic architecture of a trait defines favourable allelic combinations, and with that choice of parents and selection strategies.

7.5.1 Hexaploid genetics in a tetraploid framework

The hexaploid genetic linkage map was constructed based on the building blocks developed by the community that has worked on the disentanglement of the genetics of polysomic tetraploids. The SNP dosage calling methods were based on FitTetra (Voorrips et al., 2011), the methods for calculating linkage estimates, and reconstruction of chromosomal linkage groups and homologous chromosomes based on earlier work in tetraploids (Bourke et al., 2017, 2016; Hackett et al., 2013), very fast and accurate marker ordering with use of MDSmap (Preedy and Hackett, 2016), and reconstruction of inheritance of parental haplotypes based on a method initially developed for tetraploids (Bourke, 2014). With the solid basis from tetraploids we could move away from genetic analysis that uses methods designed for diploids to full map integration with use of all marker-types. The possibility of a situation in which multiple alleles from a single locus affect a trait is an important characteristic that is ignored if genetic analysis is performed on homologue specific maps. In chapter 6, we show that for two important quantitative traits, methods to detect QTL that account for involvement of multiple alleles are essential to disentangle the genetic architecture.

7.5.2 The need for multi-allelic markers and a reference genome

The chrysanthemum linkage map consists of SNPs. SNP assays are high-throughput, sequence based, and bi-allelic. They therefore provide an excellent methodological framework for linkage map construction and marker-trait analyses, for which a large number of markers with allelic information are

very useful. However, SNPs also have limitations, especially for QTL mapping in polyploids. As we have shown in chapter 6, multiple alleles can affect a phenotype in a polyploid, but a SNP can tag only two. This means that in single SNP analyses other non-tagged alleles are ignored. This is illustrated by the difference between single marker-trait associations and identity-by-descend (IBD) probabilities (Figure 7.4). The QTL for flowering time at CLG2 and 4 (presented in chapter 6), have different genetic architecture; the QTL on CLG2 has four underlying alleles of which two had a negative effect on the trait, and the other two a positive. The QTL at CLG4 however, has one underlying allele. The homologue responsible for the QTL at CLG4 contained many 1x0 SNPs, which resulted in a clear peak with SNP-trait associations. On CLG2 however, there were no SNPs tagging each of the associated alleles, and if they did it would not have been possible to detect the different effects of the alleles. The disability to tag multiple haplotypes in one marker-assay therefore limits marker-based QTL analysis.

Multi-allelic markers that tag haplotypes could bridge the gap between SNP markers and IBD probabilities. These type of markers can be retrieved with genotyping-by-sequencing (GBS) techniques like bait capture (Muraya et al., 2015; Uitdewilligen et al., 2013), RAD-seq (Baird et al., 2008; Schönhals, 2014), amplicon sequencing (Jamann et al., 2017), or whole genome re-sequencing (Takagi et al., 2013). Haplotypes are formed by linkages between SNP alleles through strong linkage disequilibrium. Such strong linkage disequilibrium between SNPs can be identified with read information; if two SNPs reside on the same read, they are physically very close and therefore in the same haplotype. This haplotype information can be used to tag alleles, and perform multi-allelic QTL analysis.

Haplotype inference by GBS methods is strongly aided by a reference genome sequence. Such a sequence can be used to align newly acquired sequences, and assess their position and alignment quality. A *de novo* assembly of a reference sequence for hexaploid chrysanthemum is still out of reach, despite fast developments in increasing read length and assembly software. Luckily, three efforts are on their way to sequence diploid wild relatives. One for *Chrysanthemum lavandulifolium* at Beijing Forestry University, another for *Chrysanthemum boreale* at Sahmyook University in Seoul (Cuyacot et al., 2016), and a third for *Chrysanthemum japonicum* at Plant Breeding, Wageningen University & Research. With use of the ultra-dense genetic linkage map described in chapter 6, structural variants between cultivated chrysanthemum and the diploid relatives can be identified. This can be used to investigate the origin of cultivated chrysanthemum, but also to better use the diploid genome sequence for applications in hexaploid chrysanthemum.
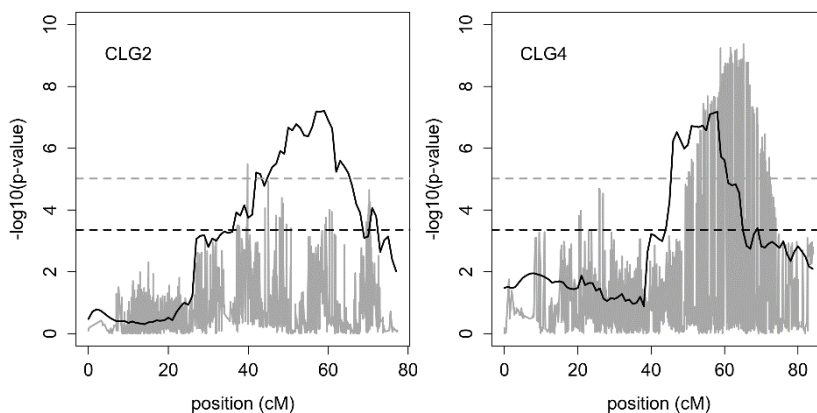

Figure 7.4. QTL plots of flowering time at CLG2 and CLG4. The IBD model (chapter 6) is represented by the black line, the single SNP analysis by the grey line. The dotted lines represent the significance threshold for the IBD model (black) and the single SNP analysis (grey).

7.5.3   Prospects of genome editing techniques

The chance that a newly acquired recessive allele is masked is very high in a polysomic hexaploid. Such recessive alleles are therefore difficult to put to use for breeding. However, recessive alleles, like alleles that enable recessive resistance (so-called susceptibility (S-) genes), are a very interesting source of newly acquired traits (Schaart et al., 2016). Fixing a new recessive trait in polyploids requires a very large number of generations, because it would need to be fully homozygous. Sequence specific nuclease technologies like CRISPR-Cas9 can induce a heritable targeted mutation on all homeologous genomes in allohexploid bread wheat (Liang et al., 2017; Y. Wang et al., 2014), and would therefore also be applicable to mutate all six homologous alleles in polysomic chrysanthemum. These relatively new methods therefore provide a very promising tool to induce (new) recessive alleles in polyploid crops, and potentially can revolutionize polyploid plant breeding.

7.6   Conclusion

The future of polyploid breeding lays wide open. The development of technologies to characterize and modify genomes, together with methods described in this thesis (and the references herein) to perform genetic analysis are bringing us closer to the point where genetic improvement of hexaploids is approaching the current progress and speed in diploids. The reconstruction of inheritance of parental haplotypes in a hexaploid full-sib population provides detailed information to model the relationship between phenotype and genotype. Together with the understanding of plant physiology to develop dedicated phenotyping methods, this enables the disentanglement of genetic architecture of important but heterogeneous traits like postharvest performance. The approach described in this thesis based on SNPs requires a very large high-quality dataset which is at time of writing quite costly. Multi-allelic markers obtained with GBS methods potentially carry more information and are cheaper. A reference genome strongly aids the development and analysis of GBS methods, and is therefore the next big step in genetic improvement of chrysanthemum and other understudied polyploid crops.

References

References

Abe, H., Nakano, M., Nakatsuka, A., Nakayama, M., Koshioka, M., Yamagishi, M., 2002. Genetic analysis of floral anthocyanin pigmentation traits in Asiatic hybrid lily using molecular linkage maps. Theor. Appl. Genet. 105, 1175–1182. doi:10.1007/s00122-002-1053-7

Acquaah, G., 2007. Principles of Plant Genetics and Breeding, 1st ed. Blackwell Publishing, Malden, USA.

Adachi, M., Kawabata, S., Sakiyama, R., 2000. Effects of temperature and stem length on changes in carbohydrate content in summer-grown cut chrysanthemums during development and senescence. Postharvest Biol. Technol. 20, 63–70. doi:10.1016/S0925-5214(00)00106-X

Adachi, M., Kawabata, S., Sakiyama, R., 1999. Changes in carbohydrate content in cut chrysanthemum [Dendranthema x grandiflorum (Ramat.) Kitamura] "Shuho-no-chikara" stems kept at different temperatures during anthesis and senescence. J. Japanese Soc. Hortic. Sci.

Adams, K.L., Cronn, R., Percifield, R., Wendel, J.F., 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. U. S. A. 100, 4649–54. doi:10.1073/pnas.0630618100

Agati, G., Azzarello, E., Pollastri, S., Tattini, M., 2012. Flavonoids as antioxidants in plants: Location and functional significance. Plant Sci. 196, 67–76. doi:10.1016/j.plantsci.2012.07.014

Albertin, W., Balliau, T., Brabant, P., Chèvre, A.M., Eber, F., Malosse, C., Thiellement, H., 2006. Numerous and rapid nonstochastic modifications of gene products in newly synthesized Brassica napus allotetraploids. Genetics 173, 1101–1113. doi:10.1534/genetics.106.057554

Albertin, W., Brabant, P., Catrice, O., Eber, F., Jenczewski, E., Chèvre, A.M., Thiellement, H., 2005. Autopolyploidy in cabbage (Brassica oleracea L.) does not alter significantly the proteomes of green tissues. Proteomics 5, 2131–2139. doi:10.1002/pmic.200401092

Allendorf, F.W., Danzmann, R.G., 1997. Secondary tetrasomic segregation of MDH-B and preferential pairing of homeologues in rainbow trout. Genetics 145, 1083–1092.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402.

Alvarez-Castellanos, P.P., Bishop, C.D., Pascual-Villalobos, M.J., 2001. Antifungal activity of the essential oil of flowerheads of garland chrysanthemum (Chrysanthemum coronarium) against agricultural pathogens. Phytochemistry 57, 99–102. doi:10.1016/S0031-9422(00)00461-1

Anderson, N., 2007. Chrysanthemum, in: Anderson, N.O. (Ed.), Flower Breeding and Genetics: Issues, Challenges and Opportunities for the 21st Century. Springer, Dordrecht, The Netherlands, pp. 389–437.

Araújo, W.L., Tohge, T., Ishizaki, K., Leaver, C.J., Fernie, A.R., 2011. Protein degradation - an alternative respiratory substrate for stressed plants. Trends Plant Sci. 16, 489–98. doi:10.1016/j.tplants.2011.05.008

Aubert, S., Gout, E., Bligny, R., Marty-Mazars, D., Barrieu, F., Alabouvette, J., Marty, F., Douce, R., 1996. Ultrastructural and biochemical characterization of autophagy in higher plant cells subjected to carbon deprivation: control by the supply of mitochondria with respiratory substrates. J. Cell Biol. 133, 1251–63.

Avivi, L., 1976. The effect of genes controlling different degrees of homoeologous pairing on quadrivalent frequency in induced autotetraploid lines of *Triticum longissimum*. Can. J. Genet. Cytol. 18, 357–364. doi:10.1139/g76-043

Bachem, C., Speckmann, G., van der Linde, P., Verheggen, F., Hunt, M.D., Steffens, J.C., Zabeau, M., 1994. Antisense expression of polyphenol oxidase genes inhibits enzymatic browning in potato tubers. Nat. Biotechnol. 12, 1101–1105.

Baena-González, E., Rolland, F., Thevelein, J.M., Sheen, J., 2007. A central integrator of transcription networks in plant stress and energy signalling. Nature 448, 938–42. doi:10.1038/nature06069

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z. a., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, e3376.

doi:10.1371/journal.pone.0003376

Bartoli, C.G., Guiamet, J.J., Montaldi, E.R., 1997. Ethylene production and responses to exogenous ethylene in senescing petals of Chrysanthemum morifolium RAM cv. Unsei. Plant Sci. 124, 15–21.

Bartoli, C.G., Simontacchi, M., Guiamet, J.J., Montaldi, E., Puntarulo, S., 1995. Antioxidant enzymes and lipid peroxidation during aging of chrysanthemum morifolium RAM petals. Plant Sci. 104, 161–168.

Bassil, N. V, Davis, T.M., Zhang, H., Ficklin, S., Mittmann, M., Webster, T., Mahoney, L., Wood, D., Alperin, E.S., Rosyara, U.R., Koehorst-vanc Putten, H., Monfort, A., Sargent, D.J., Amaya, I., Denoyes, B., Bianco, L., van Dijk, T., Pirani, A., Iezzoni, A., Main, D., Peace, C., Yang, Y., Whitaker, V., Verma, S., Bellon, L., Brew, F., Herrera, R., van de Weg, E., 2015. Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria × ananassa. BMC Genomics 16, 1–30. doi:10.1186/s12864-015-1310-1

Basta, A., Pavlovic, M., Couladis, M., Tzakou, O., 2007. Essential oil composition of the flowerheads of Chrysanthemum coronarium L. from Greece. Flavour Fragr. J. 22, 197–200. doi:10.1002/ffj

Bell, G.D.M., Kane, N.C., Rieseberg, L.H., Adams, K.L., 2013. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. Genome Biol. Evol. 5, 1309–1323. doi:10.1093/gbe/evt072

Bieleski, R., 1993. Fructan hydrolysis drives petal expansion in the ephemeral daylily flower. Plant Physiol. 103, 213–219.

Birchler, J.A., Yao, H., Chudalayandi, S., Vaiman, D., Veitia, R., 2010. Heterosis. Plant Cell 22, 2105–2112. doi:10.1105/tpc.110.076133

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Bomblies, K., Jones, G., Franklin, C., Zickler, D., Kleckner, N., 2016. The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role? Chromosoma 125, 287–300. doi:10.1007/s00412-015-0571-4

Bourke, P., 2014. QTL analysis in polyploids QTL analysis in polyploids: Model testing and power calculations. Wageningen University.

Bourke, P.M., Arens, P., Voorrips, R.E., Esselink, G.D., Koning-Boucoiran, C.F.S., van 't Westende, W.P.C., Santos Leonardo, T., Wissink, P., Zheng, C., van Geest, G., Visser, R.G.F., Krens, F.A., Smulders, M.J.M., Maliepaard, C., 2017. Partial preferential chromosome pairing is genotype dependent in tetraploid rose. Plant J. 90, 330–343. doi:10.1111/tpj.13496

Bourke, P.M., Voorrips, R.E., Kranenburg, T., Jansen, J., Visser, R.G., Maliepaard, C., 2016. Integrating haplotype-specific linkage maps in autotetraploid potato using SNP markers. Theor. Appl. Genet. 129, 2211–2226. doi:10.1007/s00122-016-2768-1

Bourke, P.M., Voorrips, R.E., Visser, R.G.F., Maliepaard, C., 2015. The double-reduction landscape in tetraploid potato as revealed by a high-density linkage map. Genetics 201, 853–863. doi:10.1534/genetics.115.181008

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–54.

Bradshaw, J.E., Hackett, C. a, Pande, B., Waugh, R., Bryan, G.J., 2008. QTL mapping of yield, agronomic and quality traits in tetraploid potato (Solanum tuberosum subsp. tuberosum). Theor. Appl. Genet. 116, 193–211. doi:10.1007/s00122-007-0659-1

Brent, R., 1973. Algorithms for minimizing without derivatives. Prentice-Hall, Englewood Cliffs, NJ.

Brouquisse, R., James, F., Pradet, A., Raymond, P., 1992. Asparagine metabolism and nitrogen distribution during protein degradation in sugar-starved maize root tips. Planta 188, 384–395. doi:10.1007/BF00192806

Brouquisse, R., James, F., Raymond, P., Pradet, A., 1991. Study of glucose starvation in excised maize root tips. Plant Physiol. 96, 619–26.

References

Buchanan-Wollaston, V., Page, T., Harrison, E., Breeze, E., Lim, P.O., Nam, H.G., Lin, J.-F., Wu, S.-H., Swidzinski, J., Ishizaki, K., Leaver, C.J., 2005. Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. Plant J. 42, 567–85. doi:10.1111/j.1365-313X.2005.02399.x

Calviño, M., Bruggmann, R., Messing, J., 2008. Screen of genes linked to high-sugar content in stems by comparative genomics. Rice 1, 166–176. doi:10.1007/s12284-008-9012-9

Carvalho, D.R.A., Koning-Boucoiran, C.F.S., Fanourakis, D., Vasconcelos, M.W., Carvalho, S.M.P., Heuvelink, E., Krens, F.A., Maliepaard, C., 2015. QTL analysis for stomatal functioning in tetraploid Rosa × hybrida grown at high relative air humidity and its implications on postharvest longevity. Mol. Breed. 35, 172. doi:10.1007/s11032-015-0365-7

Causse, M., Saliba-Colombani, V., Lesschaeve, I., Buret, M., 2001. Genetic analysis of organoleptic quality in fresh market tomato. 2. Mapping QTLs for sensory attributes. Theor. Appl. Genet. 102, 273–283.

Cavalli-Sforza, L., Bodmer, W., 1971. The genetics of human populations. WH Freeman & Co., San Fransisco.

Cervantes-Flores, J.C., Yencho, G.C., Kriegner, A., Pecota, K. V., Faulk, M.A., Mwanga, R.O.M., Sosinski, B.R., 2008. Development of a genetic linkage map and identification of homologous linkage groups in sweetpotato using multiple-dose AFLP markers. Mol. Breed. 21, 511–532. doi:10.1007/s11032-007-9150-6

Chagné, D., Crowhurst, R.N., Troggio, M., Davey, M.W., Gilmore, B., Lawley, C., Vanderzande, S., Hellens, R.P., Kumar, S., Cestaro, A., Velasco, R., Main, D., Rees, J.D., Iezzoni, A., Mockler, T., Wilhelm, L., van de Weg, E., Gardiner, S.E., Bassil, N., Peace, C., 2012. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One 7. doi:10.1371/journal.pone.0031745

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). Bot. Stud. 50, 43–55.

Chen, H., Boutros, P.C., 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12, 35. doi:10.1186/1471-2105-12-35

Chen, Z.J., 2007. Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. Annu. Rev. Plant Biol. 58, 377–406. doi:10.1146/annurev.arplant.58.032806.103835

Chen, F. Di, Li, F.T., Chen, S.M., Guan, Z.Y., Fang, W.M., 2009. Meiosis and pollen germinability in small-flowered anemone type chrysanthemum cultivars. Plant Syst. Evol. 280, 143–151. doi:10.1007/s00606-009-0173-9

Choi, K., Henderson, I.R., 2015. Meiotic recombination hotspots - A comparative view. Plant J. 83, 52–61. doi:10.1111/tpj.12870

Church, S.A., Spaulding, E.J., 2009. Gene expression in a wild autopolyploid sunflower series. J. Hered. 100, 491–495. doi:10.1093/jhered/esp008

Churchill, G.A., Doerge, R.W., 1994. Empirical threshold value for quantitative trait mapping. Genetics 138, 963–971. doi:10.1534/genetics.107.080101

Clevenger, J., Chavarro, C., Pearl, S.A., Ozias-Akins, P., Jackson, S.A., 2015. Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations. Mol. Plant 8, 831–846. doi:10.1016/j.molp.2015.02.002

Collier, D., 1997. Changes in respiration, protein and carbohydrates of tulip tepals and Alstroemeria petals during development. J. Plant Physiol. 150, 446–451. doi:10.1016/S0176-1617(97)80096-X

Couée, I., Sulmon, C., Gouesbet, G., El Amrani, A., 2006. Involvement of soluble sugars in reactive oxygen species balance and responses to oxidative stress in plants. J. Exp. Bot. 57, 449–59. doi:10.1093/jxb/erj027

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal Complex Syst. 1695. doi:10.1109/ICCSN.2010.34

Cuyacot, A.R., Won, S.Y., Park, S.K., Sohn, S.-H., Lee, J., Kim, J.S., Kim, H.H., Lim, K.-B., Hwang, Y.-J., 2016. The chromosomal distribution of repetitive DNA sequences in Chrysanthemum boreale revealed a characterization in its genome. Sci. Hortic. 198, 438–444. doi:10.1016/j.scienta.2015.12.025

Dabney, A., Storey, J.D., 2014. qvalue: Q-value estimation for false discovery rate control.

Dai, S.-L., Wang, W.-K., Li, M.-X., Xu, Y.-X., 2005. Phylogenetic Relationship of Dendranthema (DC.) Des Moul. Revealed by Fluorescent In Situ Hybridization. J. Integr. Plant Biol. 47, 783–791. doi:10.1111/j.1744-7909.2005.00068.x

De Backer, M., 2012. Characterization and detection of Puccinia horiana on chrysanthemum for resistance breeding and sustainable control. Ghent University.

De Jong, J., 1984. Genetic analysis in Chrysanthemum morifolium. I. Flowering time and flower number at low and optimum temperature. Euphytica 33, 455–463.

De Jong, J., Rademaker, W., 1986. The reaction of chrysanthemum cultivars to Puccinia horiana and the inheritance of resistance. Euphytica 35, 945–952.

Debener, T., Mattiesch, L., 1999. Construction of a genetic linkage map for roses using RAPD and AFLP markers. Theor. Appl. Genet. 99, 891–899. doi:10.1007/s001220051310

Devaux, C., Baldet, P., Joubès, J., Dieuaide-Noubhani, M., Just, D., Chevalier, C., Raymond, P., 2003. Physiological, biochemical and molecular analysis of sugar-starvation responses in tomato roots. J. Exp. Bot. 54, 1143–1151. doi:10.1093/jxb/erg113

Díaz-Troya, S., Pérez-Pérez, M.E., Florencio, F.J., Crespo, J.L., 2008. The role of TOR in autophagy regulation from yeast to plants and mammals. Autophagy 4, 851–865. doi:doi:10.4161/auto.6555

Dixon, R.A., Achnine, L., Kota, P., Liu, C.J., Reddy, M.S.S., Wang, L., 2002. The phenylpropanoid pathway and plant defence - A genomics perspective. Mol. Plant Pathol. 3, 371–390. doi:10.1046/j.1364-3703.2002.00131.x

Dixon, R.A., Paiva, N.L., 1995. Stress-Induced Phenylpropanoid Metabolism. Plant Cell 7, 1085–1097. doi:10.1105/tpc.7.7.1085

Doi, M., Nakagawa, Y., Watabe, S., Aoe, K., Inamoto, K., Imanishi, H., 2003. Ethylene-induced leaf yellowing in cut chrysanthemums (Dendranthema grandiflora Kitamura). J. Japanese Soc. Hortic. Sci. 72, 533–535.

Downs, C.G., Somerfield, S.D., Davey, M.C., 1997. Cytokinin treatment delays senescence but not sucrose loss in harvested broccoli. Postharvest Biol. Technol. 11, 93–100. doi:10.1016/S0925-5214(97)01419-1

Dowrick, G.J., 1953. The chromosomes of Chrysanthemum II. Garden Varieties. Heredity 7, 59–72.

Doyle, J.J., Sherman-Broyles, S., 2016. Double trouble: Taxonomy and definitions of polyploidy. New Phytol. doi:10.1111/nph.14276

Dubcovsky, J., Dvorak, J., 2007. Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. Science 316, 1862–1866. doi:10.1126/science.1143986

Eason, J.R., De Vré, L.A., Somerfield, S.D., Heyes, J.A., 1997. Physiological changes associated with Sandersonia aurantiaca flower senescence in response to sugar. Postharvest Biol. Technol. 12, 43–50.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. doi:10.1093/bioinformatics/btq461

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6, e19379. doi:10.1371/journal.pone.0019379

Endelman, J.B., Schmitz Carley, C.A., Douches, D.S., Coombs, J.J., Bizimungu, B., De Jong, W.S., Haynes, K.G., Holm, D.G., Miller, J.C., Novy, R.G., Palta, J.P., Parish, D.L., Porter, G.A., Sathuvalli, V.R., Thompson, A.L., Yencho, G.C., 2017. Pedigree Reconstruction with Genome-Wide Markers in Potato. Am. J. Potato Res. 1–7. doi:10.1007/s12230-016-9556-y

English-Loeb, G., Stout, M.J., Duffey, S.S., 1997. Drought stress in tomatoes: Changes in plant chemistry and potential nonlinear consequences for insect herbivores. Oikos 79, 456–468. doi:10.2307/3546888

Fanourakis, D., Pieruschka, R., Savvides, A., Macnish, A.J., Sarlikioti, V., Woltering, E.J., 2013. Sources of vase life variation in cut roses: A review. Postharvest Biol. Technol. 78, 1–15.

doi:10.1016/j.postharvbio.2012.12.001

Felcher, K.J., Coombs, J.J., Massa, A.N., Hansey, C.N., Hamilton, J.P., Veilleux, R.E., Buell, C.R., Douches, D.S., 2012. Integration of two diploid potato linkage maps with the potato genome sequence. PLoS One 7, 1–11. doi:10.1371/journal.pone.0036347

Ferrante, A., Trivellini, A., Scuderi, D., Romano, D., Vernieri, P., 2015. Post-production physiology and handling of ornamental potted plants. Postharvest Biol. Technol. 100, 99–108. doi:10.1016/j.postharvbio.2014.09.005

Flood, P.J., Harbinson, J., Aarts, M.G.M., 2011. Natural genetic variation in plant photosynthesis. Trends Plant Sci. 16, 327–335. doi:10.1016/j.tplants.2011.02.005

Flora Holland, 2015. Annual report Flora Holland [WWW Document]. URL http://jaarverslag.royalfloraholland.com (accessed 11.8.16).

Flowers, T.J., 2004. Improving crop salt tolerance. J. Exp. Bot. 55, 307–319. doi:10.1093/jxb/erh003

Fulton, T.M., Bucheli, P., Voirol, E., López, J., Pétiard, V., Tanksley, S.D., 2002. Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. Euphytica 127, 163–177. doi:10.1023/A:1020209930031

Ganal, M.W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E.S., Charcosset, A., Clarke, J.D., Graner, E.M., Hansen, M., Joets, J., Le Paslier, M.C., McMullen, M.D., Montalent, P., Rose, M., Schön, C.C., Sun, Q., Walter, H., Martin, O.C., Falque, M., 2011. A large maize (Zea mays L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6. doi:10.1371/journal.pone.0028334

Gar, O., Sargent, D.J., Tsai, C.J., Pleban, T., Shalev, G., Byrne, D.H., Zamir, D., 2011. An autotetraploid linkage map of rose (Rosa hybrida) validated using the strawberry (Fragaria vesca) genome sequence. PLoS One 6. doi:10.1371/journal.pone.0020463

Gary, C., Baldet, P., Bertin, N., Devaux, C., Tchamitchian, M., Raymond, P., 2003. Time-course of Tomato Whole-plant Respiration and Fruit and Stem Growth During Prolonged Darkness in Relation to Carbohydrate Reserves. Ann. Bot. 91, 429–438. doi:10.1093/aob/mcg037

Gaufichon, L., Reisdorf-Cren, M., Rothstein, S.J., Chardon, F., Suzuki, A., 2010. Biological functions of asparagine synthetase in plants. Plant Sci. 179, 141–153. doi:10.1016/j.plantsci.2010.04.010

Gillies, A.C., Pilar, C., Coen, E.S., Richard, A., 2002. Making rays in the Asteraceae: Genetics and evolution of radiate versus discoid flower heads, in: Cronk, Q., Bateman, R., Hawkins, J. (Eds.), Developmental Genetics and Plant Evolution. Taylor & Francis, London, pp. 233–246. doi:doi:10.1201/9781420024982.ch12

Gout, E., Rebeille, F., Douce, R., Bligny, R., 2014. Interplay of Mg2+, ADP, and ATP in the cytosol and mitochondria: Unravelling the role of Mg2+ in cell respiration. Proc. Natl. Acad. Sci. 111, E4560–E4567. doi:10.1073/pnas.1406251111

Grandke, F., Ranganathan, S., van Bers, N., de Haan, J.R., Metzler, D., 2017. PERGOLA: fast and deterministic linkage mapping of polyploids. BMC Bioinformatics 18, 12. doi:10.1186/s12859-016-1416-8

Grandke, F., Singh, P., Heuven, H.C.M., de Haan, J.R., Metzler, D., 2016. Advantages of continuous genotype values over genotype classes for GWAS in higher polyploids: a comparative study in hexaploid chrysanthemum. BMC Genomics 17, 672. doi:10.1186/s12864-016-2926-5

Grattapaglia, D., Sederoff, R., 1994. Genetic linkage maps. Genetics 137, 1121–1137.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. doi:10.1038/nprot.2013.084

Hackett, C.A., Boskamp, B., Vogogias, T., Preedy, K.F., Milne, I., 2017. TetraploidSNPMap: software for linkage

analysis and QTL mapping in autotetraploid populations using SNP dosage data. J. Hered. 108, 438–442. doi:10.1093/jhered/esx022

Hackett, C.A., Bradshaw, J.E., Bryan, G.J., 2014. QTL mapping in autotetraploids using SNP dosage information. Theor. Appl. Genet. 127, 1885–1904. doi:10.1007/s00122-014-2347-2

Hackett, C.A., Bradshaw, J.E., Meyer, R.C., McNicol, J.W., Milbourne, D., Waught, R., 1998. Linkage analysis in tetraploid species: a simulation study. Genet. Res. 71, 143–153. doi:10.1017/S0016672398003188

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, e63939. doi:10.1371/journal.pone.0063939

Hackett, C.A., Milne, I., Bradshaw, J.E., Luo, Z., 2007. TetraploidMap for Windows: Linkage map construction and QTL mapping in autotetraploid species. J. Hered. 98, 727–729. doi:10.1093/jhered/esm086

Hackett, C., Luo, Z., 2003. TetraploidMap: construction of a linkage map in autotetraploid species. J. Hered. 94, 358–359. doi:10.1093/jhered/esg066

Heap, G.A., Yang, J.H.M., Downes, K., Healy, B.C., Hunt, K.A., Bockett, N., Franke, L., Dubois, P.C., Mein, C.A., Dobson, R.J., Albert, T.J., Rodesch, M.J., Clayton, D.G., Todd, J.A., van Heel, D.A., Plagnol, V., 2009. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. Hum. Mol. Genet. 19, 122–134. doi:10.1093/hmg/ddp473

Hendry, G., 1993. Evolutionary origins and natural functions of fructans–a climatological, biogeographic and mechanistic appraisal. New Phytol. 123, 3–14.

Horie, Y., Ito, H., Kusaba, M., Tanaka, R., Tanaka, A., 2009. Participation of chlorophyll b reductase in the initial step of the degradation of light-harvesting chlorophyll a/b-protein complexes in Arabidopsis. J. Biol. Chem. 284, 17449–17456. doi:10.1074/jbc.M109.008912

Ichimura, K., Kishimoto, M., Kawabata, Y., Yamada, K., 2005. Soluble carbohydrates and variation in vase-life of cut rose cultivars "Delilah" and "Sonia." J. Hortic. Sci. Biotechnol. 80, 280–286.

Irving, D.E., Joyce, D.C., 1995. Sucrose supply can increase longevity of broccoli (Brassica oleracea) branchlets kept at 22C. Plant Growth Regul. 17, 251–256. doi:10.1007/BF00024733

Jackson, S., Chen, Z.J., 2010. Genomic and expression plasticity of polyploidy. Curr. Opin. Plant Biol. 13, 153–159. doi:10.1016/j.pbi.2009.11.004

Jamann, T.M., Sood, S., Wisser, R.J., Holland, J.B., 2017. High-Throughput Resequencing of Maize Landraces at Genomic Regions Associated with Flowering Time. PLoS One 12, e0168910. doi:10.1371/journal.pone.0168910

Jibran, R., Sullivan, K.L., Crowhurst, R., Erridge, Z.A., Chagné, D., McLachlan, A.R.G., Brummell, D.A., Dijkwel, P.P., Hunter, D.A., 2015. Staying green postharvest: how three mutations in the Arabidopsis chlorophyll b reductase gene NYC1 delay degreening by distinct mechanisms. J. Exp. Bot. 66, 6849–6862. doi:10.1093/jxb/erv390

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W.E., Tuteja, R., Spillane, C., Robinson, S.J., Links, M.G., Clarke, C., Higgins, E.E., Huebert, T., Sharpe, A.G., Parkin, I.A.P., 2014. The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. Nat. Commun. 5, 3706. doi:10.1038/ncomms4706

Ke, D., Saltveit, M.E., 1988. Plant hormone interaction and phenolic metabolism in the regulation of russet spotting in iceberg lettuce. Plant Physiol. 88, 1136–1140. doi:10.1104/pp.88.4.1136

Kempthorne, O., 1957. An introduction to genetic statistics. John Wiley & Sons, Inc., New York.

Kim, H.K., Choi, Y.H., Verpoorte, R., 2010a. NMR-based metabolomic analysis of plants. Nat. Protoc. 5, 536–49. doi:10.1038/nprot.2009.237

Kim, H.K., Saifullah, Khan, S., Wilson, E.G., Kricun, S.D.P., Meissner, A., Goraler, S., Deelder, A.M., Choi, Y.H., Verpoorte, R., 2010b. Metabolic classification of South American Ilex species by NMR-based metabolomics. Phytochemistry 71, 773–784. doi:10.1016/j.phytochem.2010.02.001

References

Kim, K.J., Kim, Y.H., Yu, H.H., Jeong, S. Il, Cha, J.D., Kil, B.S., You, Y.O., 2003. Antibacterial activity and chemical composition of essential oil of Chrysanthemum boreale. Planta Med. 69, 274–277. doi:10.1055/s-2003-38479

Kim, S., Koo, B., Kim, J., Liu, J., 2007. Metabolic discrimination of sucrose starvation from Arabidopsis cell suspension by 1H NMR based metabolomics. Biotechnol. Bioprocess Eng. 12, 653–661.

King, G.A., Woollard, D.C., Irving, D.E., Borst, W.M., 1990. Physiological changes in asparagus spear tips after harvest. Physiol. Plant. 80, 393–400. doi:10.1111/j.1399-3054.1990.tb00058.x

Klee, H.J., 2010. Improving the flavor of fresh fruits: Genomics, biochemistry, and biotechnology. New Phytol. 187, 44–56. doi:10.1111/j.1469-8137.2010.03281.x

Klie, M., Menz, I., Linde, M., Debener, T., 2016. Strigolactone pathway genes and plant architecture: association analysis and QTL detection for horticultural traits in chrysanthemum. Mol. Genet. Genomics 291, 957–969. doi:10.1007/s00438-015-1155-y

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. Front. Plant Sci. 5, 1–8. doi:10.3389/fpls.2014.00479

Koning-Boucoiran, C.F.S., Esselink, G.D., Vukosavljev, M., van 't Westende, W.P.C., Gitonga, V.W., Krens, F.A., Voorrips, R.E., van de Weg, W.E., Schulz, D., Debener, T., Maliepaard, C., Arens, P., Smulders, M.J.M., 2015. Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (Rosa L.). Front. Plant Sci. 6, 249. doi:10.3389/fpls.2015.00249

Kriegner, A., Cervantes, J., Burg, K., Mwanga, R., Zhang, D., 2003. A genetic linkage map of sweetpotato [Ipomoea batatas (L.) Lam.] based on AFLP markers. Mol. Breed. 11, 169–185. doi:10.1023/A:1022870917230

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. doi:10.1038/nmeth.1923

Langton, F.A., 1989. Inheritance in Chrysanthemum morifolium Ramat. Heredity 62, 419–423.

Leach, L.J., Belfield, E.J., Jiang, C., Brown, C., Mithani, A., Harberd, N.P., 2014. Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. BMC Genomics 15, 276. doi:10.1186/1471-2164-15-276

Leal-Bertioli, S., Shirasawa, K., Abernathy, B., Moretzsohn, M., Chavarro, C., Clevenger, J., Ozias-Akins, P., Jackson, S., Bertioli, D., 2015. Tetrasomic recombination is surprisingly frequent in allotetraploid Arachis. Genetics 199, 1093–1105. doi:10.1534/genetics.115.174607

Lee, J.H., Jang, M., Seo, J., Kim, G.H., 2011. Evaluation For Antibacterial Effects Of Volatile Flavors From Chrysanthemum Indicum Against Food-Borne Pathogens And Food Spoilage Bacteria. J. Food Saf. 31, 140–148. doi:10.1111/j.1745-4565.2010.00277.x

Levy, A. a, Feldman, M., 2004. Biological relevance of polyploidy : ecology to genomics Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. Biol. J. Linn. Soc. 82, 607–613.

Li, C., Chen, S., Chen, F., Li, J., Fang, W., 2011. Cytogenetic Study of Three Edible Chrysanthemum Cultivars. Russ. J. Genet. 47, 176–181. doi:10.1134/S1022795411010054

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. doi:10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Li, X., Han, Y., Wei, Y., Acharya, A., Farmer, A.D., Ho, J., Monteros, M.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. doi:10.1371/journal.pone.0084329

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y., Gao, C., 2017. Efficient

DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nat. Commun. 8. doi:10.1038/ncomms14261

Lin, H.X., Zhu, M.Z., Yano, M., Gao, J.P., Liang, Z.W., Su, W.A., Hu, X.H., Ren, Z.H., Chao, D.Y., 2004. QTLs for Na+ and K+ uptake of the shoots and roots controlling rice salt tolerance. Theor. Appl. Genet. 108, 253–260. doi:10.1007/s00122-003-1421-y

Lindhout, P., Meijer, D., Schotte, T., Hutten, R.C.B., Visser, R.G.F., van Eck, H.J., 2011. Towards F 1 Hybrid Seed Potato Breeding. Potato Res. 54, 301–312. doi:10.1007/s11540-011-9196-z

Lindqvist, H., Gastelo, M., Pérez, W., Forbes, G., De Koeyer, D., Bonierbale, M., 2014. Phenotypic stability and genome wide association study of late blight resistance in potato genotypes adapted to the tropical highlands. Phytopathology 1–47. doi:10.1094/PHYTO-10-13-0270-R

Liu, X., Guo, L., You, J., Liu, X., He, Y., Yuan, J., Liu, G., Feng, Z., 2010. Progress of Segregation Distortion in Genetic Mapping of Plants. Res. J. Agron. 4, 78–83. doi:10.3923/rjagr.2010.78.83

Luo, Z.W., Hackett, C. a, Bradshaw, J.E., McNicol, J.W., Milbourne, D., 2001. Construction of a genetic linkage map in tetraploid species using molecular markers. Genetics 157, 1369–85.

Macnish, A., Leonard, R., Borda, A., 2010. Genotypic Variation in the Postharvest Performance and Ethylene Sensitivity of Cut Rose Flowers. HortScience 45, 790–796.

Manrique-Carpintero, N.C., Coombs, J.J., Veilleux, R.E., Buell, C.R., Douches, D.S., 2016. Comparative Analysis of Regions with Distorted Segregation in Three Diploid Populations of Potato. G3 Genes|Genomes|Genetics 6, 2617–2628. doi:10.1534/G3.116.030031

Marcelis, L.F., 1996. Sink strength as a determinant of dry matter partitioning in the whole plant. J. Exp. Bot. 47, 1281–1291. doi:10.1093/jxb/47.Special_Issue.1281

Mardia, K. V., Kent, J.T., Bibby, J.M., 1979. Multivariate Analysis. Academic Press, London.

Markham, K.R., Ryan, K.G., Bloor, S.J., Mitchell, K. a., 1998. An increase in the luteolin: Apigenin ratio in Marchantia polymorpha on UV-B enhancement. Phytochemistry 48, 791–794. doi:10.1016/S0031-9422(97)00875-3

Massa, A.N., Manrique-Carpintero, N.C., Coombs, J.J., Zarka, D.G., Boone, A.E., Kirk, W.W., Hackett, C.A., Bryan, G.J., Douches, D.S., 2015. Genetic Linkage Mapping of Economically Important Traits in Cultivated Tetraploid Potato (Solanum tuberosum L.). G3 Genes|Genomes|Genetics 5, 2357–2364. doi:10.1534/g3.115.019646

McCallum, S., Graham, J., Jorgensen, L., Rowland, L.J., Bassil, N. V., Hancock, J.F., Wheeler, E.J., Vining, K., Poland, J.A., Olmstead, J.W., Buck, E., Wiedow, C., Jackson, E., Brown, A., Hackett, C.A., 2016. Construction of a SNP and SSR linkage map in autotetraploid blueberry using genotyping by sequencing. Mol. Breed. 36, 1–24. doi:10.1007/s11032-016-0443-5

McKinney, G.J., Waples, R.K., Seeb, L.W., Seeb, J.E., 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Mol. Ecol. Resour. 17, 656–669. doi:10.1111/1755-0998.12613

Merriam-Webster Online, 2015. Senescence [Def. 2]. Merriam-Webster.

Meyer, R.C., Milbourne, D., Hackett, C. a, Bradshaw, J.E., McNichol, J.W., Waugh, R., 1998. Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (Phytophthora infestans). Mol. Gen. Genet. 259, 150–60.

Monden, Y., Hara, T., Okada, Y., Jahana, O., Kobayashi, A., Tabuchi, H., Onaga, S., Tahara, M., 2015. Construction of a linkage map based on retrotransposon insertion polymorphisms in sweetpotato via high-throughput sequencing. Breed. Sci. 65, 145–153. doi:10.1270/jsbbs.65.145

Moreno, E., Obando, J.M., Dos-Santos, N., Fernández-Trujillo, J.P., Monforte, A.J., Garcia-Mas, J., 2008. Candidate genes and QTLs for fruit ripening and softening in melon. Theor. Appl. Genet. 116, 589–602. doi:10.1007/s00122-007-0694-y

Morkunas, I., Garnczarska, M., Bednarski, W., Ratajczak, W., Waplak, S., 2003. Metabolic and ultrastructural

responses of lupine embryo axes to sugar starvation. J. Plant Physiol. 160, 311–319. doi:10.1078/0176-1617-00696

Munns, R., James, R.A., Xu, B., Athman, A., Conn, S.J., Jordans, C., Byrt, C.S., Hare, R.A., Tyerman, S.D., Tester, M., Plett, D., Gilliham, M., 2012. Wheat grain yield on saline soils is improved by an ancestral Na$^+$ transporter gene. Nat. Biotechnol. 30, 360–4. doi:10.1038/nbt.2120

Muraya, M.M., Schmutzer, T., Ulpinnis, C., Scholz, U., Altmann, T., 2015. Targeted Sequencing Reveals Large-Scale Sequence Polymorphism in Maize Candidate Genes for Biomass Production and Composition. PLoS One 10, e0132120. doi:10.1371/journal.pone.0132120

Murray, S.C., Sharma, A., Rooney, W.L., Klein, P.E., Mullet, J.E., Mitchell, S.E., Kresovich, S., 2008. Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. Crop Sci. 48, 2165–2179. doi:10.2135/cropsci2008.01.0016

Nägele, T., Henkel, S., Hörmiller, I., Sauter, T., Sawodny, O., Ederer, M., Heyer, A.G., 2010. Mathematical modeling of the central carbohydrate metabolism in Arabidopsis reveals a substantial regulatory influence of vacuolar invertase on whole plant carbon metabolism. Plant Physiol. 153, 260–272. doi:10.1104/pp.110.154443

Narumi, T., Kanno, Y., Suzuki, M., Kishimoto, S., Ohmiya, A., Satoh, S., 2005. Cloning of a cDNA encoding an ethylene receptor (DG-ERS1) from chrysanthemum and comparison of its mRNA level in ethylene-sensitive and -insensitive cultivars. Postharvest Biol. Technol. 36, 21–30. doi:10.1016/j.postharvbio.2004.11.001

Nguepjop, J.R., Tossim, H.-A., Bell, J.M., Rami, J.-F., Sharma, S., Courtois, B., Mallikarjuna, N., Sane, D., Fonceka, D., 2016. Evidence of Genomic Exchanges between Homeologous Chromosomes in a Cross of Peanut with Newly Synthetized Allotetraploid Hybrids. Front. Plant Sci. 7, 1–12. doi:10.3389/fpls.2016.01635

Nijsse, J., 2001. Functional anatomy of the water transport system in cut chrysanthemum. Wageningen University.

Nijveen, H., van Kaauwen, M., Esselink, D.G., Hoegen, B., Vosman, B., 2013. QualitySNPng: a user-friendly SNP detection and visualization tool. Nucleic Acids Res. 41, 587–590. doi:10.1093/nar/gkt333

Onozaki, T., Ikeda, H., Yamaguchi, T., 2001. Genetic improvement of vase life of carnation flowers by crossing and selection. Sci. Hortic. 87, 107–120. doi:10.1016/S0304-4238(00)00167-9

Osborn, T.C., Chris Pires, J., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.S., Comai, L., Madlung, A., Doerge, R.W., Colot, V., Martienssen, R.A., 2003. Understanding mechanisms of novel gene expression in polyploids. Trends Genet. 19, 141–147. doi:10.1016/S0168-9525(03)00015-5

Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34, 401–437. doi:10.1146/annurev.genet.34.1.401

Paris, M.K., Zalapa, J.E., McCreight, J.D., Staub, J.E., 2008. Genetic dissection of fruit quality components in melon (Cucumis melo L.) using a RIL population derived from exotic x elite US Western Shipping germplasm. Mol. Breed. 22, 405–419. doi:10.1007/s11032-008-9185-3

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. Sci. Hortic. 192, 80–88. doi:10.1016/j.scienta.2015.05.009

Peace, C.P., 2017. DNA-informed breeding of rosaceous crops: promises, progress and prospects. Hortic. Res. 4, 17006. doi:10.1038/hortres.2017.6

Plaxton, W.C., Podestá, F.E., 2007. The Functional Organization and Control of Plant Respiration. Crit. Rev. Plant Sci. 25, 159–198. doi:10.1080/07352680600563876

Pourcel, L., Routaboul, J.-M., Cheynier, V., Lepiniec, L., Debeaujon, I., 2007. Flavonoid oxidation in plants: from biochemical properties to physiological functions. Trends Plant Sci. 12, 29–36. doi:10.1016/j.tplants.2006.11.006

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. Theor. Appl. Genet. 129, 2117–

2132. doi:10.1007/s00122-016-2761-8

Preiss, J., 2009. Biochemistry and Molecular Biology of Starch Biosynthesis, in: Starch: Chemistry and Technology. Elsevier Inc., pp. 83–148. doi:10.1016/B978-0-12-746275-2.00004-5

Put, H.M.C., 1990. Micro-organisms from freshly harvested cut flower stems and developing during the vase life of chrysanthemum, gerbera and rose cultivars. Sci. Hortic. 43, 129–144. doi:10.1016/0304-4238(90)90044-F

Qu, L., Hancock, J.F., 2001. Detecting and mapping repulsion-phase linkage in polyploids with polysomic inheritance. Theor. Appl. Genet. 103, 136–143. doi:10.1007/s001220100647

R Core Team, 2014. R: A language and environment for statistical computing.

Raessler, M., 2011. Sample preparation and current applications of liquid chromatography for the determination of non-structural carbohydrates in plants. Trends Anal. Chem. 30, 1833–1843. doi:10.1016/j.trac.2011.06.013

Rak, K., Bethke, P.C., Palta, J.P., 2017. QTL mapping of potato chip color and tuber traits within an autotetraploid family. Mol. Breed. 37, 15. doi:10.1007/s11032-017-0619-7

Ramsey, J., Schemske, D.W., 2002. Neopolyploidy in Flowering Plants. Annu. Rev. Ecol. Syst. 33, 589–639. doi:10.1146/annurev.ecolsys.33.010802.150437

Ramsey, J., Schemske, D.W., 1998. Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. Annu. Rev. Ecol. Syst. 29, 467–501. doi:10.1146/annurev.ecolsys.29.1.467

Revolution Analytics, Weston, S., 2014a. doParallel: Foreach parallel adaptor for the parallel package.

Revolution Analytics, Weston, S., 2014b. doSNOW: Foreach parallel adaptor for the snow package.

Reyes-Arribas, T., Barrett, J.E., Huber, D.J., Nell, T. a., Clark, D.G., 2001. Leaf senescence in a non-yellowing cultivar of chrysanthemum (Dendranthema grandiflora). Physiol. Plant. 111, 540–544.

Rhoads, D.M., Umbach, A.L., Subbaiah, C.C., Siedow, J.N., 2006. Mitochondrial reactive oxygen species. Contribution to oxidative stress and interorganellar signaling. Plant Physiol. 141, 357–366. doi:10.1104/pp.106.079129

Rico, D., Martín-Diana, A.B., Barat, J.M., Barry-Ryan, C., 2007. Extending and measuring the quality of fresh-cut fruit and vegetables: a review. Trends Food Sci. Technol. 18, 373–386. doi:10.1016/j.tifs.2007.03.011

Ritsema, T., Smeekens, S., 2003. Fructans: beneficial for plants and humans. Curr. Opin. Plant Biol. 6, 223–230. doi:10.1016/S1369-5266(03)00034-7

Robins, J.G., Hansen, J.L., Viands, D.R., Brummer, E.C., 2008. Genetic mapping of persistence in tetraploid alfalfa. Crop Sci. 48, 1780–1786. doi:10.2135/cropsci2008.02.0101

Rolny, N., Costa, L., Carrión, C., Guiamet, J.J., 2011. Is the electrolyte leakage assay an unequivocal test of membrane deterioration during leaf senescence? Plant Physiol. Biochem. 49, 1220–1227. doi:10.1016/j.plaphy.2011.06.010

Rosyara, U.R., De Jong, W.S., Douches, D.S., Endelman, J.B., 2016. Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. Plant Genome 9, 1–10. doi:10.3835/plantgenome2015.08.0073

Roxas, N.J.L., Tashiro, Y., Miyazaki, S., Isshiki, S., Takeshita, A., 1995. Meiosis and pollen fertility in Higo chrysanthemum (Dendranthema x grandiflorum (Ramat.) Kitam.). J. Japanese Soc. Hortic. Sci. 64, 161–168.

Saintenac, C., Jiang, D., Akhunov, E.D., 2011. Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol. 12, R88. doi:10.1186/gb-2011-12-9-r88

Sakuraba, Y., Schelbert, S., Park, S.Y., Han, S.H., Lee, B.D., Andres, C.B., Kessler, F., Hortensteiner, S., Paek, N.C., 2012. STAY-GREEN and Chlorophyll Catabolic Enzymes Interact at Light-Harvesting Complex II for Chlorophyll Detoxification during Leaf Senescence in Arabidopsis. Plant Cell 24, 507–518.

doi:10.1105/tpc.111.089474

Saliba-Colombani, V., Causse, M., Langlois, D., Philouze, J., Buret, M., 2001. Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. Theor. Appl. Genet. 102, 273–283.

Santos, J.L., Alfaro, D., Sanchez-Moran, E., Armstrong, S.J., Franklin, F.C.H., Jones, G.H., 2003. Partial Diploidization of Meiosis in Autotetraploid Arabidopsis thaliana. Genetics 165, 1533–1540.

Satoh, S., Watanabe, M., Chisaka, K., Narumi, T., 2008. Suppressed Leaf Senescence in Chrysanthemum Transformed with a Mutated Ethylene Receptor Gene mDG-ERS1 (etr1-4). J. Plant Biol. 51, 424–427.

Schaart, J.G., van de Wiel, C.C.M., Lotz, L.A.P., Smulders, M.J.M., 2016. Opportunities for Products of New Plant Breeding Techniques. Trends Plant Sci. 21, 438–449. doi:10.1016/j.tplants.2015.11.006

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. Theor. Appl. Genet. 130, 717–726. doi:10.1007/s00122-016-2845-5

Schneider, C. a, Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675. doi:10.1038/nmeth.2089

Schönhals, E., 2014. Identifying novel diagnostic SNP markers for potato (Solanum tuberosum L .) tuber starch and yield by association mapping. Universität Köln.

Schouten, R.E., Tijskens, L.M.., van Kooten, O., 2002. Predicting keeping quality of batches of cucumber fruit based on a physiological mechanism. Postharvest Biol. Technol. 26, 209–220. doi:10.1016/S0925-5214(02)00017-0

Serang, O., Mollinari, M., Garcia, A.A.F., 2012. Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. PLoS One 7, e30906. doi:10.1371/journal.pone.0030906

Shahin, A., Arens, P., Van Heusden, A.W., Van Der Linden, G., Van Kaauwen, M., Khan, N., Schouten, H.J., Van De Weg, W.E., Visser, R.G.F., Van Tuyl, J.M., 2011. Genetic mapping in Lilium: mapping of major genes and quantitative trait loci for several ornamental traits and disease resistances. Plant Breed. 130, 372–382. doi:10.1111/j.1439-0523.2010.01812.x

Shahin, A., van Gurp, T., Peters, S. a, Visser, R.G., van Tuyl, J.M., Arens, P., 2012. SNP markers retrieval for a non-model species: a practical approach. BMC Res. Notes 5, 79. doi:10.1186/1756-0500-5-79

Shirasawa, K., Tanaka, M., Takahata, Y., Ma, D., Cao, Q., Liu, Q., Zhai, H., Kwak, S.-S., Cheol Jeong, J., Yoon, U.-H., Lee, H.-U., Hirakawa, H., Isobe, S., 2017. A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweetpotato (Ipomoea batatas). Sci. Rep. 7, 44207. doi:10.1038/srep44207

Shunying, Z., Yang, Y., Huaidong, Y., Yue, Y., Guolin, Z., 2005. Chemical composition and antimicrobial activity of the essential oils of Chrysanthemum indicum. J. Ethnopharmacol. 96, 151–158. doi:10.1016/j.jep.2004.08.031

Smeekens, S., Ma, J., Hanson, J., Rolland, F., 2010. Sugar signals and molecular networks controlling plant growth. Curr. Opin. Plant Biol. 13, 274–279. doi:10.1016/j.pbi.2009.12.002

Song, K., Lu, P., Tang, K., Osborn, T.C., 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc. Natl. Acad. Sci. U. S. A. 92, 7719–23. doi:10.1073/pnas.92.17.7719

Spaargaren, J., 2015. Origin and Spreading of the Cultivated Chrysanthemum.

Stam, P., 1993. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J. 3, 739–744. doi:10.1111/j.1365-313X.1993.00739.x

Stebbins, G.L., 1947. Types of Polyploids: Their Classification and Significance. Adv. Genet. 1, 403–429. doi:10.1016/S0065-2660(08)60490-3

Stebbins, G.L., 1940. The significance of polyploidy in plant evolution. Am. Nat. 74, 54–66. doi:10.1007/s12134-

012-0237-5

Stevens, L., Davelaar, E., 1996. Isolation and characterization of blackspot pigments from potato tubers. Phytochemistry 42, 941–947.

Stevens, M.A., 1986. Inheritance of Tomato Fruit Quality Components. Plant Breed. Rev. 4, 273–311.

Stickland, R.G., 1972. Changes in anthocyanin, carotenoid, chlorophyll, and protein in developing florets of the chrysanthemum. Ann. Bot. 36, 459–469.

Stift, M., Berenos, C., Kuperus, P., van Tienderen, P.H., 2008. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to Rorippa (yellow cress) microsatellite data. Genetics 179, 2113–23. doi:10.1534/genetics.107.085027

Sybenga, J., 1995. Meiotic pairing in autohexaploid Lathyrus: a mathematical model. Heredity 75, 343–350. doi:10.1038/hdy.1995.145

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., Innan, H., Cano, L.M., Kamoun, S., Terauchi, R., 2013. QTL-seq: Rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 74, 174–183. doi:10.1111/tpj.12105

Tang, N., van der Lee, T., Shahin, A., Holdinga, M., Bijman, P., Caser, M., Visser, R.G.F., van Tuyl, J.M., Arens, P., 2015. Genetic mapping of resistance to Fusarium oxysporum f. sp. tulipae in tulip. Mol. Breed. 35, 1–17. doi:10.1007/s11032-015-0316-3

Thomas, H., Howarth, C.J., 2000. Five ways to stay green. J. Exp. Bot. 51, 329–337. doi:10.1093/jexbot/51.suppl_1.329

Thomas, H., Ougham, H., 2014. The stay-green trait. J. Exp. Bot. 65, 3889–3900. doi:10.1093/jxb/eru037

Thompson, J.E., Froese, C.D., Madey, E., Smith, M.D., Hong, Y., 1998. Lipid metabolism during plant senescence. Prog. Lipid Res. 37, 119–41.

Thomson, M.J., Ocampo, M., Egdane, J., Rahman, M.A., Sajise, A.G., Adorada, D.L., Tumimbang-Raiz, E., Blumwald, E., Seraj, Z.I., Singh, R.K., Gregorio, G.B., Ismail, A.M., de Ocampo, M., Egdane, J., Rahman, M.A., Sajise, A.G., Adorada, D.L., Tumimbang-Raiz, E., Blumwald, E., Seraj, Z.I., Singh, R.K., Gregorio, G.B., Ismail, A.M., 2010. Characterizing the Saltol Quantitative Trait Locus for Salinity Tolerance in Rice. Rice 3, 148–160. doi:10.1007/s12284-010-9053-8

Thornton-Wells, T.A., Moore, J.H., Haines, J.L., 2006. Dissecting trait heterogeneity: a comparison of three clustering methods applied to genotypic data. BMC Bioinformatics 7, 204. doi:10.1186/1471-2105-7-204

Tiwari, B., Belenghi, B., Levine, A., 2002. Oxidative stress increased respiration and generation of reactive oxygen species, resulting in ATP depletion, opening of mitochondrial permeability transition, and programmed cell death. Plant Physiol. 128, 1271–1281. doi:10.1104/pp.010999.1

Trivellini, A., Jibran, R., Watson, L.M., O'Donoghue, E.M., Ferrante, A., Sullivan, K.L., Dijkwel, P.P., Hunter, D.A., 2012. Carbon deprivation-driven transcriptome reprogramming in detached developmentally arresting Arabidopsis inflorescences. Plant Physiol. 160, 1357–72. doi:10.1104/pp.112.203083

Truco, M.J., Ashrafi, H., Kozik, A., van Leeuwen, H., Bowers, J., Reyes Chin Wo, S., Stoffel, K., Xu, H., Hill, T., Van Deynze, A., Michelmore, R.W., 2013. An Ultra High-Density, Transcript-Based, Genetic Map of Lettuce. Genes|Genomes|Genetics 3, 617–631. doi:10.1534/g3.112.004929

Trusty, S., Miller, W., 1991. Postproduction carbohydrate levels in pot chrysanthemums. J. Am. Soc. Hortic. Sci. 116, 1013–1018.

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, 10–14. doi:10.1371/journal.pone.0062355

Ukoskit, K., Thompson, P., 1997. Autopolyploidy versus allopolyploidy and low-density randomly amplified polymorphic DNA linkage maps of sweetpotato. J. Am. Soc. Hortic. Sci. 122, 822–828.

References

van der Graaff, E., Schwacke, R., Schneider, A., Desimone, M., Flügge, U.-I., Kunze, R., 2006. Transcription analysis of arabidopsis membrane transporters and hormone pathways during developmental and induced leaf senescence. Plant Physiol. 141, 776–792. doi:10.1104/pp.106.079293

van Dijk, T., Noordijk, Y., Dubos, T., Bink, M.C., Meulenbroek, B.J., Visser, R.G., van de Weg, E., 2012. Microsatellite allele dose and configuration establishment (MADCE): an integrated approach for genetic studies in allopolyploids. BMC Plant Biol. 12, 25. doi:10.1186/1471-2229-12-25

van Dijk, T., Pagliarani, G., Pikunova, A., Noordijk, Y., Yilmaz-Temel, H., Meulenbroek, B., Visser, R.G., van de Weg, E., 2014. Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map. BMC Plant Biol. 14, 55. doi:10.1186/1471-2229-14-55

van Doorn, W.G., Cruz, P., 2000. Evidence for a wounding-induced xylem occlusion in stems of cut chrysanthemum flowers. Postharvest Biol. Technol. 19, 73–83. doi:10.1016/S0925-5214(00)00069-7

van Doorn, W.G., Van Meeteren, U., 2003. Flower opening and closure: a review. J. Exp. Bot. 54, 1801–12. doi:10.1093/jxb/erg213

van Geest, G., van Meeteren, U., Arens, P., 2015. Can Phenotyping for Water Balance Improve Breeding for Vase Life? Acta Hortic. 1087, 149–154.

Van Ieperen, W., Nijsse, J., Keijzer, C.J., Van Meeteren, U., 2001. Induction of air embolism in xylem conduits of pre-defined diameter. J. Exp. Bot. 52, 981–91.

van Meeteren, U., 1992. Role of air embolism and low water temperature in water balance of cut chrysanthemum flowers. Sci. Hortic. 51, 275–284. doi:10.1016/0304-4238(92)90125-V

Van Meeteren, U., Van Gelder, A., Van Ieperen, W., 2005. Effect of growth conditions on post harvest rehydration ability of cut chrysanthemum flowers. Acta Hortic. 669, 287–296.

Van Meeteren, U., Van Gelder, H., 1999. Effect of time since harvest and handling conditions on rehydration ability of cut chrysanthemum flowers. Postharvest Biol. Technol. 16, 169–177. doi:10.1016/S0925-5214(99)00020-4

Van Meeteren, U., Van Gelder, H., Van Ieperen, W., 2000. Reconsideration of the use of deionized water as vase water in postharvest experiments on cut flowers. Postharvest Biol. Technol. 2, 175–187. doi:10.1016/S0925-5214(99)00050-2

Van Ooijen, J., 2006. Joinmap ® 4: Software for the calculation of genetic linkage maps in experimental populations.

Van Ooijen, J.W., Jansen, J., 2013. Genetic Mapping in Experimental Populations. doi:10.1017/CBO9781139003889

Vanlerberghe, G.C., 2013. Alternative Oxidase: A Mitochondrial Respiratory Pathway to Maintain Metabolic and Signaling Homeostasis during Abiotic and Biotic Stress in Plants. Int. J. Mol. Sci. 14, 6805–47. doi:10.3390/ijms14046805

Vergauwen, R., Van den Ende, W., Van Laere, a, 2000. The role of fructan in flowering of Campanula rapunculoides. J. Exp. Bot. 51, 1261–6.

Verspreet, J., Pollet, A., Cuyvers, S., Vergauwen, R., Van den Ende, W., Delcour, J.A., Courtin, C.M., 2012. A simple and accurate method for determining wheat grain fructan content and average degree of polymerization. J. Agric. Food Chem. 60, 2102–7. doi:10.1021/jf204774n

Vogelstein, B., Kinzler, K.W., Yan, H., Yuan, W., Velculescu, V.E., 2002. Allelic Variation in Human Gene Expression. Science 297, 1143. doi:10.1126/science.1072545

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

Voorrips, R.E., Maliepaard, C., 2012. The simulation of meiosis in diploid and tetraploid organisms using various genetic models. BMC Bioinformatics 13, 248. doi:10.1186/1471-2105-13-248

Vos, P.G., Paulo, M.J., Voorrips, R.E., Visser, R.G.F., van Eck, H.J., van Eeuwijk, F.A., 2017. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theor. Appl. Genet. 130, 123–135. doi:10.1007/s00122-016-2798-8

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. Theor. Appl. Genet. 128, 2387–2401. doi:10.1007/s00122-015-2593-y

Vukosavljev, M., Arens, P., Voorrips, R.E., van 't Westende, W.P., Esselink, G., Bourke, P.M., Cox, P., van de Weg, W.E., Visser, R.G., Maliepaard, C., Smulders, M.J., 2016. High-density SNP-based genetic maps for the parents of an outcrossed and a selfed tetraploid garden rose cross, inferred from admixed progeny using the 68k rose SNP array. Hortic. Res. 3, 16052. doi:10.1038/hortres.2016.52

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M., Salvi, S., Milner, S.G., Cattivelli, L., Mastrangelo, A.M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A.R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M.C., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K.J., Hayden, M., Akhunov, E., 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol. J. 12, 787–796. doi:10.1111/pbi.12183

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., Qiu, J.-L., 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat. Biotechnol. 32, 947–952. doi:10.1038/nbt.2969

Watanabe, K., 1983. Studies on the control of diploid-like meiosis in polyploid taxa of Chrysanthemum. Theor. Appl. Genet. 66, 9–14.

Wernett, H., Sheehan, T.J., Wilfret, G., Marousky, F., Lyrene, P., Knauft, D., 1996. Postharvest longevity of cut-flower Gerbera. I. Response to selection for vase life components. J. Am. Soc. Hortic. Sci. 121, 216–221.

Wolfram Research, 2015. Mathematica.

Wu, K., Burnquist, W., Sorrells, M., Tew, T., 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83, 294–300.

Wu, R., Ma, C.-X., Casella, G., 2002. A Bivalent Polyploid Model for Linkage Analysis in Outcrossing Tetraploids. Theor. Popul. Biol. 62, 129–151. doi:10.1006/tpbi.2002.1608

Yakimova, E., Kapchina-toteva, V., Alexieva, V., Sergiev, I., Karanov, E., Str, B., 1996. Effect of Chlorsulfuron (Glean-75) and Sucrose on Some Post-Harvest Physiological Events in Cut Flowers. Bulg. J. Plant Physiol. 22, 74–87.

Yeo, A.R., Yeo, M.E., Flowers, S.A., Flowers, T.J., 1990. Screening of rice (Oryza sativa L.) genotypes for physiological characters contributing to salinity resistance, and their relationship to overall performance. Theor. Appl. Genet. 79, 377–384. doi:10.1007/BF01186082

Zhan, L., Hu, J., Ai, Z., Pang, L., Li, Y., Zhu, M., 2013. Light exposure during storage preserving soluble sugar and l-ascorbic acid content of minimally processed romaine lettuce (Lactuca sativa L.var. longifolia). Food Chem. 136, 273–8. doi:10.1016/j.foodchem.2012.07.123

Zhang, F., Chen, S., Chen, F., Fang, W., Chen, Y., Li, F., 2011a. SRAP-based mapping and QTL detection for inflorescence-related traits in chrysanthemum (Dendranthema morifolium). Mol. Breed. 27, 11–23. doi:10.1007/s11032-010-9409-1

Zhang, F., Chen, S., Chen, F., Fang, W., Deng, Y., Chang, Q., Liu, P., 2011b. Genetic analysis and associated SRAP markers for flowering traits of chrysanthemum (Chrysanthemum morifolium). Euphytica 177, 15–24. doi: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. Sci. Hortic. 125, 422–428. doi: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, 1–9. doi:10.1371/journal.pone.0083023

Zhang, F., Jiang, J., Chen, S., Chen, F., Fang, W., 2012a. Detection of quantitative trait loci for leaf traits in chrysanthemum. J. Hortic. Sci. Biotechnol. 87, 613–618.

Zhang, F., Jiang, J., Chen, S., Chen, F., Fang, W., 2012b. Mapping single-locus and epistatic quantitative trait loci for plant architectural traits in chrysanthemum. Mol. Breed. 30, 1027–1036. doi:10.1007/s11032-011-9686-3

Zhang, F.Z., Wagstaff, C., Rae, A.M., Sihota, A.K., Keevil, C.W., Rothwell, S.D., Clarkson, G.J.J., Michelmore, R.W., Truco, M.J., Dixon, M.S., Taylor, G., 2007. QTLs for shelf life in lettuce co-locate with those for leaf biophysical properties but not with those for leaf developmental traits. J. Exp. Bot. 58, 1433–1449. doi:10.1093/jxb/erm006

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M., Hulse-Kemp, A.M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., Chang, L., Zhang, W., Song, Q., Kirkbride, R.C., Chen, X., Dennis, E., Llewellyn, D.J., Peterson, D.G., Thaxton, P., Jones, D.C., Wang, Q., Xu, X., Zhang, H., Wu, H., Zhou, L., Mei, G., Chen, S., Tian, Y., Xiang, D., Li, X., Ding, J., Zuo, Q., Tao, L., Liu, Y., Li, J., Lin, Y., Hui, Y., Cao, Z., Cai, C., Zhu, X., Jiang, Z., Zhou, B., Guo, W., Li, R., Chen, Z.J., 2015. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat. Biotechnol. 33, 531–537. doi:10.1038/nbt.3207

Zhang, Y., Dai, S., Hong, Y., Song, X., 2014. Application of genomic SSR locus polymorphisms on the identification and classification of chrysanthemum cultivars in China. PLoS One 9, e104856. doi:10.1371/journal.pone.0104856

Zhao, N., Yu, X., Jie, Q., Li, H., Li, H., Hu, J., Zhai, H., He, S., Liu, Q., 2013. A genetic linkage map based on AFLP and SSR markers and mapping of QTL for dry-matter content in sweetpotato. Mol. Breed. 32, 807–820. doi:10.1007/s11032-013-9908-y

Zheng, C., Voorrips, R.E., Jansen, J., Hackett, C.A., Ho, J., Bink, M.C.A.M., 2016. Probabilistic multilocus haplotype reconstruction in outcrossing tetraploids. Genetics 203, 119–131. doi:10.1534/genetics.115.185579

Zhu, X., Li, H., Ye, M., Jiang, L., Sang, M., Wu, R., 2016. AlloMap6: an R package for genetic linkage analysis in allohexaploids. Brief. Bioinform. 1–9. doi:10.1093/bib/bbw085

Zor, T., Selinger, Z., 1996. Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies. Anal. Biochem. 236, 302–8. doi:10.1006/abio.1996.0171

Zych, K., Voorrips, R.E., Jansen, R.C., Gort, G., n.d. FitTetra 2.0 – improved genotype calling for tetraploids with multiple population and parental data support.

Summary

DNA-informed selection can strongly improve the process of plant breeding. It requires the detection of DNA polymorphisms, calculation of genetic linkage, access to reliable phenotypes and methods to detect genetic loci associated with phenotypic traits of interest. Cultivated chrysanthemum is an outcrossing hexaploid with an unknown mode of inheritance. This complicates the development of resources and methods that enable the detection of trait loci. Postharvest performance is an essential trait in chrysanthemum, but is difficult to measure. This makes it an interesting but challenging trait to phenotype and detect associated genetic loci. In this thesis I describe the development of resources and methods to enable phenotyping for postharvest performance, genetic linkage map construction and detection of quantitative trait loci in hexaploid chrysanthemum.

Postharvest performance is a complicated trait because it is related to many different disorders that reduce quality. One of these disorders in chrysanthemum is disk floret degreening, which occurs after long storage. In **chapter 2**, we show that degreening can be prevented by feeding the flower heads with sucrose, suggesting carbohydrate starvation plays a role in the degreening process. To investigate the response to carbohydrate starvation of genotypes with different sensitivity to disk floret degreening, we investigated the metabolome of sugar-fed and carbohydrate-starved disk florets by ^{1}H-NMR and HPAEC. We show that the metabolome is severely altered at carbohydrate starvation. In general, starvation results in an upregulation of amino acid and secondary metabolism. Underlying causes of genotypic differences explaining variation in disk floret degreening in the three investigated genotypes remained to be elucidated, but roles of regulation of respiration rate and camphor metabolism were posed as possible candidates.

In **chapter 3**, disk floret degreening was found to be the most important postharvest disorder after 3 weeks of storage among 44 white chrysanthemum cultivars. To investigate the inheritance of disk floret degreening, we crossed two genotypes with opposite phenotypic values of both disk floret degreening and carbohydrate content to obtain a population segregating for disk floret degreening. To phenotype the cultivar panel and the bi-parental population precisely and in a high throughput manner, we developed a method that quantified colour of detached capitula over time. This method was validated with visual observations of disk floret degreening during vase life tests. In a subset of the bi-parental population we measured carbohydrate content of the disk florets at harvest. The amount of total carbohydrates co-segregated with sensitivity to degreening, which shows that the difference in disk floret degreening sensitivity between the parents could be explained by their difference in carbohydrate content. However, the correlation was rather weak, indicating carbohydrate content is not the only factor playing a role.

In order to develop resources for DNA-informed breeding, one needs to be able to characterize DNA polymorphisms. In **chapter 4**, we describe the development of a genotyping array containing 183,000 single nucleotide polymorphisms (SNPs). These SNPs were acquired by sequencing the transcriptome of 13 chrysanthemum cultivars. By comparing the genomic dosage based on the SNP assay and the dosage as estimated by the read depth from the transcriptome sequencing data, we show that alleles are expressed conform the genomic dosage, which contradicts to what is often found in disomic polyploids. In line with this finding, we conclusively show that cultivated chrysanthemum exhibits genome-wide hexasomic inheritance, based on the segregation ratios of large numbers of different types of markers in two different populations.

Tools for genetic analysis in diploids are widely available, but these have limited use for polyploids. In **chapter 5**, we present a modular software package that enables genetic linkage map construction in tetraploids and hexaploids. Because of the modularity, functionality for other ploidy levels can be easily

added. The software is written in the programming language R and we named it polymapR. It can generate genetic linkage maps from marker dosage scores in an F1 population, while taking the following steps: data inspection and filtering, linkage analysis, linkage group assignment and marker ordering. It is the first software package that can handle polysomic hexaploid and partial polysomic tetraploid data, and has advantages over other polyploid mapping software because of its scalability and cross-platform applicability.

With the marker dosage scores of the bi-parental F1 population from the genotyping array and the developed methods to perform linkage analysis we constructed an integrated genetic linkage map for the hexaploid bi-parental population described in chapter 3 and 4. We describe this process in **chapter 6**. With this integrated linkage map, we reconstructed the inheritance of parental haplotypes for each individual, and expressed this as identity-by-descent (IBD) probabilities. The phenotypic data on disk floret degreening sensitivity that was acquired as described in chapter 3, was used in addition to three other traits to detect quantitative trait loci (QTL). These QTL were detected based on the IBD probabilities of 1 centiMorgan intervals of each parental homologue. This enabled us to study genetic architecture by estimating the effects of each separate allele within a QTL on the trait. We showed that for many QTL the trait was affected by more than two alleles.

In **chapter 7**, the findings in this thesis are discussed in the context of breeding for heterogeneous traits, the implications of the mode of inheritance for breeding and the advantages and disadvantages of polyploidy in crop breeding. In conclusion, this thesis provides in general a significant step for DNA-informed breeding in polysomic hexaploids, and for postharvest performance in chrysanthemum in particular.

Acknowledgements

About the author

Geert (Gerrit Adriaan) van Geest was born on June 25th 1987 in 's-Gravenzande, the Netherlands. As the son of a grower of cut-flowers, he grew up in an environment where horticulture and plants played a central role. After finishing secondary school in 2005, he travelled the world for a year, and studied cultural anthropology at Leiden University for the next year. Coming to realize his passion was in horticulture and plant breeding, he enrolled in 2007 in Wageningen University to study plant sciences. After receiving his cum-laude BSc degree in 2010, he continued his masters' degree combining the specializations Plant Breeding and Genetic Resources and Greenhouse Horticulture. He graduated in 2012, and soon after that started working at Deliflor Chrysanten B.V. as a researcher at the breeding department. In 2013, his PhD project started as a cooperation between Deliflor Chrysanten B.V. and the research groups Plant Breeding and Horticulture and Product Physiology at Wageningen UR. The results of this PhD project are treated in this thesis.

Education statement EPS

Education Statement of the Graduate School
Experimental Plant Sciences

Issued to: **Geert van Geest**
Date: **15 September 2017**
Group: **Plant Breeding**
University: **Wageningen University & Research**

1) Start-up phase	*date*
► **First presentation of your project**	
Title: Linkage mapping and analysis of QTL in hexaploid chrysanthemum	26 Nov 2012
► **Writing or rewriting a project proposal**	
Title: Linkage mapping and analysis of quantitative trait loci (QTL) in hexaploid chrysanthemum (Chrysanthemum x morifolium)	Sep-Dec 2012
► **Writing a review or book chapter**	
► **MSc courses**	
BIF 30806 - Advanced Bioinformatics	Feb-Mar 2014
Laboratory use of isotopes	
Subtotal Start-up Phase	*11.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
EPS PhD Student day, Leiden, NL	29 Nov 2013
EPS Get2Gether 2015, Soest, NL	29-30 Jan 2015
EPS Get2Gether 2016, Soest, NL	28-29 Jan 2016
► **EPS theme symposia**	
EPS Theme 4 symposium 'Genome Biology', Wageningen, NL	16 Dec 2016
EPS Theme 2 symposium 'Interactions between Plants and Biotic Agents', Wageningen, NL	23 Jan 2017
Annual meetings (Lunteren days) and other National	
► **Platforms**	
Annual meeting Experimental Plant Sciences, Lunteren, NL	14-15 Apr 2014
Annual meeting Experimental Plant Sciences, Lunteren, NL	11-12 Apr 2016
Annual meeting Experimental Plant Sciences, Lunteren, NL	10-11 Apr 2017
Seminars (series), workshops and symposia	
Christine Hackett: Linkage analysis and QTL mapping in autotetraploid potato using SNP dosage data	07 Sep 2015
Jeffrey Endelman: Genome-wide association analysis and prediction in tetraploid potato	18 Nov 2016
► **Seminar plus**	
► **International symposia and congresses**	
Next generation plant breeding congress, Ede, NL	13-14 Nov 2012
24th EUCARPIA SYMPOSIUM 2012 – Section Ornamentals "Ornamental Breeding Worldwide", Warsaw, Poland	02-05 Sep 2012
2013 International Symposium on Chrysanthemum, Beijing, China	25-29 Sep 2013
QMSCO2015 Kermanshah, Iran	01-04 May 2015
25th EUCARPIA SYMPOSIUM 2015- Section Ornamentals "Crossing Borders", Melle, Belgium	29 Jun-01 Jul 2015
20th General Eucarpia congres, Zurich, Switzerland	29 Aug-01 Sep 2016
Phenodays 2016 Conference, Berlin, Germany	26-27 Oct 2016
► **Presentations**	
Talk: QMSCO	01-04 May 2015
Talk: EUCARPIA ornamentals 2015	29 Jun-01 Jul 2015
Talk: EPS Get2Gether 2015	29-30 Jan 2015
Talk: General Eucarpia congres	29 Aug-01 Sep 2016
► **IAB interview**	
► **Excursions**	
Subtotal Scientific Exposure	*15.6 credits**

3) In-Depth Studies	_date_
► **EPS courses or other PhD courses**	
Multivariate Analysis	22-24, 30 & 31 Oct 2013
The power of RNAseq	16-18 Dec 2013
Phenodays phenomatics workshop, Berlin, Germany	24-25 Oct 2016
► **Journal club**	
Member of literature discussion group at HPC (FLOP)	2014-2015
Member of literature discussion group at PBR (Polyploid genetics)	2016
► **Individual research training**	
Subtotal In-Depth Studies	_4.3 credits*_

4) Personal development	_date_
► **Skill training courses**	
	28-29 Oct & 21 Nov 21 2013
Communication with the media and general public	
Competence assessment	21 May & 18-20 Jun 2013
Voice Matters - Voice and Presentation Skills Training	01 & 15 Oct 2013
WGS workshop carousel 2014	06 Jun 2014
► **Organisation of PhD students day, course or conference**	
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	_2.0 credits*_

TOTAL NUMBER OF CREDIT POINTS*	**33.4**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

* A credit represents a normative study load of 28 hours of study.