University of Alberta

Linkage mapping, phenotypic characterization, and introgression analysis of crosses with safflower *(Carthamus tinctorius)*

by

Catherine J. Archibald

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Plant Biology

Department of Biological Sciences

Spring 2011
Edmonton, Alberta

Examining Committee

Allen Good, Supervisor, Department of Biological Sciences

James C. Cahill, Department of Biological Sciences

David Coltman, Department of Biological Sciences

Dean Spaner, Department of Agriculture, Food, and Nutritional Sciences

Abstract

An inter-specific linkage map of a backcross (BC1) population derived from a cross between *C. tinctorius* and C. *oxyacanthus* was developed using primarily microsatellite markers. A composite map, which aligned the inter-specific (*C. tinctorius* x *C. oxyacanthus)* map with an intra-specific linkage map previously developed from an F2 population of *C. tinctorius*, was created to search for syntenic regions. Results indicate that despite low marker saturation, there is substantial colinearity between the two linkage maps, and one translocation or inversion event. Upon subsequent self-fertilization events, phenotypes of the inter-specific backcross population were characterized in both the field (BC1S2 generation) and growth chamber (BC1S2 and BC1S3 generations), and identified several lines of agronomic interest. Introgression analysis was performed (BC1S3 generation) to assess the level of integration of *C. oxyacanthus* DNA into the *C. tinctorius* genome, and results are suggestive of less inter-specific recombination than expected.

Acknowledgements

First, I would like to acknowledge and thank Dr. Allen Good for his supervision and counsel throughout my graduate studies, without which this work would not be possible. It has been a pleasure and a privilege to study in your laboratory, and I wish you much success in all your future endeavours. Second, I would like to thank Dr. Reinhold Mayerhofer for his assistance and contributions, and especially for developing the genetic resources I used in linkage mapping. Third, Dr. Hans-Henning Mündel has offered his expertise and advice on safflower phenotyping in the 2009 field trial in Warner, Alberta. Thank you for all the trips you have taken to the field and for your photography contributions. Thank you also to Brian Otto for providing space on his farm in Warner, Alberta for the safflower field trial.

To Victoria Bowles, I would like to say thank you for offering me all the insight you have, and for helping me adjust to the laboratory when I first arrived. I wish also to acknowledge and thank the laboratory technicians Marion Mayerhofer and Laila Muzammal for their daily cooperation and ongoing help with our work in safflower. To Marion in particular, thank you for developing the linkage mapping population. To Dr. Steven Rothstein and Dr. Yongmei Bi from the University of Guelph, I thank you for encouraging my interest in research and introducing me to Dr. Allen Good. To the many Good lab members with whom I have worked, thank you very much for your great company, support, and for all the good times. I wish you all the best.

Last but not least, I want to acknowledge and thank my family members and close friends who have been a constant support for me. I have appreciated all your care, phone calls, e-mails, and visits to Edmonton. To my Mum and Dad, and sisters Heather and Jen, I thank you for taking an interest in my work; it never felt like you were far away. Thank you to my Nana and Grandpa Owen for all your care and concern. Thank you to Heidi Andre for your delightful day to day company. Especially, I'd like to gratefully acknowledge Christopher Greve, who rearranged his life and geography in order to be here with me while I work and study. Thank you for your patience and love.

Table of Contents

List of Tables

List of Figures

List of Symbols, Abbreviations and Nomenclature

AFLP – amplified fragment length polymorphism

BC1 – first backcross generation

BC1S2 –backcross generation that has been subsequently self-fertilized twice

BC1S3 – backcross generation that has been subsequently self-fertilized three times

BLASTN – nucleotide basic local alignment search tool

cDNA – complementary deoxyribonucleic acid

Cent – Centennial cultivar of *Carthamus tinctorius*

CGP – Compositae Genome Project

cM – centimorgans

CTAB – Cetyl trimethylammonium bromide

d – days

dATP – deoxyadenosine triphosphate

dCTP –deoxycytosine triphosphate

DF – days to flowering

dGTP – deoxyguanosine triphosphate

DM – days to maturity

DNA – deoxyribonucleic acid

dTTP – deoxythymidine triphosphate

EST – expressed sequence tag

F1 – first filial generation

F2 – second filial generation

HIP – hexane-isopropanol solution

IBPGR – International Board for Plant Genetic Resources

ISSR – inter-simple sequence repeat

ITS – internal transcribed spacers

LG – linkage group

LOD – log of ODDs ratio

n – haploid chromosome number

2n – diploid chromosome number

NP12 – NP12 cultivar of *Carthamus tinctorius*

Ns – non-significant

Oxy – *Carthamus oxyacanthus*

P - probability

PCR – polymerase chain reaction

pers. observation – personal observation

PMPs – plant made pharmaceuticals

QTL – quantitative trait loci

R – Pearson correlation coefficient

RAPD – random amplified polymorphic DNA

rep – replicate

RFLP – restriction fragment length polymorphism

rmax – maximum recombination frequency

SARF – sum of adjacent recombination fractions

SD – standard deviation

SNP – single nucleotide polymorphism

SSR – simple sequence repeat

STR – short tandem repeat

S-317 – S-317 cultivar of *Carthamus tinctorius*

S5 – a population resulting from five generations of self-fertilization

TCM – traditional Chinese medicine

X^2 – Chi-square statistic

Linkage mapping, phenotypic characterization and introgression analysis of crosses with safflower (*Carthamus tinctorius*)

1. General Introduction

Safflower (*Carthamus tinctorius* L.) has been cultivated as a crop for over 4,500 years. It originated in the eastern Mediterranean, and spread to Egypt, Ethiopia, southern Europe, south Asia and the Far East early in its evolution (Smith, 1996). *C. tinctorius* is now grown in over 60 countries, with India being the top commercial producer. Safflower was grown on an estimated 690 km^2 in North America in 2009, producing 708 Mg of seed (FAO, 2009). In Canada, safflower cultivars have been adapted to grow in the brown and dark brown soil zones of the Prairie region (Figure 1.1). It is grown largely for production of high quality cooking oil, but now is being used in molecular pharming, and certain cultivars have also shown potential for use as a biofuel.

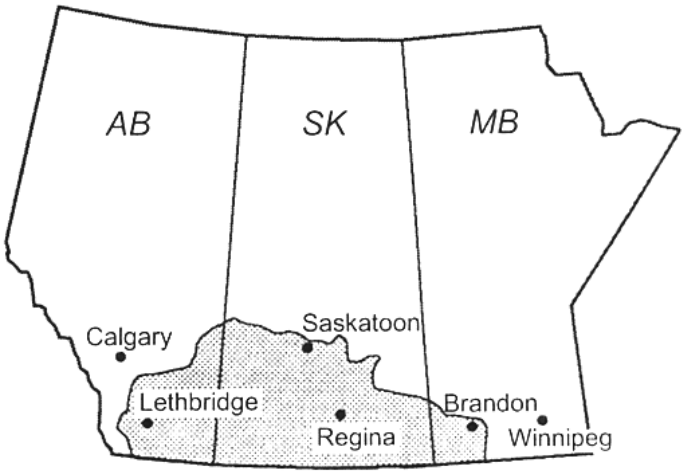

Figure 1.1: Areas of adaptation for safflower in Canada.
Figure adapted from Mündel *et. al*., 2004.

C. tinctorius is a diploid plant (2n=24) belonging to the Cynareae (thistle) tribe and the Compositae (Asteraceae) family. The Compositae does not yet have a well-characterized model plant like *Arabidopsis* from which genetic material can be sourced, and the commercial crops most closely related to safflower are lettuce (*Lactuca sativa*) and sunflower (*Helianthus annus*). However, there are numerous wild relatives in the *Carthamus* genus which are available to study, including several which have been labelled noxious weeds in North America (*C. lanatus, C. leucocaulos, C. oxyacanthus*). In order to further the development of safflower as a crop, it is necessary to have both sufficient genetic material available, and to understand the introgression ability of safflower with weedy species. Developing genetic material provides plant breeders with the resources needed for improvement of *C. tinctorius*, and an understanding of the introgression between safflower and its wild relatives is helpful to evaluate the potential use of these species for crop improvement. The work in this thesis contributes to this research first by the development of an inter-specific genetic map, which constitutes the first in-depth linkage analysis performed for any of the species in the *Carthamus* genus, and second with a study of the introgression pattern and phenotypic characterization of plant populations derived from an inter-specific cross between safflower and its wild relative.

1.1. Biology of Safflower

Safflower is a dicotyledonous, herbaceous, annual or winter annual, thistle-like plant that is highly self-fertilizing, with out-crossing rates of less than

10% (Knowles, 1969). While safflower has numerous wild relatives in the *Carthamus* genus (*C. creticus, C. glaucus, C. lanatus, C. leucocaulos, C. oxyacanthus, C. palaestinus, C. persicus, C. turkestanicus*), it is not a weed, as it only exists as a cultivated crop (Smith, 1996). Its most prominent features are colourful flower heads, a deep taproot, and the production of white, oil-bearing fruits. Safflower is suited to grow in hot, dry climates, where soils are moist in early spring but generally well-drained. Depending on environmental conditions, the typical generation length of safflower varies from about 17-20 weeks (Smith, 1996), and the growth cycle is divided into the following stages: emergence, rosette, stem elongation, branching, flowering and maturity (Figure 1.2; Mündel *et. al.*, 2004).

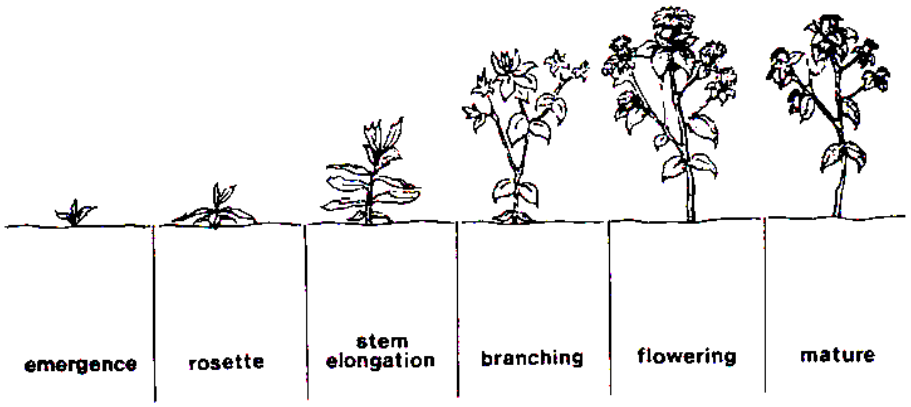

Figure 1.2: Growth cycle of safflower plants.
Figure adapted from Mündel *et. al.*, 2004.

Upon seeding, emergence occurs between 3 days to 3 weeks, depending on moisture availability and temperature. Safflower emerges once soil temperatures rise to 4.4°C, with emergence rates far higher when temperatures are 15.6°C or greater (Kaftka, 1965). Safflower yield is greatest when planted in

deep, well-drained, sandy loam soil with good water holding capacity well below surface level (Mündel *et. al*., 2004). The amount of fertilizer needed depends on the desired yield and the other crops used in rotation, and soil testing is needed to determine if any additional nutrients are necessary. If rotated with cereal crops, safflower can recover residual nitrogen fertilizer found below the root depth of cereal crops. Otherwise, high yields are usually obtained with 112 to 134 kg/ha of nitrogen application (Oelke *et. al*., 1992). Once the cotyledons of emerged seedlings have spread, leaves begin to grow and the plant enters a rosette stage, where the stem does not elongate but the plant develops a long taproot that can grow 2-3m in depth. During this period of time, safflower is tolerant to cold weather and frost, but highly susceptible to weeds (Dajue and Mündel, 1996). The nature of the rooting system in safflower, being that it has one deep taproot and several successive lateral roots, enables safflower to obtain water and nutrients from well below the surface soil level, which gives safflower a degree of drought resistance.

After spending 2-3 weeks in a rosette stage, stem elongation proceeds very quickly, and much branching off the main axis occurs. The first branches off the main axis are called the primary branches; the branches stemming from the primary branches are called the secondary branches, and so on. Safflower can grow to a height of 0.5-1.8m (Smith, 1996). The leaf size varies greatly among cultivars, and leaf shape can be linear, lanceloate, ovate or oblong, with increasing spininess towards the top of the plant. Normally spines are present on the leaves and bracts of safflower (Bradley *et. al*., 1999), and the presence of spines is

largely determined by a single, dominant gene (Pahlavani *et. al*., 2004). The role of spines in safflower is still unclear, but there has been speculation about their value in deterring birds and herbivores. Non-spiny varieties of safflower have been developed for the purposes of hand harvesting floral parts and seeds (Singh, 2007). Flowering begins once stems are elongated and branching is extensive. Groups of flowers are contained in a flower head known as a capitulum, a characteristic trait of all plants in the Compositae family. A safflower plant has many capitula, each of which is found at the end of a stem and surrounded by involucral bracts. A given capitulum can have 20 to 180 flowers or florets. Within a capitulum, flowering starts at the outer florets and proceeds centripetally towards the centre.

Flowering begins at the primary capitulum of a safflower plant, which is located at the top of the main stem, and then continues toward the outer stems. Florets are tubular and contain five corolla lobes, which spread during anthesis (Pille and Knowles, 1975). On each floret are five stamens which are united in the anther tube. As the style and stigma grow up through the surrounding anther tube, self-pollination occurs. An ovary is located at the base of each floret, which develops into a single-seeded achene or fruit (Knowles, 1980). Flower colour varies from white, which is rarely seen, to deep red, with shades of yellow and orange being most common (Figure 1.3). Wilted flower colour is typically darker than flowers in the full bloom stage, which can last up to four weeks (Dajue and Mündel, 1996).

Figure 1.3: Diversity of flower colour in *C. tinctorius* cultivars, showing clockwise from top left: white, pale yellow, medium yellow, orange and deep red capitula. Flowers were photographed from field plots in Warner, Alberta (2009).

The achenes mature within 30 to 35 days of flowering, after which safflower requires about another two weeks to dry, prior to harvest. Once the leaves are brown and the capitula have dried such that very little green, if any, remains on the bracts of the last developing capitula, plants are ready for harvest (Mündel *et. al.*, 2004). If weather is excessively moist prior to harvest, sprouting may occur in the capitula. This is avoided with prompt harvest directly upon physiological maturity with a small grain combine and subsequent drying of the fruits. The oil-bearing fruits contain 33-60% hull and 40-67% kernel, with oil content typically ranging from 20-45% (Dajue and Mündel, 1996). The tough, fibrous hull acts to protect the embryo and cotyledons which comprise the kernel (Smith, 1996). The achenes, which herein will simply be called the seed, are four-sided and normally a glossy, white colour. Most seeds of modern safflower cultivars lack the tuft-like appendage or pappus that is characteristic of thistles.

During the growth cycle, safflower plants are subject to numerous diseases, with Alternaria leaf blight caused by *Alternaria carthami* and Sclerotinia head rot caused by *Sclerotinia sclerotiorum* being the most common (Kaftka, 1965). Both can significantly affect yield, particularly in unseasonably wet conditions. Safflower can also suffer damage by pests and herbivores, including grasshoppers, thrips, and deer, and require cultural and chemical controls to avoid problems with weeds.

1.2. Historical and Current Uses of Safflower

Safflower has been cultivated for many purposes. As early as 4,500 B.C.E. in Egypt, Morocco, China and India, safflower was used for carthamin, the yellow or red coloured dye in the flowers (Cannon and Cannon, 2003). The dye was used to colour cotton, silk, cheese, cosmetics, and even to anoint mummies' tombs in Egypt (Ekin, 2005), though more inexpensive synthetic dyes are now preferred for these purposes. Safflower has been used as an herb in teas, tinctures and decoctions in traditional Chinese medicine (TCM) for thousands of years. In India, the flowers of safflower are known as koosumbha, while they are called hong hua in TCM. Safflower has traditionally been used to treat disorders due to blood stasis, including amenorrhea and dysmenorrhea (Moloney, 1998). Blossoms of safflower have also been used in Chinese medicine to treat fever, measles, phlegm, and skin rashes. Unprocessed safflower oil is an effective laxative, while processed safflower has no laxative properties (Lizhong, 1993; Zhang and Chai, 1997).

It was not until the 1800s that safflower began to be used for its oil. Safflower oil was first used in paints, varnishes, and burned for lighting where electricity was unavailable, but later became a valuable source of high quality cooking oil as the interest in healthful, polyunsaturated fats increased. Experimental work provided by SemBioSys Genetics Inc. (2008, unpublished) offers a comparison of fatty acid profiles in mature seeds of several *Carthamus* species, including *C. tinctorius*. Results of the fatty acid profiling experiment, where lipids were extracted with a 3:2 hexane:isopropanol (HIP) solution and subsequently transesterified with methanol in acidic conditions (1.5M HCl), indicate low levels of linolenic acid in all *Carthamus* species tested (Table 1.1). All analyzed species show linoleic acid content in excess of 70% of total fatty acids, apart from the S-317 cultivar of *C. tinctorius*, which is a cultivar bred specifically to be a high producer of oleic acid. Not surprisingly, the commercial safflower, *C. tinctorius* cv. Centennial, had the highest average percentage of oil per embryo while the weedy species *C. leucocaulos* had the lowest. Percentage of oil per embryo was not calculated for *C. tinctorius* cv. S-317. These results are comparable with other fatty acid profiling experiments, except values of oleic acid in *C. lanatus* were higher in other studies (Murthy and Anjani, 2008; Sabzalian *et. al*., 2008). The high percentage of polyunsaturated fat (linoleic acid) in *C. tinctorius* cv. Centennial makes its oil one of the healthiest edible oils available today. Its use has been recommended for reduction in serum cholesterol levels (Smith, 1996), and safflower oil is now also used in salad dressings and margarine.

Table 1.1: Mean fatty acid profiles (%) from HIP extracted total lipids with calculated standard deviation, courtesy of SemBioSys Genetics Inc. (2008, unpublished).

Species		Palmitic Acid	Stearic Acid	Oleic Acid	Linoleic Acid	Linolenic Acid	% Oil/ Embryo
C. creticus	Mean	8.59	3.20	16.58	71.01	0.00	40.09
	SD	0.12	0.15	2.82	3.25	0.00	4.08
C. glaucus	Mean	6.73	1.72	18.64	71.90	0.17	37.39
	SD	0.19	0.06	2.84	3.39	0.02	2.37
C. lanatus	Mean	7.19	2.39	8.02	81.42	0.00	43.39
	SD	0.31	0.26	1.36	1.90	0.00	2.63
C. leucocaulos	Mean	9.04	4.40	8.56	74.75	0.00	27.78
	SD	0.65	0.78	1.17	3.45	0.00	4.81
C. oxyacanthus	Mean	7.07	1.81	11.56	78.36	0.00	37.98
	SD	0.04	0.23	0.29	0.39	0.00	0.68
C. palaestinus	Mean	7.12	1.88	10.43	79.93	0.00	47.55
	SD	0.28	0.28	2.21	2.85	0.00	2.12
C. tinctorius	Mean	7.76	1.49	9.82	80.70	0.10	46.89
cv. Centennial	SD	0.38	0.15	0.85	1.31	0.09	1.72
C. tinctorius	Mean	5.91	2.21	74.10	16.88	0.00	-
cv. S-317	SD	0.00	0.03	0.12	0.01	0.00	

Molecular pharming is the production of pharmaceutically important compounds by use of transgenic DNA. The Calgary-based biotech company SemBioSys Genetics Inc. uses safflower as their model plant species to produce high-value recombinant compounds that attach to oilbodies in the seed. The use of safflower in transgenic experiments has resulted in the production of synthetic insulin, which is currently in phase 2 human clinical trials (Markley *et. al*., 2006), and biologically active Apolipoprotein Al Milano ($ApoAl_{Milano}$), which is believed to play an important role in the prevention of coronary heart disease (Nykiforuk *et. al.*, 2010).

In addition to the production of plant-made pharmaceuticals (PMPs), safflower oil has shown potential for use as a biofuel or fuel extender. Biodiesel fuel is attractive because of its potential to reduce greenhouse gas emissions and

ability to be used directly in diesel engines with only minor alterations. Deciding upon the ideal oilseed crop for biodiesel feedstock in Canada has been an important consideration since both the introduction of the Renewable Fuel Standard in May of 2008 in the House of Commons, which states that biodiesel needs to be one of Canada's transportation fuel sources, and the development of cost-effective methods of esterification by the BIOX Corporation (http://www.bioxcorp.com/index.php). While canola (*Brassica napus*) is currently the preferred oilseed crop for biodiesel feedstock in Canada, safflower may be a superior choice. Both canola and safflower oil have the ideal low melting point for biodiesel production, however, high oleic acid producing cultivars of safflower have greater oxidative stability. In a two-year comparative study in Montana, USA with sunflower, canola, flax (*Linum usitatissimum*), soybean (*Glycine max*), and camelina (*Camelina sativa*) oilseed crops, safflower had the highest seed and oil yield, with 3404 kg/ha and 267 L/ha, respectively (Bergman and Flynn, 2008). Furthermore, two of the three bi-products created from processing safflower feedstock, glycerine and fat, can be resold to generate additional income, while methanol, the other bi-product, is recycled for use in supplementary processing (van Gerpen, 2005). High oleic acid producing cultivars of safflower do have a possible pollutant-reducing effect because the oil is biodegradable, largely free of sulphur, and lacks the fossil fuel carbon dioxide (Flynn and Berman, 2001). Still, economical production of biofuel from safflower necessitates the development of greater income sources for the bi-products of processing and a more efficient method of crop production.

On Canadians farms today, commercial safflower is grown primarily for birdseed for caged birds. Safflower also supplies meal used in animal forage and poultry. Safflower forage is of comparable quality to cereal or alfalfa, and is desired by cattle, sheep and goat (Landaua *et. al*., 2004). Despite all these uses and applications of safflower plants, it remains a low acreage crop compared to other oilseed crops, in part because of a need for further crop improvement requiring genetic material that has been largely unavailable until now.

1.3. Genetic Material

An understanding of the genetic makeup of a plant species is highly beneficial for the purposes of breeding wild germplasm into established cultivars and analyzing inter-specific hybrid populations. A description of the genome can be provided with high density linkage maps or full genome sequencing, of which the former is typically done first in plant species where there is a clear economic reason. Linkage maps, also known as genetic maps, are graphical representations of a genome, based on the concept that the recombination frequency between genes decreases as does the distance between them. Distances recorded are genetic distances measured in centimorgans (cM), where 1 cM is defined as the distance between genes or markers such that one product of meiosis in 100 is recombinant, which is a recombination frequency equivalent to 1% (Kosambi, 1944). Linkage maps employ various marker systems and statistical functions to establish the most probable genome layout.

The earliest marker systems used in genetic mapping were morphological and could be genotyped based on phenotypic information. The first linkage map was created in *Drosophila* based on sex-linked morphological characters (Sturtevant, 1913). Later, alternate forms of enzymes called isozymes or isoenzymes were used as markers. These are enzymes which vary in amino acid sequence but not in function, and were valued for their low cost and fast development. Isozyme variation is detected with gel electrophoresis and enzyme-specific stains (Soltis and Soltis, 1990). Early linkage map construction for the tomato (*Lycopersicon esculentum*) genome used the isozyme marker system since isozyme variants did not appear to affect phenotypes, unlike morphological markers which were impractical for use in plant breeding programs because they were generally recessive and had undesirable phenotypes (Rick and Tanksley, 1980; Bernatzky and Tanksley, 1986). Since then, there have been many further technological developments and there are now saturated or near-saturated linkage maps developed for many plant species including *Brassica napus* (Lombard and Delourme, 2000; Zudhong *et. al*., 2007), *Oryza sativa* (rice) (Causse *et. al*., 1994; Harushima *et. al*., 1998), *Hordeum vulgare* (barley) (Liu *et. al*., 1996), *Zea mays* (maize) (Vuylsteke *et. al.*, 1999), *Triticum aestivum* (winter wheat) (Palliard *et. al*., 2003), *Glycine max* (Song *et. al.*, 2004), *Helianthus annus* (Yu *et. al*., 2003), and a myriad of other plant species, many of which have also been sequenced. The current methodology by which linkage maps are created no longer relies on morphological and enzyme data, but rather uses DNA-based molecular markers, which are generally more accurate.

The first molecular markers to be used in linkage mapping were restriction fragment length polymorphisms (RFLPs), which detect the varying lengths of the genetic sequence that exists between restriction sites. While such variation within a genome is very common, RFLPs have largely been replaced with other marker systems because of the technical labour involved. Randomly amplified polymorphic DNA markers (RAPDs) and amplified fragment length polymorphisms (AFLPs) are dominant PCR-based (polymerase chain reaction) molecular marker systems, which can be quickly developed but cannot distinguish chromosomal differences. RAPDs use primers that are 6-12 nucleotides in length and amplify random DNA segments (Virk *et. al.,* 1995). The resulting polymorphism is due to variation in the DNA sequence at the site of primer binding and in differing lengths of DNA between primer binding sites (Powell *et. al*., 1996). The use of RAPDs has eliminated some of the technical difficulties associated with RFLPs, but results are not repeatable among laboratories. Unlike RAPDs, AFLPs are largely reproducible among laboratories (Jones *et. al*., 1997). They detect variation between restriction sites by restriction digestion, ligation of adaptors to the sticky ends of restriction fragments, selective PCR amplification of these fragments with primers complementary to the adaptor, and analysis of the banding pattern (Vos *et. al*., 1995). The use of simple sequence repeats (SSR), also known as microsatellite markers or short tandem repeats (STR), has become increasingly popular. These are codominant markers which differ in the number of di-, tri- tetra- hepta- or hexanucleotide repeats present. Detecting polymorphism in SSRs requires the use of unique flanking primers in PCR

reactions, which ensures site specificity and make this marker system highly reliable. Simple sequence repeats do require time to develop, especially in species where genomic sequences or expressed sequence tag (EST) sequences derived from sequencing genomic and cDNA libraries have not been made public. The most expensive molecular marker system is the use of single nucleotide polymorphisms (SNPs), which can be mapped to precise areas in the genome where there is variation in a single nucleotide; however, their initial development is highly time-consuming.

Genetic maps have innumerable applications, including their use in mapping quantitative trait loci (QTL), marker assisted selection, DNA fingerprinting for the purposes of cultivar identification, and studying the introgression pattern in hybrid populations and the evolutionary divergence between species. Building comparative maps from developed linkage maps greatly aids in understanding their evolutionary process (Panjabi *et. al*., 2008), identifying conserved regions of potential importance (Kowalski *et. al*., 1994), and highlighting regions of divergence (Cheung *et. al*., 2009). Until recently, limited genetic mapping had been attempted for safflower, except for the attempts of Ma and Smith (1985) and Ravikumar *et. al.* (2008), where the former map characterized chloroplast DNA based on four restriction endonucleases and the latter map consisted of just nine RAPD markers. While the safflower genome ought to ultimately be fully sequenced, the initial provision of a detailed linkage map should afford insight into the safflower genome in a relatively quick and inexpensive manner.

1.4. Breeding Strategies

Safflower production in Canada was considered risky prior to the introduction of early-maturing Canadian varieties and the registration of effective pesticides in 1985 (Mündel *et. al*., 2004). While several Canadian cultivars have now been established (i.e. Saffire, S-317, AC Sunset), safflower breeding efforts have been limited in Canada. Internationally, mass selection and pure-line selection are the breeding strategies most broadly used for safflower, both of which are ancient methods of crop improvement (Mündel and Bergman, 2009).

Mass selection simply involves saving seed from plants with desirable phenotypes for planting in the following season. The idea is that in so doing, the population frequency of desirable genotypes increases, resulting in an average improvement in the population's performance over time (Aquaah, 2007). Selection can be directly targeted toward traits of interest or made indirectly on correlated traits. Mass selection has been used in Montana, USA to develop safflower cultivars with resistance to several diseases, including Alternaria leaf blight (Bergman *et. al*. 1985, 1987, 1989). While mass selection is a quick and inexpensive method of crop improvement, it can result in correlated responses in secondary traits because of linkage or pleiotropy, and only works on traits with high heritability. Pure line selection, however, can improve traits of low heritability, as selection is based on progeny performance (Aquaah, 2007). A mixed population of different genotypes is self-fertilized for a number of generations to generate genetically pure or near pure lines (also known as near isogenic lines). These lines are then selected by testing over a number of years at

multiple locations. This breeding system has been very popular in India, with many safflower cultivars developed from selection in local land races. Such lines or cultivars have high uniformity but very little genetic variation. While uniformity is necessary for cultivars used in processing markets which require certain safflower qualities, this lack of variation means crops are more susceptible to pathogenic outbreaks and rely on phenotypic plasticity for production response and stability in various environmental conditions (Aquaah, 2007). Although pure line and mass selection have been widely used in safflower breeding, the pedigree method has been the most commonly used practice. This approach first involves hybridization between parents chosen to generate population variation, and then repetitive selection and documentation in successive generations, beginning with the F2 generation. The pedigree plant breeding method has been used in safflower to handle segregating populations and to select for highly heritable and simply inherited traits (Knowles, 1989).

Since *C. tinctorius* has been known to form fertile hybrids in both natural and artificial settings with several related wild species, including *C. oxyacanthus*, *C. palaestinus*, and *C. persicus* (Ashri, 1957; Ashri and Knowles, 1960; Ashri and Rudich, 1965; Deshpande, 1952; Knowles, 1969), hybrid and backcross breeding strategies may be particularly effective approaches either for developing new cultivars or for introducing variation into domesticated cultivars. While both hybrid breeding and backcross breeding strategies use out-crossing techniques for the initial production of a hybrid F1 population, they differ in that the focus of hybrid breeding is to develop new cultivars based on heterosis in hybrid

populations, while backcross breeding aims to incorporate specific characters (i.e. genes for disease resistance) into otherwise high-quality commercial cultivars through repeated crossing of hybrids with the recurrent domesticated parental species (Aquaah, 2007; Mündel and Bergman, 2009). Upon hybridization or backcrossing, single seed descent can be used to quickly attain homozygosity. In backcross breeding schemes, feral traits are introduced or reintroduced into the population. In safflower, these may include traits such as increased seed number, capitula, branching, spinescence, disease resistance, competitive ability, environmental plasticity, time to bolting and maturity, as well as changes in flower and seed colour (Basu *et. al*., 2004; Mayerhofer *et. al.*, submitted). For example, backcrossing has been used in India to transfer wilt (*Fusarium oxysporum*) resistance genes to the Nira cultivar of *C. tinctorius* (Singh *et. al.*, 2003). Both backcross and hybrid breeding strategies have the benefit that they are repeatable, as when the same parents are used again, the hybrid or backcross population can be recovered. However, transferring recessive traits of interest through backcrossing requires additional time, and undesirable linkages may be propagated with this strategy. The best crop improvement program is ultimately one that incorporates multiple breeding strategies.

Through a combination of approaches, safflower breeding to date has improved many traits, including the development of high-yielding cultivars, cultivars with resistances to various pests and diseases, spineless safflower varieties, and high oleic acid producing varieties (Mündel and Bergman, 2009). However, safflower breeders have not yet been able to employ molecular

strategies of crop improvement such as marker assisted selection because of a lack of genetic information. In addition, important traits related to yield, oil content, early maturity, and resistance to biotic and abiotic stresses need to be further improved upon. In order to achieve these objectives of safflower production, crop improvement efforts will need to focus on the following: comparative analysis of stress responses of safflower cultivars in various environments; improving fatty acid composition and percentage oil content in various early-maturing cultivars; increasing the number of capitula and seeds per capitula on plants, which are characters responsible for high-yielding lines (Abel and Driscoll, 1976); and investigation of hybrid or backcross populations for additional variation or traits of interest.

1.5. Genome Introgression

The phenomenon of fertile hybrid production between *C. tinctorius* and several wild *Carthamus* species indicates there is some degree of introgression that occurs between the species' genomes. Introgression, or introgressive hybridization, is defined as infiltration of germplasm from one species to another, and in recent years has been distinguished from transient gene flow in that introgression is considered stable or permanent, although what exactly constitutes permanent gene flow remains ambiguous and is not easily determined (Rieseberg and Wendel, 1993). Naturally, the more closely related the hybridizing species, the more gene introgression there will be. Introgression may not be equal throughout the genome, and in-depth studies are useful for identifying regions of

low homology between species' genomes, where recombination is suppressed or genetic divergence has occurred. Introgression studies can also shed light on the process of plant evolution in closely related species and the relative fitness of certain traits in backcross populations (Mayerhofer *et. al*., submitted).

The out-crossing capabilities of *C. tinctorius* with certain sympatric *Carthamus* species has likely contributed to its evolution, as the allopolyploids *C. turkestanicus* and *C. creticus* are known to have evolved from hybridizations of *C. lanatus* with *C. glaucus* and *C. leucocaulos*, respectively (Vilatersana *et. al*., 2007). The natural hybridization of *C. tinctorius* with wild *Carthamus* species, especially with those which are noxious weeds, does raise concern both about the possible introduction of feral traits into cultivated safflower and transgene escape from domesticated safflower plants. Indeed, hybridization between transgenic cultivated crops and their wild or weedy relatives is the primary concern with marketing genetically engineered crops (Ellstrand and Hoffman, 1990; Linder and Schmitt, 1994). While these risks do exist in regions of geographical overlap (i.e. *C. tinctorius* and *C. oxyacanthus* overlap in California), efforts to eradicate noxious weeds are ongoing. Furthermore, *C. tinctorius*' low seed dormancy, large degree of self pollination, low propensity to weediness, and overall low acreage as a crop species may minimize out-cropping problems, supplying apposite conditions for controlled introgression studies (Mündel and Bergman, 2009).

The nature and degree of introgression between *C. tinctorius* and its wild *Carthamus* relatives is not understood, and the stability of introgression in successive generations after hybridization has not been studied in detail.

Research in this area can provide information regarding the safety of *C. tinctorius* for use as a genetically engineered crop species, as well as the possibilities and limitations of drawing from the gene pool of its wild relatives for improvement of the *C. tinctorius* crop.

1.6. Thesis Objectives

1. The first objective of this thesis research was to identify polymorphic markers in the parental plant species of a *C. oxyacanthus* x *C. tinctorius* cross using the SSR markers previously developed in our laboratory. Markers showing polymorphism among the parental species were then chosen for further analysis in a backcross (BC1) population produced by crossing F1 plants with the recurrent *C. tinctorius* parent.
2. Polymorphic markers were amplified and analyzed in the backcross population, with the goal of creating an inter-specific (*C. oxyacanthus* x *C. tinctorius*) linkage map. The resulting inter-specific linkage map was then used to produce a composite linkage map with an intra-specific *C. tinctorius* map previously created in our laboratory, in order to identify syntenic regions.
3. Next, I focused on characterizing the phenotypes of the backcross plant population in both field and growth chamber experiments after two and three further generations of self-fertilization (BC1S2 and BC1S3). This was done so that the potential for use of these

populations in crop improvement programs, such as those incorporating hybrid or backcross breeding strategies, could be evaluated. My hypothesis is that there will be phenotypic variation among lines in these populations due to the use of two parental plant species, and as a result of this, some lines may show phenotypes of agronomic interest.

4. Finally, marker analysis of the BC1S3 plant population was performed in order to track the introgression of *C. oxyacanthus* DNA into the domesticated *C. tinctorius* genome, so that any areas of selection or allelic preference could be highlighted, and the stability of the introgression assessed.

1.7. Acknowledgements

I wish to gratefully acknowledge Dr. Cory Nykiforuk of SemBioSys Genetics Inc. for performing the fatty acid composition analysis of the *Carthamus* species (Table 1.1).

1.8. References

Abel, G.H. and Driscoll, M.F. 1976. Sequential trait development and breeding for high yields in safflower. *Crop Sci*. **16**: 213-216.

Acquaah, G. 2007. Principles of plant genetics and breeding. Oxford: Blackwell Publishing Ltd. pp. 282-312.

Ashri, A. 1957. Cytogenetics and morphology of *Carthamus* L. species and

hybrids. Ph.D. Dissertation, University of California, Davis, USA. pp. 186.

Ashri, A. and Knowles, P.F. 1960. Cytogenetics of safflower (*Carthamus* L.) species and their hybrids. *Agron. J.* **52**(1): 11-17.

Ashri, A. and Rudich, Y. 1965. Unequal reciprocal natural hybridization rates between two *Carthamus* L. species. *Crop Sci*. **5**: 190-191.

Basu, C., Halfhill, M.D., Mueller, T.C., and Stewart, C.N. 2004. Weed genomics: new tools to understand weed biology. *Trends Plant Sci*. **9**(8): 391-398.

Bergman, J.W., Carlson, G., Kushnak, G., Riveland, N.R., and Stallknecht, G. 1985. Registration of 'Oker' safflower. *Crop Sci.* **25**: 1127-1128.

Bergman, J.W., Baldridge, D.E., Brown, P.L., Dubbs, A.L., Kushnak, G.D., and Riveland, N.R. 1987. Registration of 'Hartman' safflower. *Crop Sci*. **27**: 1090-1091.

Bergman, J.W., Carlson, G., Kushnak, G., Riveland, N.R., Stallnecht, G., Welty, L.E., and Wichman, D. 1989. Registration of 'Girard' safflower. *Crop Sci*. **29**: 828-829.

Bergman, J.W. and Flynn, C. R. 2008. Evaluation of safflower and other oilseed crops grown in the United States Northern Plains Regions for biofuels/biobased products. *In* Proceedings of the 7th International Safflower Conference, Wagga Wagga, NSW, Australia, 3-6 November 2008.

Bernatzky, R. and Tanksley, SD. 1986. Toward a saturated linkage map in

tomato based on isozymes and random cDNA sequences. *Genetics* **112**: 887-898.

Bradley, V.L., Guenther, R.L., Johnson, R.C., and Hannan, R.M. 1999. Evaluation of safflower germplasm for ornamental use. In Janik, J. (Ed.), *Perspectives on new crops and uses*. ASHS Press: Alexandria, USA, pp. 433-435.

Cannon, J. and Cannon, M. 2003. Dye Plants and Dyeing. Timber Press Inc.: Kew, London, pp. 128.

Causse, M.A., Fulton, T.M., Cho, Y.G., Ahn, S.N., Chunwongse, J., Wu, K., Xiao, J., Yu, Z., Ronald, P.C., Harrington, S.E., Second, G., McCouch, S.R., and Tanksley, S.D. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* **138**(4): 1251-1274.

Cheung, F., Trick, M., Drou, N., Lim, Y.P., Park, J.-Y., Kwon, S.-J., Kim, J.-A., Scott, R., Pires, J.C., Paterson, A.H., Town, C. and Bancroft, I. 2009. Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. *Plant Cell* **21**(7): 1912-1928.

Dajue, L., and Mündel, H.-H. 1996. Safflower. *Cathamus tinctorius* L. Promoting the conservation of underutilized and neglected crops. Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany and International Plant Genetic Resources Institute, Rome, Italy.

Deshpande, R. B. 1952. Wild safflower (*Carthamus oxyacantha* Bieb.) – A

possible oilseed crop for the desert and arid regions. *Ind. J. Genet. Pl. Br.* **12**(1): 10-14.

Dille, J.E. and Knowles, P.F. 1975. Histology and inheritance of the closed flower in *Carthamus tinctorius* (Compositae). *A. J. Botany* **62**(3): 209-215.

Ekin, Z. 2005. Resurgence of safflower (*Carthamus tinctorius* L.) utilizaton: A global view. *J Agron*. **4**(2): 83-87.

Ellstrand, N.C. and Hoffman, C.A. 1990. Hybridization as an avenue of escape for engineered genes: strategies for risk reduction. *BioScience.* **40**: 438-442.

Flynn, C. and Bergman, J. Analytical chemistry methods used in the research and development of safflower varieties for the United States Northern Great Plains region. *In* Proceedings of the 5th International Safflower Conference, Williston, North Dakota and Sidney, Montana, 23-27 July 2001, pp. 295-299.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kobuki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G.S., and Saskati, T. 1998. A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* **148**(1): 479-494.

Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N.,

Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A., and Karp, A. 1997. Reproducibility tested of RAPD, AFLP, and SSR in plants by a network of European laboratories. *Mol. Breed.* **3**: 381-390.

Kaftka, S. 1965. Safflower production in California. Oakland, CA: University of California, Communication Services Publications, pp. 6.

Knowles, P.F. 1969. Centres of plant diversity and conservation of crop germplasm: Safflower. *Econ. Botany* **23**: 324-329.

Knowles, P.F. 1980. "Safflower." Hybridization of Crop Plants: Safflower. Madison, USA: American Society of Agronomy-Crop Science of America, pp. 535-548.

Knowles, P.F. 1989. Safflower. In Röbbelen, G. and Downey, R.K. (Eds.), *Oil Crops of the World.* McGraw-Hill, New York, pp. 363-374.

Kosambi D. D. 1944. The estimation of map distances from recombination values. *Ann. Eugen.* **12**: 172-175.

Kowalski, S.P., Lan, T.-H., Feldmann, K.A., and Paterson, A.H. 1994. Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics* **138**: 499-510.

Landaua, S., Friedmana, S., Brennera, S., Bruckentalb, I., Weinbergc, Z.G., Ashbellc, G., Henc, Y., Dvasha, L., and Lehsem, Y. 2004. The value of safflower (*Carthamus tinctorius* L.) hay and silage grown under Mediterranean conditions as forage for dairy cattle. *Livestock Prod. Sci.* **88**: 263-271.

Linder, C.R. and Schmitt, J. 1994. Assessing the risks of transgene escape through time and crop-wild hybrid persistence. *Mol Ecol.* **3**: 23-30.

Lizhong, Y. 1993. The study of new and medicinal oil of safflower. *In* Proceedings of the 3rd International Safflower Conference, Beijing, China, 14-18 June 1993, pp. 896.

Liu, Z.-W., Biyashev, R.M., and Saghai-Maroof, M.A. 1996. Development of simple sequence DNA markers and their integration into a barley linkage map. *Theor. Appl. Genet*. **93**: 869-876.

Lombard, V. And Delourme, R. A consensus linkage map for rapeseed (*Brassica napus* L.): construction and integration of three individual maps from DH populations. *Theor. Appl. Genet.* **103**: 491-507.

Ma, C. and Smith, M.A. 1985. Physical and genetic mapping of safflower chloroplast DNA. *Plant Molec. Biol.* **4**: 385-388.

Mayerhofer, M., Mayerhofer., R, Topinka., D, Christianson, J., and Good, A.G. (submitted). Introgression potential between cultivated safflower (*Carthamus inctorius*) and wild relatives within the *Carthamus* family.

Markley, N., Nykiforuk, C., Boothe, J., and Moloney, M. Producing proteins using transgenic oilbody-oleosin technology. BioPharm International, June 2006. Available at http://www.sembiosys.ca/Docs/Biopharma.pdf [Accessed Apr. 2010].

Moloney, D. 1998. The American Association of Oriental Medicine's Complete Guide to Chinese Herbal Medicine: How to treat illnesses and maintain wellness with Chinese herbs. Berkeley Books, NY, 292 pp.

Mündel, H.-H and Bergman, J.W. 2009. Chapter 14: Safflower. In Vollmann, J. and Rajcan, I. (Eds.), Handbook of Plant Breeding. *Oil Crops.* **4**: 423-447.

Mündel, H.-H., Blackshaw, R.E., Byers, J.R., Huang, H.C., Johnson, D.L., Keon, R., Kubik, J., McKenzie, R., Otto, B., Roth, B., and Stanford, K. 2004. Safflower production on the Canadian Prairies: revisited in 2004. Agriculture and Agri-Food Canada, Lethbridge Research Centre, Lethbridge, AB, 36 pp.

Murthy, I.Y.L.N. and Anjani, K. 2008. Fatty acid composition in *Carthamus* species. *In* Proceedings of the 7th International Safflower Conference, Wagga Wagga, NSW, Australia, 3-6 November 2008.

Nykiforuk, C.L., Shen, Y. Murray, E.W., Boothe, J.G., Busseuil, D., Rhéaume, E., Tardif, J.C., Reid, A., and Moloney, M.M. 2010. Expression and recovery of biologically active recombinant Apolipoprotein Al_{Milano} from transgenic safflower (*Carthamus tinctorius*) seeds. *Plant Biotech.* (in print) 14 pp.

Oelke, E.A., Oplinger, E.S., Teynor, T.M., Putnam, D.H., Doll, J.D., Kelling, K.A., Durgan, B.R., and Noetzel, D.M. 1992. Safflower. Alternate field crops manual. Available at http://www.hort.purdue.edu/newcrop/afcm/safflower.html. [Accessed Aug. 2010].

Pahlavani, M.H., Mirlohi, A.F. and Saeidi, G. 2004. Inheritance of Flower

Colour and Spininess in Safflower (*Carthamus tinctorius* L.). *J. Heredity* **95**(3): 265-267.

Palliard, S., Schnurbusch, T., Winzeler, M., Messmer, M., Sourdille, P., Abderhalden, O., Keller, B., and Schachermayr, G. 2003. An integrated genetic linkage map of winter wheat (*Triticum aestivum* L.). *Theor. Appl. Genet*. **107**(7): 1235-1242.

Panjabi, P., Jagannath, A., Bisht, N.C., Padmaja, K.L., Sharma, S., Gupta, V., Pradhan, A.K., and Pental, D. 2008. Comparative mapping of *Brassica juncea* and *Arabiopsis thaliana* using Intron Polymorphism (IP) markers: homoelogous relationships, diversification and evolution of the A, B and C Brassica genomes. *BMC Genomics* **9**:113.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalski, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* **2**: 225-238.

Ravikumar, R.L., Roopa, V.K., Soregaon, C.D., and Satish, D. 2008. Molecular diversity in *Carthamus* species and development of inter-specific mapping population for the first molecular map in safflower. *In* 7th International Safflower Conference, Wagga Wagga, Australia, 3-6 November 2008.

Rick, C. and Tanksley, SD. 1980. Isozymic gene linkage map of the tomato: applications in genetics and breeding. *Theor. Appl. Genet.* **57**: 161-170.

Rieseberg, L.H. and Wendel, J.F. 1993. Introgression and its consequences in plants. In Harrison, R.G. (Ed.), *Hybrid zones and the evolutionary process*. New York: Oxford University Press, pp. 70-86.

Sabzalian, M.R., Saeidi, G., and Mirlohi, A. 2008. Oil content and fatty acid composition in seeds of three safflower species. *J. Am. Oil. Chem. Soc.* **85**: 717-721.

Singh, R.J. 2007. Genetic resources, chromosome engineering, and crop improvement. CRC Press, Boca Raton, USA.

Singh, V., Rathod, D.R., Deshpande, M.B., Deshmukh, S.R., and Nimbkar, N. 2003. Breeding for wilt resistance in safflower. In: *Extended Summaries. National Seminar on "Stress Management in Oilseeds for Attaining Self-Reliance in Vegetable Oils"*. Hyderabad, India, 28-30 January 2003, ISOR, Directorate of Oilseeds, Rajendranagar, Hyderabad, pp. 368-370.

Smith, J.R. 1996. Safflower. AOCS Press, USA, pp. 40-45.

Soltis, D.E. and Soltis, P.S. 1990. Isozymes in Plant Biology. Chapman and Hall, London. pp. 5-11.

Song, Q.J., Marek, L.F., Shoemaker, R.C., Lark, K.G., Concibido, V.C., Delanney, X., Specht, J.E., and Cregan, P.B. 2004. A new integrated genetic linkage map of soybean. *Theor. Appl. Genet*. **109**: 122-128.

Sturtevant, A. H. 1913. The linear arrangement of the six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* **14**:43-59.

Van Gerpen J.V., 2005. Biodiesel fuel processing and production. *Fuel Process. Technol.* **86**: 1097-1107.

Vilatersana, R., Brysting, A.K., and Brochmann, C. 2007. Molecular evidence

for hybrid origins of invasive polyploids *Carthamus creticus* and *C. turkestanicus* (Cardueae, Asteraceae). *Mol. Phylogenetic. Evol.* **44**: 610-621.

Virk, P.S., Ford-Lloyd, B.V., Jackson, M.T., and Newbury, H.J. 1995. Use of RAPD for the study of diversity within plant germplasm collections. *Heredity* **74**: 170-179.

Vuylsteke, M., Mank, R., Antoinse, R., Bastiaans, E., Senior, M.L., Stuber, C.W., Melchinger, A.E., Lübberstedt T., Xia, X.C., Stam, P., Zabeau, M., and Kuiper, M. 1999. Two high density AFLP maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theor. Appl. Genet*. **99**: 921-935.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**(21): 4407-4414.

Yu, J.K., Tang, S., Slabaugh, M.B., Heesacker, A., Cole, G., Herring, M., Soper, J., Han, F., Chu, W.C., Webb, D.M., Thompson, L., Edwards, K.J., Berry, S., Leon, A.J. Olungu, C., Maes, N., Knapp, S.J. 2003. Towards a saturated molecular genetic linkage map for cultivated sunflower. *Crop Sci.* **43**: 367-387.

Zhang, L., Yan, J., and Chai, Q. 1997. Pharmacological study of safflower. *In* Proceedings of the 4th International Safflower Conference, Bari, Italy, 2-7 June 1997, pp. 339-346.

Zhudong, S., Wang, Z., Tu, J., Zhang, J., Yu, F., McVetty, P.B.E., and Li, G.

2007. An ultradense genetic recombination map for *Brassica napus*, consisting of 13551 SRAP markers. *Theor. Appl. Genet*. **114**: 1305-1317.

2. Linkage mapping of the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*[1]

2.1. Introduction

Safflower is used in North America today for birdseed, as a source of high quality cooking oil, and as livestock feed. SemBioSys Genetics Inc., the industrial collaborator that has funded much of this research, has been using safflower as their model organism for creating transgenic plants to produce plant-made pharmaceuticals (Markley *et. al*., 2006). High oleic acid producing cultivars of safflower have also been shown to have potential as feedstock in the production of biodiesel (Bergman and Flynn, 2008). Still, safflower is used much less frequently than other oilseed crops like canola and sunflower, largely because of its low yield, but also due to its vulnerability toward several diseases and poor competitive ability with weeds (Mündel *et. al*., 2004). Breeding programs to improve these traits have been restricted by the limited knowledge of genetic variability in *C. tinctorius* and a lack of genomic resources. Several studies have initiated the molecular characterization in safflower by developing DNA markers and using them to evaluate genetic diversity (Sehgal and Raina, 2005; Zhang *et. al*., 2006; Johnson *et. al*., 2007) and elucidate phylogenetic relationships (Vilatersana *et. al*., 2000, 2005; Garcia-Jacas *et. al*, 2001; Chapman and Burke, 2007; Bowles *et. al*., 2008; Sehgal *et. al*., 2008). Still, the safflower genome has

[1] A version of this chapter has been published.
Mayerhofer, M., Archibald, C., Bowles, V., and Good, A. 2010. Development of molecular markers and linkage maps for the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*. *Genome* **53**(4): 266-276.

not yet been sequenced, or until now, mapped. Genetic maps are useful for providing valuable knowledge to plant breeders regarding the genomic location of markers and closely linked traits, which one could target with a map based cloning approach. The objective of this chapter is to develop genetic material useful for further characterization of the safflower genome through the creation of an inter-specific linkage map between *C. tinctorius* and *C. oxyacanthus*.

Linkage mapping requires genetic variation such that the mapping population of plants has segregating genotypes at specific marker loci. Landraces of *C. tinctorius* are self-pollinating domesticated crops that are likely to become highly homozygous over time. It is expected that by using a wild *Carthamus* species in crossing, valuable genes that have been lost during domestication can be uncovered, and this may help restore the safflower gene pool. Jeweled distaff thistle (*C. oxyacanthus*) is a proposed progenitor of *C. tinctorius* with comparable oil content (Ashri and Knowles, 1960; Ravikumar *et. al*, 2008; Sabzalian *et. al*., 2008; Sehgal *et. al*., 2008) and was chosen as the wild *Carthamus* species for crossing with safflower. As *C. oxyacanthus* frequently cross-pollinates and has considerable morphological variation, its inclusion in the mapping population should add genetic variability and increase marker polymorphism, and as such, maximize the utility of the available marker information (Deshpande, 1952; Sabzalian *et. al*., 2009; Sabzalian *et. al*., 2010). There were two different *C. tinctorius* cultivars ('Centennial' and 'S-317') used in the backcross scheme for generating the inter-specific mapping population. The levels of intra-specific polymorphism between these two highly related cultivars could be simultaneously

tracked while screening for marker polymorphisms between the *C. tinctorius* and *C. oxyacanthus* parents.

The linkage mapping process involved comparing two-point recombination frequencies and subsequently grouping markers into linkage groups based on the likelihood of recombination between them. Chi-square tests were used to verify that Mendelian segregation was occurring between individuals for a given marker, and the Log of ODDs score method (LOD) was used to calculate the likelihood of marker linkage based on the observed pattern of segregation (Morton, 1955). The resulting linkage map is one that is most probable given the population size and the results of genotyping. The inter-specific map created here is based on markers developed from six sources, consisting primarily of microsatellite markers (SSR), with the exception of 40 restriction fragment length polymorphism (RFLP) markers.

In the laboratory, an intra-specific map was simultaneously prepared by Reinhold Mayerhofer using an F2 population derived from a cross between *C. tinctorius* cultivars Centennial and NP12. Since the two maps shared a common Centennial parent, they could be directly compared for syntenic regions. Thus, the two mapping populations were used to develop a composite map, which is the first detailed linkage map for any of the species in the *Carthamus* genus. This chapter describes the inter-specific map in detail, as it was created as part of the thesis project.

2.2. Materials and Methods

2.2.1. Developing the mapping population

Seeds were obtained from Richard Johnson of the US Department of Agriculture (USDA, Pullman, WA). The *C. tinctorius* cultivars Centennial (USDA PI 538779), S-317 (USDA PI 599253) as well as *C. oxyacanthus* (USDA PI 426185) were used as parents of the mapping populations. The *C. tinctorius* cultivar Centennial was first crossed with pollen from *C. oxyacanthus*. Two F1 plants were then backcrossed with pollen from two *C. tinctorius* cv. S-317 plants to create a population of 120 BC1 individuals. Forty-two BC-1-7 plants (1-7-1 to 1-7-42) were generated from crossing F1 plant 1 and S-317 plant 7, 60 BC-2-9 plants came from crossing F1 plant 2 and S-317 plant 9 (2-9-1 to 2-9-61), and 18 BC-1-9 plants came from crossing F1 plant 1 and S-317 plant 9 (1-9-1 to 1-9-18) (Figure 2.1). The growth chamber conditions used were 16 hour day lengths with 21°C day and 18°C night temperatures. Sunshine Mix4 soil was used and plants were fertilized every 2 weeks with 20-20-20 fertilizer. The chamber had an average light intensity of 349.3µmol/m^2/s. Plants were grown in 6" plastic pots and watered as needed. Plants were bagged before producing flowers to prevent out-crossing. While there were two cultivars of the *Carthamus tinctorius* species used, polymorphism among them was expected to be very low. In preparation for mapping, large-scale DNA isolations from the individuals of the backcross population were performed using a modified CTAB method (Doyle and Doyle, 1990).

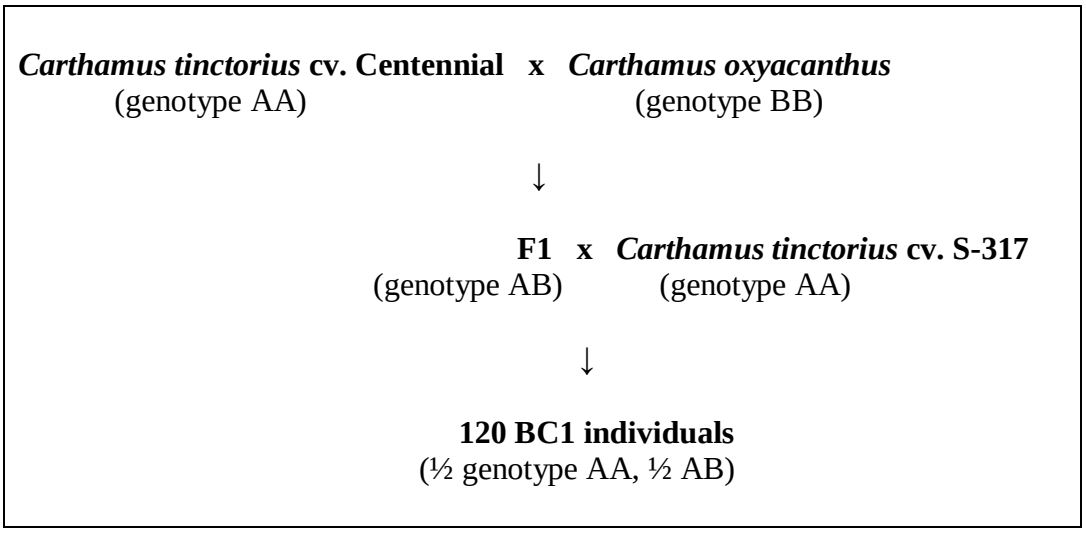

Figure 2.1: Generation of the BC1 mapping population for the inter-specific linkage map of *C. tinctorius* and *C. oxyacanthus*.

2.2.2. Genetic Material

Molecular markers used in linkage mapping were developed from a combination of sources, as described in Mayerhofer *et. al*. (2010). In short, this study used 972 SSR markers labelled 'ct' that were developed from EST unigenes of *C. tinctorius* cv. AC Sunset, publicly available from the Compositae Genome Project (CGP) at http://cgpdb.ucdavis.edu/cgpb2/est_info_library.php. EST unigenes from the CGP were also used to design 62 dinucleotide SSR primer pairs of spotted knapweed (*Centaurea maculosa*), 27 of which have been screened for creating the inter-specific map and are designated 'cm'. As both spotted knapweed and safflower are from the same 'thistle' tribe, it was hypothesized that this was an appropriate organism from which to source genetic material. A subset of 36 'VL' microsatellite markers derived from a genomic library of *C. tinctorius* cv. S-317, enriched for dinucleotide repeats, was also used in this study (Bowles

et. al., 2008). The subset of 'VL' markers chosen for this study were unique markers that previously showed polymorphism in an intra-specific parental screen among *C. tinctorius* cultivars Centennial and NP12 (Mayerhofer *et. al*., 2010). Another source of *C. tinctorius* sequences was provided by SemBioSys Genetics Inc. Thirty-seven SSR markers labelled 'ct' and 62 labelled 'gd' were designed from EST and genomic sequences of *C. tinctorius* cv. S-317, respectively. The detection of simple sequence repeats (SSRs) and the design of primers annealing around the SSR loci while incorporating the M13 (-21) sequence (5'-TGT AAA ACG ACG GCC AGT-3') are described in Mayerhofer *et. al.* (2010). The program PrimerPro (http://www.cs.ualberta.ca/~yifeng/primerpro/), which integrates the microsatellite identification tool MISA (http://pgrc.ipk-gatersleben.de/misa/) and the primer design program Primer3 (Rozen and Skaletsky, 2000), was used for this process. RFLP markers were developed from a cDNA library of *C. tinctorius* cv. S-317 and mapped in the backcross mapping population by Reinhold Mayerhofer prior to beginning my work. The cDNA library was made from 14-day-old seedlings using the Superscript Plasmid System with Gateway Technology for cDNA Synthesis and Cloning (Invitrogen, Carlsbad, California). There were 75 random clones chosen for amplification with polymerase chain reaction (PCR) using T7 and SP6 primers. Of these, 33 RFLP probes were polymorphic at 40 loci, resulting in 40 'S30' markers that were mapped in a subset of 30 BC1 individuals (Mayerhofer *et. al.*, 2010).

2.2.3. Microsatellite Marker Screening

A total of 1134 SSR markers were screened for polymorphism in the inter-specific mapping population. The microsatellite loci were amplified and simultaneously labelled with fluorescent dyes using a protocol adapted from Schuelke (2000). PCR reactions were performed in a Gene Amp 9700 thermocycler (Perkin-Elmer, Norwalk, CT) in 15µL reaction volumes containing 10 mM Tris-HCL (pH 8.0); 50 mM KCl; 0.01% gelatin; 7.5 mM $MgCl_2$; 0.2 mM each of dATP, dCTP, dGTP, and dTTP; 0.064 µM of the M13 primer; 2.5 units/reaction Taq DNA polymerase; and 25 ng of template DNA. The cycle parameters were 94°C/5 min; 30 cycles of 94°C/30 sec, 56°C/45 sec, 72°C/45 sec; followed by 8 cycles of 94°C/30 sec, 53°C/45 sec, 72°C/45 sec; and a final extension of 72°C/10 min. The M13 primers were labelled with fluorescent dyes FAM, VIC, NED or PET (Applied Biosystems, Foster City, CA), allowing for multiplexing of the PCR products before fragment analysis. The PCR products were diluted 10x in water and 2 µL were added to 8 µL formamide and 0.2 µL LIZ600 (Applied Biosystems) size standard. The samples were subsequently run on a 3730 DNA Analyzer (Applied Biosystems). Genemapper 4.0 software (Applied Biosystems) was used to display results.

2.2.4. Marker Genotyping

Markers showing polymorphism in parental DNA were then amplified in 66 individuals of the BC1 population using the same protocol, including the 30 BC1 individuals already used in RFLP analysis. The F1 DNA was also amplified

at each marker loci in order to verify locus heterozygosity. Markers with clean, co-dominant amplification for BC1 individuals were used in mapping. Homozygous genotypes with two *C. tinctorius* alleles were termed 'A', while heterozygous genotypes with one *C. tinctorius* allele and one *C. oxyacanthus* allele were called 'H'. Scoring of all mapped markers was triple-checked prior to making the final maps. Detailed marker information, including forward and reverse primer sequences, as well as the source library and reference information is provided in supplementary tables 2 and 3 of Mayerhofer *et. al.* (2010) for SSR and RFLP markers, respectively.

2.2.5. Mapping Software

The software MapDisto v1.7 Beta 132 for MS Windows was used to create maps from the scoring matrix (Lorieux, 2007). The final map displayed used the default settings of LOD score 3.0 and a maximum recombination frequency (rmax) of 0.3. The software calculates 2-point recombination frequencies ('Find Groups' command) and converts them into map distances in centimorgans (cM) by using Kosambi's mapping function (Kosambi, 1944). Linkage groups (LGs) are created by minimizing the sum of adjacent recombination fractions (SARF criterion), and using the Seriation II algorithm ('Order Sequence' command), which selects for a highly probable map based on expected genotypic ratios, as opposed to simply forming the shortest map. MapDisto calculates Chi-square tests to measure the deviation from the expected 1:1 (A:H) Mendelian ratio for all markers within an individual, and also to

measure segregation distortion between individuals within linkage groups. Markers with segregation ratio probabilities of P<0.05 were decidedly distorted. These are indicated on the map with asterisks (*0.01<P<0.05, **0.001<P<0.01, ***0.0001<P<0.001, ****1e-05<P<0.0001). The order of neighbouring loci within a linkage group was then checked with the 'Bootstrap Order' and 'Check Inversions' commands. A composite map, created for the purposes of comparing synteny between the *C. tinctorius* intra-specific map with this inter-specific map is illustrated in this study. However, only results of marker screening and mapped loci for the backcross population will be discussed in detail. Finally, linkage groups of the inter-specific map were also analyzed for colinearity with the *Arabidopsis thaliana* linkage map using BLASTN searches (www.ncbi.nlm.nih.gov).

2.3. Results

2.3.1. Marker Screening and Mapped Markers

Parental screening of 1134 SSR markers resulted in 221 polymorphic markers (19.5%). Rates of polymorphism and mapped markers varied among the different sources of SSRs, as summarized in Table 2.1. The highest rate of polymorphism, 38.9%, was found for the 'VL' markers developed from genomic DNA, followed by the 'ct' markers that originated from the CGP unigene sequences, of which 19.4% were polymorphic. The 'ct' markers from SemBioSys Genetics Inc. were polymorphic at a rate of 18.9%, while 'gd' and 'cm' markers showed the lowest level of polymorphism, at just 12.9% and 11.1%, respectively.

More than any of the SSR markers, the RFLP markers showed the most polymorphism in the backcross population, with 40 polymorphic loci resulting from the 75 cDNA randomly selected clones (53.3%).

In summary, only 126 of the 1134 screened SSR markers (11.1%) could be used in the map. A list of these markers and their respective amplification products is summarized in Table 2.2. Not surprisingly, the 'ct' markers from the CGP comprised the majority of mapped markers, with 109 of the 972 (11.2%) screened markers being mapped. Seven markers of each the SemBioSys Genetics Inc. 'ct' markers and the 'VL' markers could be mapped (18.9% and 19.4%, respectively). Finally, only two 'gd' markers and one 'cm' marker could be mapped (3.2% and 3.7%, respectively). Of the 95 SSR markers that showed parental polymorphism but were not used in the inter-specific map, the backcross individuals in 49 of these instances were monomorphic for this locus, as they were missing the *C. oxyacanthus* allele. It is expected that the frequency of the *C. oxyacanthus* allele would be reduced from the F1 to BC1 generation by 25%, which is consistent with these results, where 21% of *C. oxyacanthus* alleles were lost by the BC1 generation. In 30 instances where markers showing polymorphism in the parental screening were not mapped, marker amplification either failed in the mapping population, was too unclear to score, or there were too many missing data points for that marker. The remaining cases went unmapped because genotyping was too complex, as in the case of new alleles of unknown parentage.

Of the 126 SSR markers finally mapped, only 29 showed polymorphic alleles between the two *C. tinctorius* parents. A list of these markers, showing their fragment sizes and location on the linkage groups is given in Table 2.3. Due to this low level of polymorphism, no additional intra-specific map of the Centennial and S-317 cultivars of *C. tinctorius* was created.

Some intriguing observations were made for certain SSR markers. For marker ct984, it was found that the presence of the *C. oxyacanthus* allele changed the shape of the stutter bands for the *C. tinctorius* allele, but not the length of the amplified PCR fragment. For eight of the mapped microsatellite markers (ct266, ct285, ct297, ct383, ct390, ct476, ct657, and ct802), two *C. oxyacanthus* alleles were observed in the parental screening, while just one was observed in the F1 and BC1 offspring of the mapping cross. Interestingly, in the selfed progeny of the *C. oxyacanthus* parent only the alternate allele was amplified (Table 2.4). The reason for this phenomenon is not clear but it was hypothesized that it may be due to genomic instability in the hybrid plants at certain loci.

Table 2.1: The number of microsatellite markers analyzed, showing rates of polymorphism, number of mapped markers, source species, and genetic libraries used in marker development.

Marker Designation	Source Library	Source Species	Analyzed Markers	Polymorphic Markers	Mapped Markers
cm	EST	*C. maculosa*	27	3 (11.1%)	1 (3.7%)**
ct*	EST	*C. tinctorius* cv. AC Sunset	972	189 (19.4%)	109 (11.2%)
ct*	EST	*C. tinctorius* cv. S-317	37	7 (18.9%)	7 (18.9%)
gd	genomic	*C. tinctorius* cv. S-317	62	8 (12.9%)	2 (3.2%)
VL	enriched genomic	*C. tinctorius* cv. S-317	36	14 (38.9%)	7 (19.4%)
			1134	221 (19.5%)	126 (11.1%)

* The first set of 'ct' markers was sourced from EST libraries of the CGP and the second from SemBioSys Genetics Inc. (Mayerhofer *et. al.*, 2010).

**The percent values here indicate percent of analyzed markers that were able to be mapped.

Table 2.2: A list of SSR markers mapped in the BC1 population, giving the dye used and the size of each allele in base pairs (bp) for each of the parents *C. tinctorius* cv. Centennial (Cent) and cv. S-317 (S-317), and *C. oxyacanthus* (Oxy).

Marker	Dye	Cent	Oxy	S-317
cm022	FAM	157	163	157
ct006	FAM	279	264	279
ct015	VIC	285	279	281
ct024	FAM	280	307	280
ct026	VIC	141	135	137
ct032	PET	115	110	127
ct043	FAM	148	161	150
ct044	FAM	111	101	116
ct047	NED	129	120	129
ct124	VIC	406	414	406
ct125	NED	451	479	451
ct137	VIC	203	189	203
ct138	VIC	184	182	184
ct139	NED	175	163	175
ct156	VIC	329	322	329
ct167	PET	263	259	263
ct168	VIC	155	149	155
ct169	FAM	198	185	198
ct181	PET	212	224	219
ct185	FAM	189	192	189
ct195	NED	191	179	191
ct196	NED	286	284	286
ct201	PET	239	253	239
ct208	NED	148	138	148
ct216	FAM	114	107	114
ct222	VIC	247	241	247
ct227	FAM	178	180	178
ct233	FAM	151	147	151
ct239	FAM	241	269	266
ct266	VIC	140	145	140
ct274	VIC	538	528	538
ct279	NED	217	213	209
ct285	PET	234	231	234
ct297	PET	218	212	218
ct309	PET	177	173	177

Table 2.2 Continued: A list of SSR markers mapping in the BC1 population.				
Marker	**Dye**	**Cent**	**Oxy**	**S-317**
ct316	PET	184	178	180
ct331	FAM	139	137	139
ct333	VIC	217	219	217
ct335	VIC	294	296	297
ct337	FAM	211	230	211
ct351	NED	261	265	261
ct353	PET	253	255	253
ct360	FAM	262	271	262
ct361	FAM	262	265	262
ct370b	NED	284	286	284
ct381	NED	183	185	177
ct383	VIC	184	178	184
ct384	PET	233	221	233
ct390	FAM	198	200	198
ct405	NED	206	202	206
ct408	NED	266	258	266
ct410	PET	240	252	240
ct415	FAM	186	184	186
ct419	VIC	246	238	246
ct423	VIC	188	194	196
ct440	NED	213	235	213
ct448	NED	249	243	249
ct458	FAM	235	223	235
ct467	NED	250	252	250
ct473	PET	444	480	444
ct474	VIC	205	202	205
ct476	FAM	248	293	248
ct483	FAM	203	200	203
ct490	FAM	163	157	168
ct495	VIC	281	263	281
ct497	PET	218	191	218
ct504	NED	194	188	194
ct512	NED	248	245	248
ct518	PET	185	183	185
ct520	NED	230	212	234
ct531	VIC	185	163	185
ct535	VIC	424	422	409

Table 2.2 Continued: A list of SSR markers mapped in the BC1 population.

Marker	Dye	Cent	Oxy	S-317
ct536	PET	275	236	258
ct537	NED	436	444	436
ct540	NED	237	231	237
ct549	VIC	368	358	368
ct558	PET	270	231	272
ct563	VIC	324	334	321
ct588	VIC	394	364	376
ct590	NED	313	303	313
ct595	PET	432	436	432
ct598	FAM	368	344	374
ct605	FAM	250	229	250
ct619	FAM	323	315	323
ct642	VIC	374	366	373
ct643	NED	280	259	280
ct655	FAM	453	444	453
ct657	VIC	331	266	331
ct684	VIC	300	307	300
ct693	NED	194	198	194
ct698	VIC	427	384	424
ct706	PET	209	218	209
ct715	NED	169	128	167
ct754	PET	196	204	196
ct756	PET	158	152	158
ct763	PET	193	204	193
ct780	FAM	238	237	238
ct783	VIC	343	345	343
ct785	VIC	256	232	256
ct788	PET	243	233	237
ct802	FAM	335	308	335
ct804	NED	325	337	325
ct811	FAM	461	468	461
ct812	NED	230	232	230
ct816	VIC	225	231	225
ct820	PET	190	201	190
ct828	NED	197	181	197
ct830	FAM	223	221	223
ct831	VIC	174	170	178

Table 2.2 Continued: A list of SSR markers mapped in the BC1 population.

Marker	Dye	Cent	Oxy	S-317
ct833	PET	361	357	361
ct857	NED	222	221	222
ct858	PET	325	320	314
ct859	VIC	453	427	453
ct861	PET	317	303	317
ct895	FAM	225	223	226
ct977	FAM	193	197	193
ct981	NED	152	150	152
ct984	VIC	435	429	435
gd0010	VIC	265	254	265
gd0062b	PET	185	182	185
VL052	VIC	171	156	173
VL066	FAM	229	154	175
VL097	FAM	196	193	196
VL098b	NED	148	146	148
VL102	PET	266	260	276
VL108	PET	236	229	236
VL109	VIC	207	201	207

Note:
SSR Markers with 'b' after them are so called because another allele from the same marker was amplified previously in intra-specific mapping performed by R. Mayerhofer (Mayerhofer *et. al*., 2010).

Table 2.3: Markers showing polymorphism between *C. tinctorius* cv. Centennial and *C. tinctorius* cv. S-317.

SSR marker	Cent	S-317	Linkage Group
ct015	285	281	O7
ct026	141	137	O4a
ct032	115	127	O5
ct043	148	150	O3
ct044	111	116	O6
ct181	212	219	O2
ct239	241	266	O2
ct316	184	180	O1
ct335	294	297	O4b
ct381	183	177	O6
ct423	188	196	unlinked
ct490	163	168	O11
ct520	230	234	O4b
ct535	424	409	O3
ct536	275	258	O4b
ct558	270	272	O8b
ct563	324	321	O5
ct588	394	376	O4b
ct598	368	374	O1
ct642	374	373	O5
ct698	427	424	O7
ct715	169	167	O9
ct788	243	237	O4b
ct831	174	178	O8a
ct858	325	314	O3
ct895	225	226	O7
VL052	171	173	O4a
VL066	229	175	O1
VL102	266	276	O7

Table 2.4: Mapped microsatellite markers with two amplified *C. oxyacanthus* (Oxy) parental alleles, which showed alternate amplification in BC1 and F1 DNA compared to the selfed offspring of the *C. oxyacanthus* parent.

SSR Marker	Parental Oxy Alleles	Oxy Allele in F1 and BC1	Allele in Selfed Oxy	Linkage Group
ct266	135, 145	135	145	O5
ct285	231, 233	233	231	O5
ct297	212, 222	222	212	O3
ct383	178, 188	188	178	O3
ct390	194, 200	194	200	O3
ct476	269, 293	269	293	O3
ct657	266, 289	289	266	O1
ct802	303, 308	303	308	O5

2.3.2. Linkage Mapping

Using the default settings in MapDisto of LOD 3.0 and rmax of 0.3, thirteen linkage groups were found. There are three markers from the inter-specific map that remain unlinked, including two RFLPs (S30E5 and S30A04) and one SSR (ct423). Using the same MapDisto settings, the composite map was created by aligning the microsatellite markers that were mapped in both the intra- and inter-specific map, as shown in Figure 2.2 with dotted lines. The 'T' linkage groups are those belonging to the intra-specific *C. tinctorius* map, while the 'O' linkage groups belong to the inter-specific *C. oxyacanthus* map. All microsatellite markers mapped in common in the two plant populations mapped to the same linkage group (i.e. common microsatellite markers were found in linkage groups T1 and O1 or T4a and O4b, but not in T1 and O2 or T2 and O3, etc.). For markers ct370, gd0062, and VL098 different alleles of the same marker amplified in the two populations, which were called 'a' (i.e. ct370a) on the *C. tinctorius* and 'b' (i.e. ct370b) on the *C. oxyacanthus* linkage groups. Upon aligning the two maps, there is some evidence that the original 13 linkage groups developed from the backcross population might coalesce into 11 linkage groups. Linkage groups 04a and O4b may be part of one longer linkage group, as might O8a and O8b. The density of markers in both maps is not sufficient to definitively correlate 12 linkage groups with the 12 chromosomes of both *C. tinctorius* and *C. oxyacanthus*. Still, marker colinearity appeared to be well-conserved in regions where there were an adequate number of shared loci between the linkage groups. For linkage groups T5 and O5, the position of loci, particularly of ct137, ct266

and ct419, suggest that there has been a translocation or inversion event in that region. Loci divided among linkage groups T4a, T4b, O4a, and O4b may be indicative of evolutionary divergence between species, or simply a result of a lack of connecting markers.

The lengths of the linkage groups for the inter-specific map, calculated by summation of adjacent genetic distances, are as follows: 92.55cM (O1), 11.26cM (O2), 64.69cM (O3), 22.12cM (O4a), 83.82cM (O4b), 124.49cM (O5), 49.16cM (O6), 42.99cM (O7), 4.59cM (O8a), 18.97cM (O8b), 15.48cM (O9), 47.35cM (O10), and 2.33cM (O11). The inter-specific map covers a total genetic distance of 580cM, which is shorter than the 954cM of the intra-specific map and indicative of suppressed recombination in the inter-specific cross. Only on the inter-specific map is there dense clustering of markers, found on linkage groups O4b and O7 where 12 and 7 markers co-segregate, respectively, and thus map to precisely the same location. BLASTN searches using the sequences of the markers on the linkage groups did not reveal any colinearity with *Arabidopsis thaliana*. An example of BLASTN results is shown for linkage group O3 in Table 2.5. This table shows that for this LG, the top BLASTN hits come from multiple *Arabidopsis* chromosomes and as such, the two sections of DNA cannot be directly aligned between these species. The situation is the same for all linkage groups.

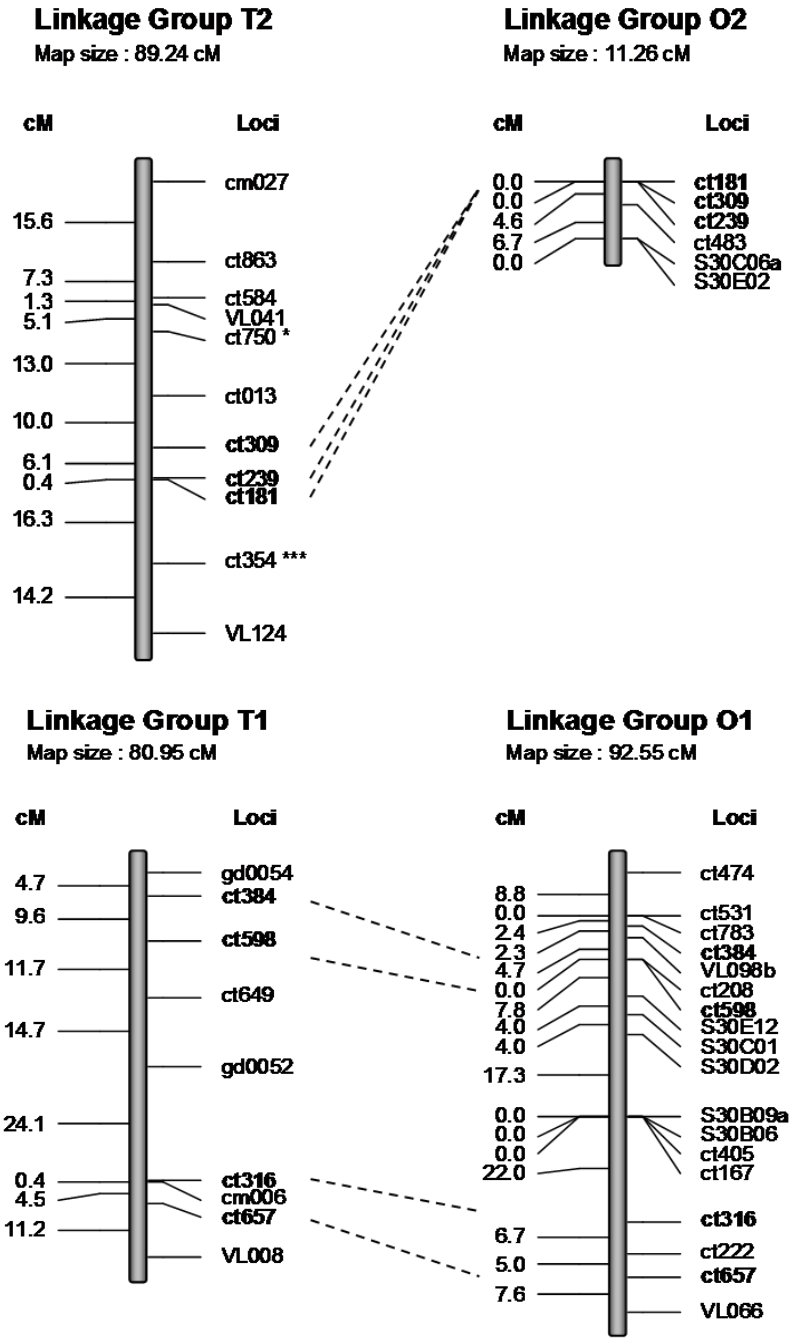

Figure 2.2: The composite linkage map as generated in MapDisto (LOD 3.0 and rmax 0.3) from combining an intra-specific map previously made from an F2 cross between *C. tinctorius* cv. Centennial and NP12 with 116 mapped microsatellite markers scored for 138 individuals (left hand side, 'T' linkage groups) and the inter-specific map of a BC1 generation from a cross with *C. oxyacanthus* and the recurrent *C. tinctorius* parent with 166 markers (126 SSRs and 40 RFLPs) scored for 66 individuals (right side, 'O' linkage groups). Dotted lines indicate the relative positions of microsatellite markers mapping in both populations.

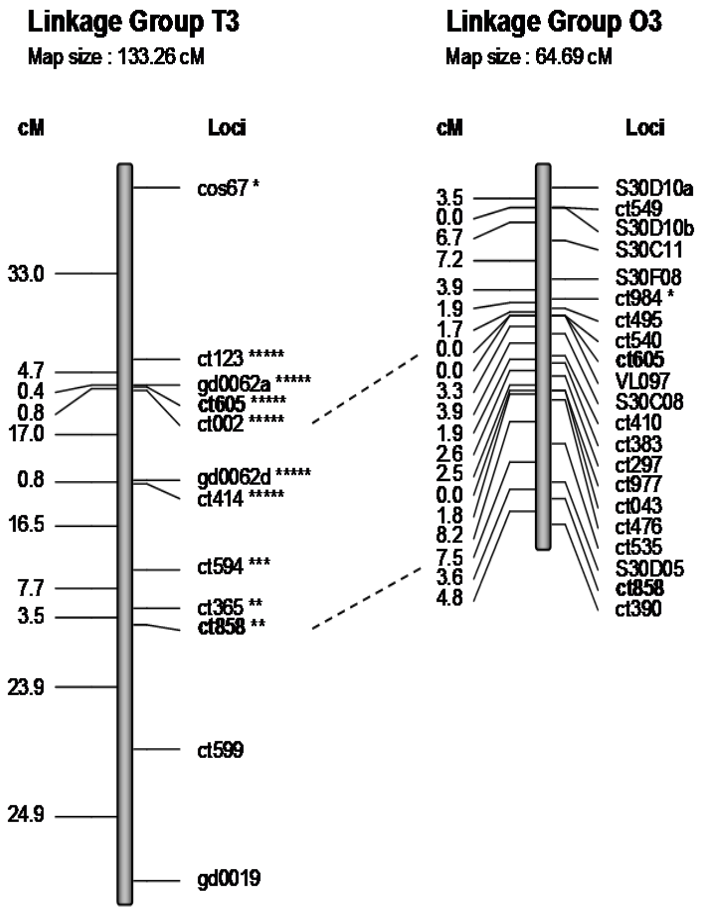

Figure 2.2 continued: The composite linkage map as generated in MapDisto.

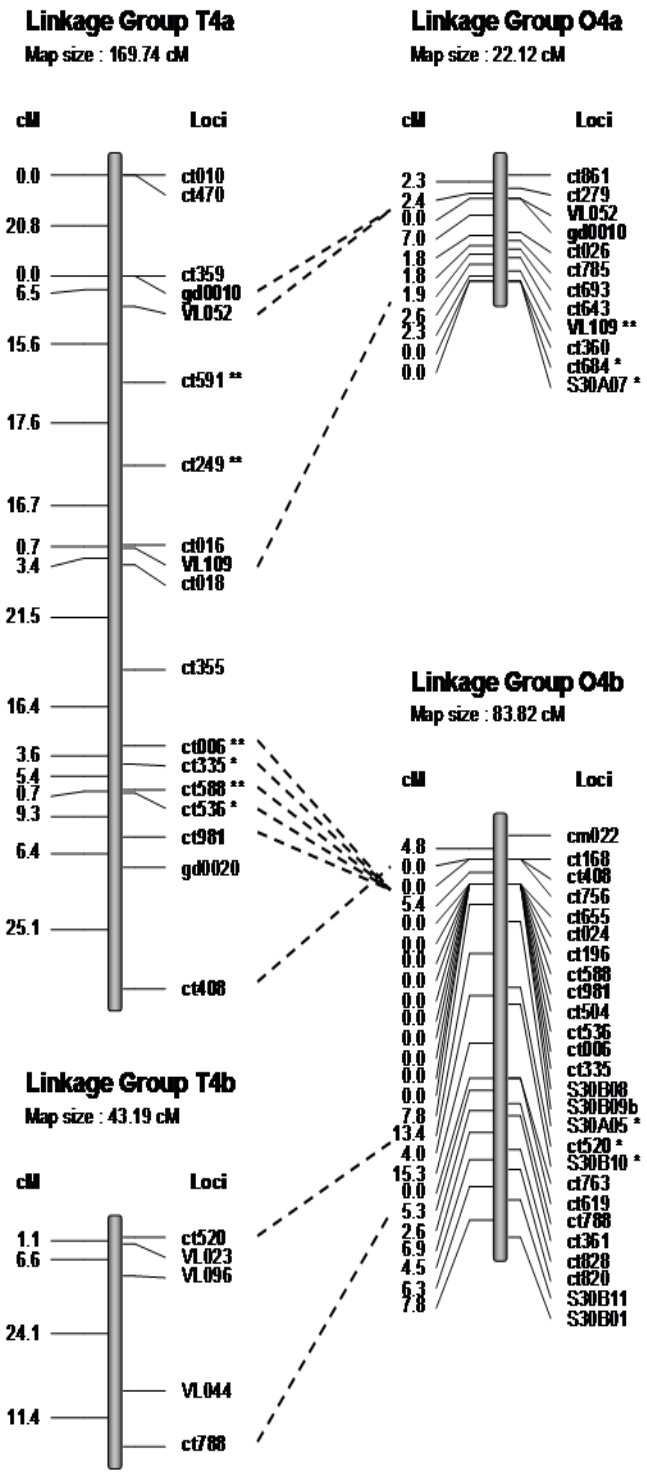

Figure 2.2 continued: The composite linkage map as generated in MapDisto.

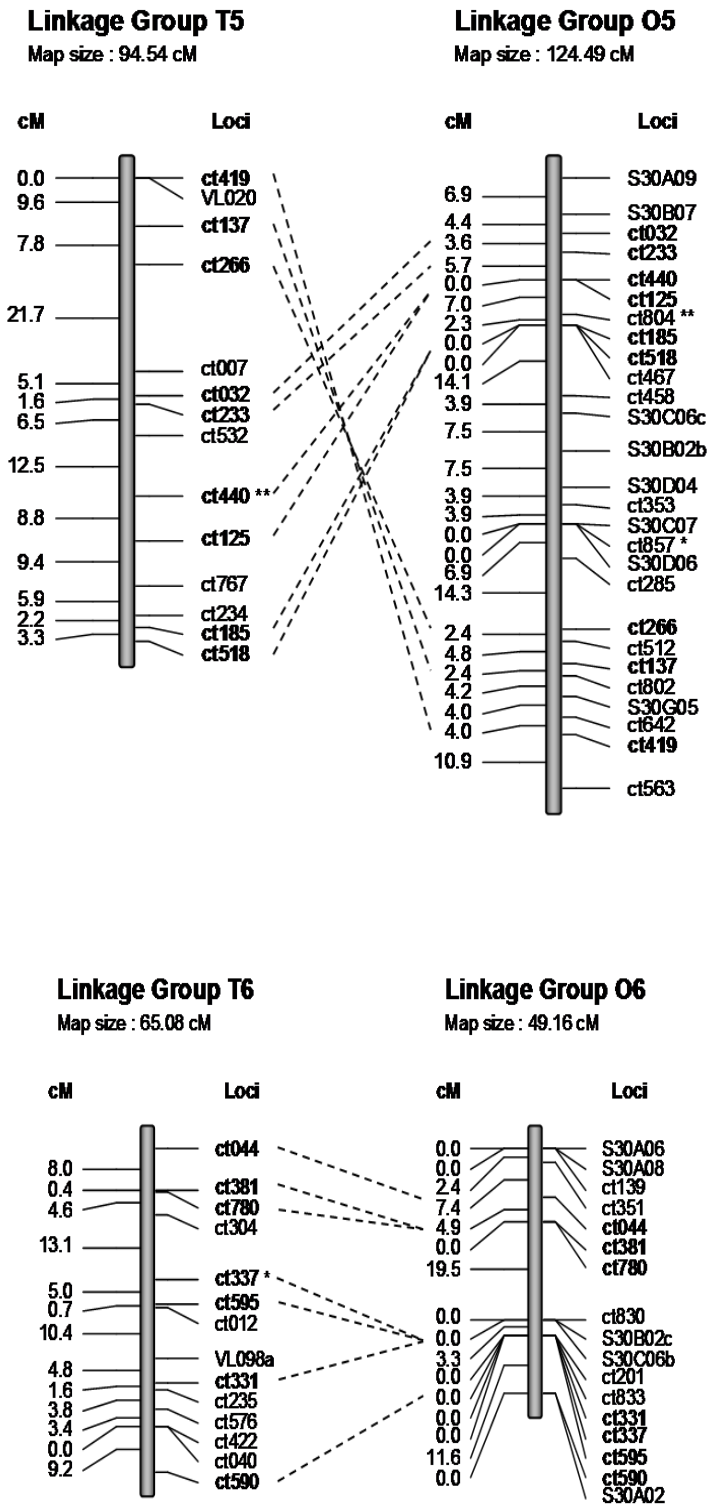

Figure 2.2 continued: The composite linkage map as generated in MapDisto.

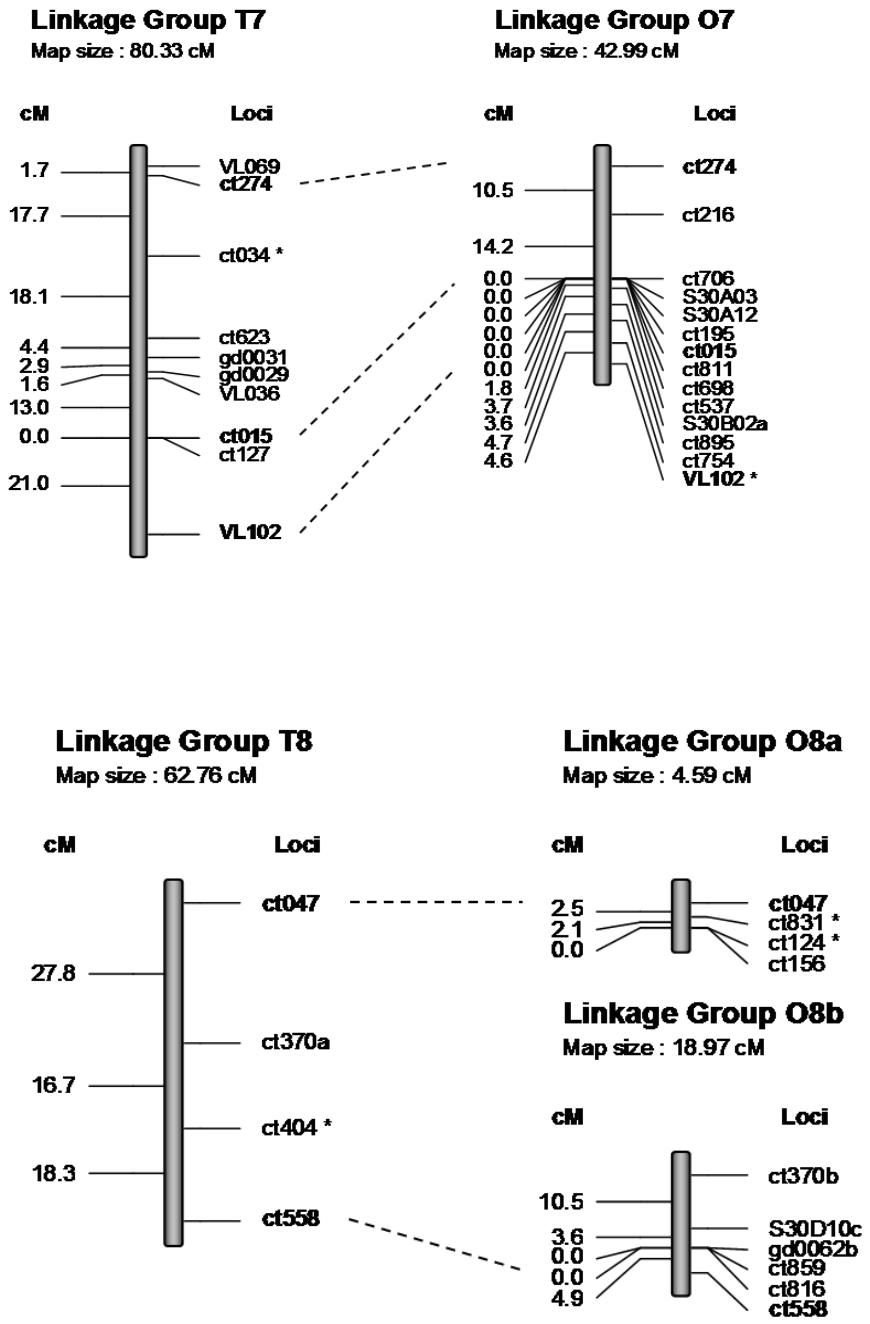

Figure 2.2 continued: The composite linkage map as generated in MapDisto.

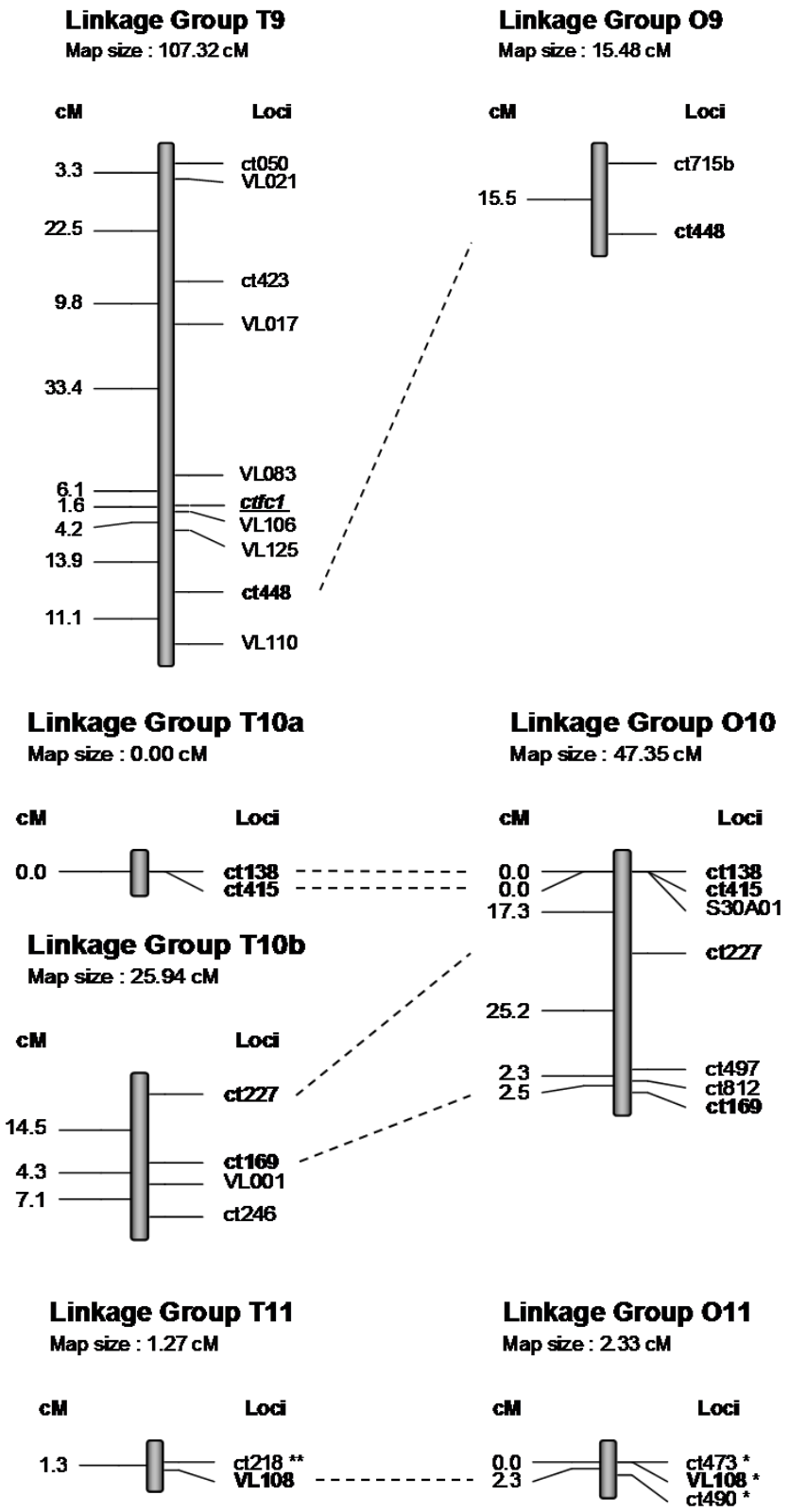

Figure 2.2 continued: The composite linkage map as generated in MapDisto.

Table 2.5: Top BLASTN results for nucleotide sequence similarity of linkage group O3, starting at the top of the linkage group, with *Arabidopsis thaliana*.

Marker	Top BLASTN Hit
ct390	gene BTR1L near At5G04420
ct858	At1G54450
S30D05	At5G05113
ct535	At5G40645
ct476	At5G67390
ct043	At4G10260
ct977	gene STE1 near At3G02590
ct297	At5G65960
ct383	At5G65960
ct410	At4G10260
S30C08	no hits
VL097	no hits
ct605	At2G35050
ct540	gene CDC25 near At5G03460, gene ATBAG3 near At5G07475
ct495	At4G13730
ct984	At2G22180, At2G46390
S30F08	At3G17845, At3G54360, At3G47100, At3G12140
S30C11	gene CYP97A3 near At1G31790
S30D10b*	At3G51730
ct549	At1G17910, At1G74456
S30D10a*	At3G51730

*These nucleotide sequences are the same.

2.4. Discussion

The work from this study and the resulting composite map marks the beginning of the first major linkage analysis for the *Carthamus* species. The markers used in the inter-specific map were SSR and RFLP markers, which are generally codominant markers that are transferable between mapping populations and species, particularly when they are sourced from EST sequences or cDNA clones. Prior to this work, largely dominant markers that depend upon PCR-amplified genomic sequences were used for various purposes in studies of

Carthamus species. Internal transcribed spacers (ITS) of nuclear ribosomal DNA and chloroplast gene sequences were used primarily for phylogenetic studies. More commonly, randomly amplified polymorphic DNA (RAPD), amplified fragments length polymorphism (AFLP), and inter simple sequence repeats (ISSR) marker types were used. In contrast to SSR and RFLP markers, which can be costly and time-consuming to develop, these markers are simple and inexpensive to develop. However, both technical difficulties of the assay and reproducibility among laboratories can be a problem and as such, maps based solely on anonymous markers are generally only useful for the populations from which they were created.

The project of mapping the backcross population has now exhausted six sources of genetic material. The molecular markers used took advantage of published EST sequences and PCR primers in addition to sequences provided by our industrial collaborator and those developed in our laboratory. Of the six marker types used on the inter-specific map, RFLP markers were by far the most polymorphic. This is not surprising, as RFLP markers have the potential to cross-hybridize to all the orthologs and paralogs of a gene, while amplification of SSR markers only occurs at the locus where the specific flanking primers anneal. It was observed that nearly all of the RFLP probes used bound to numerous restriction fragments in both *C. tinctorius* and *C. oxyacanthus*, with a range of 1 to 14 bands and an average of 5.3 alleles per marker. The reduced sequence-specificity and cross hybridizing potential of RFLP markers makes them more likely to detect polymorphic loci than PCR-based markers. However, RFLP

analysis is very laborious and requires the isolation of large amounts of high-quality genomic DNA, which can be challenging given the high polysaccharide content of the extracts, particularly of *C. oxyacanthus* (R. Mayerhofer, pers. observation).

The obvious difficulty in creating a saturated linkage map has been the lack of polymorphism between *C. oxyacanthus* and both *C. tinctorius* varieties and therefore the low percentage of mappable markers. Genetic variation between domesticated safflower and numerous wild relatives of the *Carthamus* genus has been explored previously in several studies using ISSRs (Ash *et. al*., 2003; Sehgal and Raina, 2005; Yang *et. al*., 2007; Sabzalian *et. al*., 2009; Sehgal *et. al*., 2009), RAPDs (Sehgal and Raina, 2005; Amini *et. al*., 2008; Ravikumar *et. al*, 2008; Khan *et. al*., 2009; Sehgal *et. al*., 2009), AFLPs (Sehgal *et. al*., 2005; Zhang *et. al*., 2006; Johnson *et. al*., 2007; Sehgal *et. al*., 2009), and conserved, intron-spanning PCR markers (Chapman and Burke, 2007). These studies included up to 193 safflower and wild accessions and reported up to 83% polymorphism. It is, however, difficult to correlate these results with the results of the present marker analyses, as different marker systems were used and none of the surveys included accessions from which our mapping populations were derived. Genetic diversity in *Carthamus* species has been evaluated using SSR markers in two other studies thus far. One is Bowles *et. al*. (2008), where the 'VL' markers, which have been used here in our mapping project, were exploited for the purpose of phylogenetic analysis, and the other is in Ravikumar *et. al*. (2008), where just 6.9% polymorphism was found between *C. tinctorius* and *C. palaestinus*. After the

submission of this study for publication, a paper was published on the development, polymorphism and cross-taxon utility of EST SSRs developed from the same CGP safflower ESTs (Chapman *et. al*, 2009). There was a huge difference in the amount of polymorphism found within safflower between the two studies. Chapman *et. al*. (2009) found that 89.4% of their accessions were polymorphic, compared with 8% found in the F2 intra-specific population (Mayerhofer *et. al*., 2010). Though the PCR protocols used differed slightly between the two experiments and many polymorphisms found in the parental screens could not be replicated in the mapping populations, the reason for this discrepancy is most likely due to the choice of germplasm. Unfortunately, there is no common safflower accession between the two studies to verify this, and *C. oxyacanthus* was not included in their work. While low levels of polymorphism do occur in self-pollinating cultivated crops, it was unexpected to see it to that extent, especially in an inter-specific cross. However, the observed level of polymorphism in the backcross population was still higher than the intra-specific cross of safflower (Mayerhofer *et. al.*, 2010). The overall lack of polymorphism between the two species is not entirely unexpected, as they both come from the same section (Carthamus) and are known to produce fertile F1 hybrids (MacPherson *et. al*., 2004). The low polymorphism found suggests that much of the morphological variation known to occur between *C. tinctorius* and *C. oxyacanthus* (Deshpande, 1952; Sabzalian *et.al*., 2009) is due to variation in single genes. This is consistent with the work of Ashri and Efron (1964) who, upon studying the mode of inheritance of morphological characters, concluded

that many of the morphological differences were caused by differences in single genes.

Of the SSR markers screened, it is not surprising that the 'VL' markers were the most polymorphic, given that the markers used were a subset chosen because they showed polymorphism in the intra-specific screen. Still, 50% of polymorphic 'VL' markers were not able to be mapped, largely because they failed to amplify PCR products in the mapping population. In the end, only 9 useful markers came from the genomic sequences (the 'gd' and 'VL' markers), likely because random non-coding sequences are less conserved among species than sequences from expressed genes. The remaining 118 mapped SSR markers were from EST sequences, with the majority of mapped markers being the 'ct' markers sourced from the CGP. The 'cm' markers from *Centaurea maculosa* were developed to help remedy the problem of low polymorphism, as polymorphism in more distantly related species is expected to be higher. While spotted knapweed was a logical choice from which to source additional markers given that it belongs to the same subtribe (Centaureinae) as safflower and has been studied by the CGP, these markers showed the lowest level of polymorphism in the backcross mapping population, and only one marker could be mapped. Other markers developed within the Compositae research community for comparative mapping and phylogenetic analysis (Chapman *et. al*., 2007; Heesacker *et. al*., 2008) were not screened in the backcross population as very low polymorphism was found in screens of the intra-specific mapping population from our lab.

Of the marker loci which were polymorphic among the *C. tinctorius* and *C. oxyacanthus* parents, only 67% could be mapped. This made the process of mapping quite inefficient as even with over 1100 SSR markers screened the final map was still sparse, showing gaps of up to 25.2cM between markers on the inter-specific map. As previously stated, the most common reason for this problem was that the backcross individuals were monomorphic for the locus despite the parental polymorphism, meaning that generally the *C. oxyacanthus* allele was not found in the backcross population. While it was expected that *C. oxyacanthus* allele frequencies would be reduced from the F1 to the backcross generation, it is also possible that the population size used was not sufficient to show inherent segregation. This is, however, highly improbable. It may be that certain *C. oxyacanthus* alleles were deleted upon crossing twice with the domesticated *C. tinctorius* species due to instability or preferential amplification. In other cases, polymorphic markers could not be mapped because amplification was unclear, or there were too many missing points in the data due to failed amplification. This could be because the annealing temperature for the PCR reactions was consistently set at 56°C for the first 30 cycles and 53°C for the last 8 cycles. Some markers have slightly different ideal annealing temperatures; however, optimizing and running different PCR conditions for every primer pair is impractical for the large number of markers that were screened and mapped in this project. It is noteworthy that all of the polymorphic 'ct' markers from EST libraries of *C. tinctorius* cv. S-317 were able to be mapped. These markers may have been easily amplified in the individuals of the backcross population because

of a combination of two effects, one being that of the highly-conserved nature of expressed gene sequences in relation to random non-coding sequences and the second that the S-317 cultivar was used as the recurrent parent in the backcross population and thus, S-317 alleles are found in all of the individuals.

For the eight SSR markers which had two amplified *C. oxyacanthus* alleles in the parental DNA and just one *C. oxyacanthus* allele in the F1 and BC1 DNA, there has been another case where SSR markers have been known to show genetic changes. Gaeta *et. al*. (2007) studied the genomic changes in the polyploid *Brassica napus* and found 71% of their SSR markers exhibited some changes among S5 lines. Genetic changes occurred on 36 of 38 chromosomes, and were found to be in higher frequencies in large regions of homology between the two genomes. If we consider that marker clustering, where 7 or more markers map to precisely the same location, occurs in regions of low homology between the *C. tinctorius* and *C. oxyacanthus* genomes, then the present study may support the idea that genomic changes are more common in homologous regions, as none of the eight microsatellite markers showing genomic changes mapped to the highly clustered linkage groups O4b and O7. Regardless, it does appear that there is allelic preference for one of the parental *C. oxyacanthus* alleles in the F1 and subsequently, BC1 individuals, and for the other allele in the selfed progeny of the *C. oxyacanthus* parent.

MapDisto created this map based on expected segregation ratios, under the conditions of rmax 0.3 and LOD 3.0. Given that a LOD of 3.0 means that the probability of seeing the observed segregation pattern is 1000x more probable

under a model of linkage than under independent assortment, the default MapDisto settings were sufficient for this population size. There is reasonable coverage of markers for half of the linkage groups, while the other linkage groups only carry up to only 6 loci. Better saturation of these regions will need to be carried out in the future. The Compositae family does not have a well-characterized and sequenced model plant like *Arabidopsis* that can be used to exploit marker colinearity and sequence databases in order to target specific chromosomal regions in the species being mapped (Mayerhofer *et. al*., 2005). That there were no syntenic regions found between the linkage groups in this study and *Arabidopsis* is not surprising, given the evolutionary divergence of the species. Even for the best-characterized members of the Compositae, sunflower and lettuce, only synteny at a fine scale covering regions of <5cM has been found with *A. thaliana* (Timms *et. al*., 2006). While sunflower and lettuce would likely show more homology with safflower, comparative mapping was not possible since the sunflower EST SSRs and 'universal' markers were not used in making the inter-specific map. Nonetheless is it expected that any synteny between these species would probably be fragmented and complex given that the chromosome number of sunflower (n=17) and lettuce (n=9) differs from that of safflower (n=12). Thus, the best approach to improving the saturation of the present composite map may be to exploit marker systems such as AFLPs and SNPs, which were not used in the present study.

There were limitations as to the level of detail with which the two linkage maps could be compared for synteny, as the specific SSR markers showing

polymorphism frequently differed between the two mapping populations. On linkage groups which had multiple shared markers, colinearity between the maps was well-preserved, despite low marker saturation. This supports earlier hypotheses that *C. oxyacanthus* is a progenitor of cultivated safflower, which are based on cytogenetic studies and molecular marker analysis (Ashri and Knowles, 1960; Ravikumar *et. al*, 2008; Sehgal *et. al*., 2008). Furthermore, the multi-locus character of the RFLP markers and most SSR markers indicates that the genomes of *C. oxyacanthus* and *C. tinctorius* are highly duplicated and complex. Distinct patterns of duplications and chromosomal rearrangements could not be established, as typically only one allele was polymorphic for each marker.

The dense clustering of marker loci in the inter-specific map on linkage groups O4b and O7 potentially highlights non-homologous regions of the species' genomes. These regions may be important for future studies interested in exploring the species genetic divergence. Alternatively, a portion of marker clustering may be because the population size was not sufficient to identify low levels of recombination events in these regions. The use of a larger mapping population in future linkage mapping attempts could clarify this issue. Even presently, the linkage maps may be used in plant breeding strategies, such as marker assisted selection as a means for crop improvement. Given an appropriately sized safflower mapping population and an improvement in marker saturation, the maps can be used for QTL analysis to identify candidate gene loci. Of particular interest would be the genes involved in oleic acid production because of its importance in biodiesel, and identifying genes related to drought

tolerance. Already, the linkage maps have been used in the laboratory for introgression studies between the *C. tinctorius* and *C. oxyacanthus* genomes (Chapter 3), and for mapping the gene for white flower colour in *C. tinctorius* (Mayerhofer *et. al*., 2010).

2.5. References

Amini, F., Saeidi, G., and Arzani, A. 2008. Study of genetic diversity in safflower genotypes using agro-morphological traits and RAPD markers. *Euphytica* **163**(1): 21-30.

Ash, G.J., Raman, R., and Crump, N.S. 2003. An investigation of genetic variation in *Carthamus lanatus* in New South Wales, Australia, using inter simple sequence repeats (ISSR) analysis. *Weed Res.* **43**(3): 208-213.

Ashri, A., and Efron, Y. 1964. Inheritance studies with fertile inter-specific hybrids of three *Carthamus* L. species. *Crop Sci* **4**(5): 510-514.

Ashri, A., and Knowles, P.F. 1960. Cytogenetics of safflower (*Carthamus* L.) species and their hybrids. *Agron. J.* **52**(1): 11-17.

Bergman, J.W. and Flynn, C. R. 2008. Evaluation of safflower and other oilseed crops grown in the United States Northern Plains Regions for biofuels/biobased products. *In* Proceedings of the 7th International Safflower Conference, Wagga Wagga, NSW, Australia, 3-6 November 2008.

Bowles, V., Hall, J., and Good, A. 2008. Creation of microsatellite markers for

investigation of relationships among closely related *Carthamus* species. *In* Proceedings of the 7th International Safflower Conterence, Wagga Wagga, NSW, Australia, 3-6 November 2008.

Chapman, M.A. and Burke, J.M. 2007. DNA sequence diversity and the origin of cultivated safflower (*Carthamus tinctorius* L.; Asteraceae) *BMC Plant Biol*. **7**: 60.

Chapman, M.A., Chang, J., Weisman, D., Kesseli, R.V., and Burke, J.M. 2007. Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theor. Appl. Genet.* **115**(6): 747-755.

Chapman, M.A., Hvala, J., Strever, J., Matvienko, M., Kozik, A., Michelmore, R.W., *et. al*. 2009. Development, polymorphism and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.). *Theor. Appl. Genet.* **120**(1): 85-91.

Deshpande, R. B. 1952. Wild safflower (*Carthamus oxyacantha* Bieb.) – A possible oilseed crop for the desert and arid regions. *Ind. J. Genet. Pl. Br.* **12**(1): 10-14.

Doyle, J.J. and Doyle, J.L. 1990. A rapid total DNA preparation procedure for fresh plant tissue. *Focus* **12:** 13-15.

Gaeta, R.T., Pires, J.C., Iniguez-Luy, F., Leon, E., and Osborn, T.C. 2007. Genomic changes in Resynthesized *Brassic napus* and their effect on gene expression and phenotype. *Plant Cell* **19:** 3403-3417.

Garcia-Jacas, N., Susanna, A., Garnatje, T., and Vilatersana R. 2001. Generic

delimitation and phylogeny of the subtribe Centaureinae (Asteraceae): a combined nuclear and chloroplast DNA analysis. *Ann. Bot. (Lond.)* **87**(4): 503-515.

Heesacker, A., Kishore, V.K., Gao, W., Tang, S., Kolkman, J.M., Gingle, A., *et. al*. 2008. SSRs and INDELS mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor. Appl. Genet*. **94**(5): 682-691.

Khan, M.A., von Witzke-Ehbrecht, S., Maass, B.L., and Becker, H.C. 2009. Relationships among different geographical groups, agro-morphology, fatty acid composition and RAPD marker diversity in safflower (*Carthamus tinctorius*). *Genet. Resour. Crop Evol.* **56**(1): 19-30.

Kosambi, D. D. 1944. The estimation of map distances from recombination values. *Ann. Eugen.* **12**: 172-175.

Johnson, R.C., Kisha, T.J., and Evans, M.A. 2007. Characterizing safflower germplasm with AFLP molecular markers. *Crop Sci.* **47**(4): 1728-1736.

Lorieux, M. 2007. MapDisto, A Free User-Friendly Program for Computing Genetic Maps. *In* Plant and Animal Genome XV Conference, San Diego, CA, 13-17 January 2007. Available from mapdisto.free.fr [Accessed Oct. 2009].

Macpherson, M.A., Good, A.G., Topinka, A.K.C., and Hall, L.M. 2004. Theoretical hybridization potential of transgenic safflower (*Carthamus tinctorius* L.) with weedy relatives in the New World. *Can. J. Pl. Sci.* **84**(3): 923-934.

Markley, N., Nykiforuk, C., Boothe, J., and Moloney, M. Producing proteins using transgenic oilbody-oleosin technology. BioPharm International, June 2006. Available from http://www.sembiosys.ca/Docs/Biopharma.pdf [Accessed Apr. 2010].

Mayerhofer, R., Archibald, C., Bowles, B., and Good, A. 2010. Development of molecular markers and linkage maps for the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*. *Genome* **53**(4): 266-276.

Mayerhofer, R., Wilde, K., Mayerhofer, M., Lydiate, D., Bansal, V.K., Good, A.G. and Parkin, I.A. 2005. Complexities of chromosome landing in a highly duplicated genome: toward map-based cloning of a gene controlling blackleg resistance in *Brassica napus*. *Genetics* **171**(4): 1977-1988.

Morton, N.E. 1955. Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**: 277-318.

Mündel, H., Blackshaw, R.E., Byers, J.R., Huang, H.C., Johnson, D.L., Keon, R., Kubik, J., McKenzie, R., Otto, B., Roth, B., and Stanford, K. 2004. Safflower Production on the Canadian Prairies: revisited in 2004. Agriculture and Agri-Food Canada, Lethbridge Research Centre, Lethbridge, AB, 36 pp.

Ravikumar, R.L., Roopa, V.K., Soregaon, C.D., and Satish, D. 2008. Molecular diversity in *Carthamus* species and development of inter-specific mapping population for the first molecular map in safflower. *In* 7th International Safflower Conference, Wagga Wagga, Australia, 3-6 November 2008.

Rozen, S., and Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. *In* Bioinformatics methods and protocol: methods in molecular biology. Edited by S. Krawetz and S. Misener. Humana Press. pp. 365-386.

Sabzalian, M., Saeidi, G., and Mirlohi, A. 2008. Oil content and fatty acid composition in seeds of three safflower species. *J. Am. Oil. Chem. Soc.* **85**(8): 717-721.

Sabzalian, M. R., Mirlohi, A., Saeidi, G., and Rabbani, M. 2009. Genetic variation among populations of wild safflower, *Carthamus oxyacanthus* analyzed by agro-morphological traits and ISSR markers. *Genet. Resour. Crop. Evol.* **56**(8): 1057-1064.

Sabalian, M.R., Saeidi, G., Mirlohi, A., and Hatami, B. 2010. Wild safflower species (*Carthamus oxyacanthus*): A possible source of resistance to the safflower fly (*Acanthiophilus helianthi*). *Crop Prot.* **29**(6): 550-555.

Schuelke, M. 2000. An economic method for the fluorescent labelling of PCR fragments. *Nat Biotechnol*. **18**(2): 233-234.

Seghal, D., and Raina, S. 2005: Genotyping safflower (*Carthamus tinctorius*) cultivars by DNA fingerprints. *Euphytica* **146**(1): 67-76.

Sehgal, D., Rajpal, V.R., and Raina, S.N 2008. Chloroplast DNA diversity reveals the contribution of two wild species to the origin and evolution of diploid safflower (*Carthamus tinctorius* L.) *Genome* **51**(8): 638-643.

Sehgal, D., Rajpal, V.R., Raina, S.N., Sasanuma, T., and Sasakuma, T. 2009.

Assaying polymorphism at DNA level for genetic diversity diagnostics of the safflower (*Carthamus tinctorius* L.) world germplasm resources. *Genetica* **135**(3): 457-470.

Timms, L., Jimenez, R., Chase, M., Lavelle, D., McHale, L., Kozik, A., *et. al*. 2006. Analysis of synteny between *Arabidopsis thaliana* and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements. *Genetics* **173**(4): 2227-2235.

Vilatersana, R., Susanna, A., Garcia-Jacas, N., and Garnatje, T. 2000. Generic delimitation and phylogeny of the *Carduncellus-Carthamus* complex (Asteraceae) based on ITS sequences. *Plant Syst. Evol.* **221**(1): 89-105.

Vilatersana, R., Garnatje, T., Susanna, A., and Garcia-Jacas, N. 2005. Taxonomic problems in *Carthamus* (Asteraceae): RAPD markers and sectional classifications. *Bot J. Linn. Soc.* **147**(3): 375-383.

Yang, Y.X., Wu, W., Zheng, Y.L., Chen, L., Liu, R.J., and Huang, C.Y. 2007. Genetic diversity and relationships among safflower (*Carthamus tinctorius* L.) analyzed by inter-simple sequence repeats (ISSRs). *Genet. Resour. Crop Evol.* **54**(5): 1043-1051.

Zhang, L., Huang, B.B., Kai, G.Y., and Guo, M.L. 2006. Analysis of intraspecific variation of Chinese *Carthamus tinctorius* L. using AFLP markers. *Yao Xue Xue Bao* **41**(1): 91-96. [In Chinese]

3. Phenotypic characterization and introgression analysis of populations developed from a cross between *C. tinctorius* and *C. oxyacanthus*

3.1. Introduction

The potential of safflower for use in transgenic experiments and as a supplier of biofuel (Bergman and Flynn, 2008), in addition to its use as birdseed and a healthy edible oil, may create new markets for safflower production in the Canadian prairies. In order to use safflower for these purposes, the plants need to be early maturing, have low height, high oil content and seed yield, with superior seed quality. Early maturity is especially important in Canada as it allows for harvesting before subjecting the crops to the stress of cold weather (Poehlman and Sleper, 1995). Reducing plant height has been necessary for the mechanization of safflower production (Weiss, 2000). High oil and seed yield are important for economical reasons, and superior seed quality is reflected in high germination rates and seed weight, as well as in the colour of the seed, where a solid white seed is preferred for use as birdseed. It is also important to consider certain morphological characters, such as leaf shape, which are often simply inherited and may influence the physiological processes that determine yield, by increasing crop photosynthesis for example (Thurling, 2003). The plant breeding efforts on safflower which have occurred in Canada to date have largely focused on promoting early maturity, high seed yield and low plant height. However, these limited breeding efforts have not improved qualities such as oil content and yield

enough for safflower utilization to be comparable with other major oilseed crops (Dajue and Mündel, 1996).

Previous breeding strategies employed for *Brassica* species have looked to related plant species and wild relatives to improve traits related to yield and disease resistance, and to broaden the gene pool through inter-specific crosses (Chungru *et. al*., 1999; Rashid *et. al*., 1994; Roy, 1977). Breeding strategies used in safflower can only be successful if there is enough genetic variation to allow for the incorporation of agronomically important traits in the germplasm. Genetic diversity of plant populations tends to be reduced with domestication, due to the population bottleneck that occurs in the process, especially in self-fertilizing plants (Doebly *et. al*., 2006). Since wild relatives can carry useful agronomic traits for resistance to disease, biotic and abiotic stress, and be used to re-introduce genetic diversity into established plant populations, safflower breeding programs would do well to take advantage of these additional gene pools. Upon identification and examination of novel traits in hybrid plant populations, crop development programs can focus on the introgression of these traits into the cultivated species with strategies such as backcross breeding, where hybrid plants are repeatedly crossed with the cultivated plant species and simultaneously selected for the desired 'wild' genotype (Acquaah, 2007). For studying the introgression of genes from wild relatives into cultivated species, the use of a species with shared common ancestry will allow for normal genetic exchange between homologous chromosomes by homologous pairing. *C. oxyacanthus* has been the proposed wild progenitor of *C. tinctorius* (Ashri and Knowles, 1960;

Ravikumar *et. al*., 2008; Sehgal *et. al*., 2008), and has the same diploid chromosome number (2n=24). It has comparable oil content with commercial safflower and exhibits drought-resistance (Sabzalian *et. al*., 2008). Recently, it has demonstrated potential for resistance to safflower fly, a pest that has limited safflower's expansion in numerous countries (Sabzalian *et al*., 2010). Furthermore, *C. oxyacanthus* has a higher number of capitula per plant than commercial safflower varieties which is an important consideration for plant yield, though it has not been subject to the selection and domestication efforts of increasing yield or earliness of maturity as has cultivated safflower (Able and Driscoll, 1976; Sabzalian *et. al*., 2009). Earlier studies have concluded that inter-specific hybridization with *C. oxyacanthus* could indeed introduce novel traits to cultivated safflower (Ashri and Efron, 1964; Sabzalian *et. al*, 2009), and F1 hybrids of this cross have previously shown a high degree of hybrid vigour (Deshpande, 1952). However, the phenotypic and genetic effects of this hybridization have not yet been examined in any generation beyond an F2 or BC1S1 population.

In this chapter, I examine the phenotypic characteristics of populations derived from single seed descent of an inter-specific backcross population which was developed from hybridization of commercial safflower, *C. tinctorius*, with its wild relative, *C. oxyacanthus*. Specifically, the BC1S2 generation was grown and analyzed in a field trial, and both the BC1S2 and BC1S3 generations were studied in growth chamber experiments. Phenotypic characterization was performed in order to search for novel phenotypic variants and the appearance of desirable

phenotypes, particularly relating to plant maturity and yield. The examination of phenotypes in these populations can answer questions about the correlation among certain traits, which is important to consider prior to performing selection of certain morphological characters in plant breeding. The experiments will reveal any transgressive segregation occurring between parental plants and their offspring, as well as the phenotypic diversity and average morphological values within the populations. Should the characterized plants display wide phenotypic variation or novel phenotypes of interest, the population can further be propagated to produce more homozygous recombinant inbred lines available for future experiments related to plant breeding. Genetic diversity in the BC1S3 plant population was explored with an introgression study, where microsatellite markers selected from the inter-specific *C. oxyacanthus* linkage map were used (Mayerhofer *et. al*., 2010). This analysis marks a continuation of the introgression study performed on the BC1S1 generation by Bowles (2010), where a subset of microsatellite markers were chosen based on their repeatability and approximately even distribution along a linkage group from each of the five linkage groups most saturated with microsatellite markers. The same markers were amplified in this study, but on a larger population size and after two further generations of self-fertilization, where ambiguities associated with heterozygosity should be significantly reduced. Through studying the introgression patterns in a BC1S3 population, the stability of the integration of the *C. oxyacanthus* genome into the *C. tinctorius* genome can be assessed, and regions showing selection for either species' genotype identified.

3.2. Materials and Methods

3.2.1. Development of Plant Populations

There were three separate studies performed, all using plant populations derived from the backcross population used previously in linkage mapping (Chapter 2). There were BC1S2 and BC1S3 plant populations used in these studies, which were created as described in Chapter 2. The *C. tinctorius* cultivars Centennial (USDA PI 538779), S-317 (USDA PI 599253) as well as *C. oxyacanthus* (USDA PI 426185) were the three parental plants used. First, *C. tinctorius* cultivar Centennial was crossed with *C. oxyacanthus*. Next, two F1 hybrid plants were backcrossed to two *C. tinctorius* cv. S-317 plants to create a population of 120 BC1 individuals. These BC1 individuals consisted of 42 BC-1-7 plants (1-7-1 to 1-7-42) that were generated from crossing F1 plant 1 and S-317 plant 7, 60 BC-2-9 plants that came from crossing F1 plant 2 and S-317 plant 9 (2-9-1 to 2-9-61), and 18 BC-1-9 plants resulting from crossing F1 plant 1 and S-317 plant 9 (1-9-1 to 1-9-18). Next, these 120 individuals comprising the backcross population were propagated by single seed descent (by bagging to ensure self fertilization), where a single seed produced from each BC1 plant was chosen and planted, producing 120 BC1S1 plants. The BC1S1 plants were then grown and upon self-fertilization, single seeds from each plant were again chosen and planted, which resulted in 103 BC1S2 plants. These BC1S2 plants were used in the current study. Finally, 98 of the 103 BC1S2 plants produced seeds. Single seeds produced from 48 of the seed-bearing BC1S2 plants were then chosen and grown for use in the BC1S3 growth chamber experiments on phenotypic

characterization. In addition, the *C. tinctorius* and *C. oxyacanthus* parental plants were also grown in the growth chamber and phenotypes were measured. The development of the plant populations is illustrated in Figure 3.1.

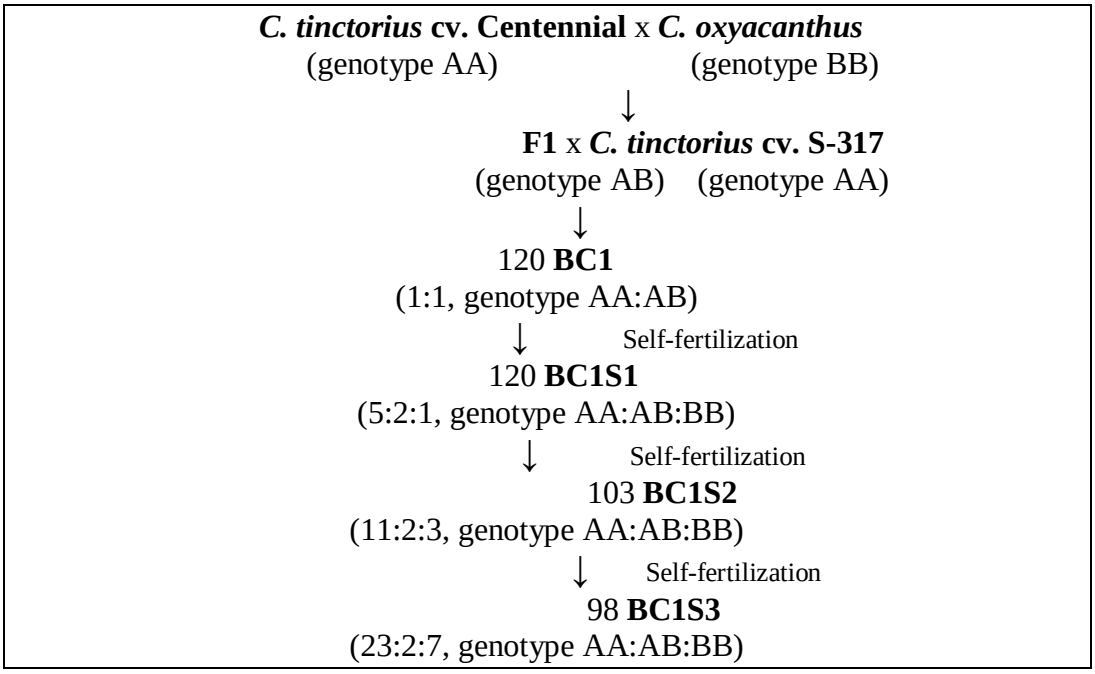

Figure 3.1: Development of the BC1S2 and BC1S3 plant generations through single seed descent of the BC1 mapping population, showing the expected Mendelian ratios of marker segregation.

3.2.2. Selection of Traits

The traits chosen for characterization were those which reflect plant growth, vigour, earliness and yield, as well as those showing notable variation in the population. In the early stages, germination rates and the number of pairs of true leaves were measured to reflect upon seed quality and seedling health. Due to the requirement of low height for the mechanization of safflower production, plant height was measured in both BC1S2 and BC1S3 generations. The time to first flowering and plant maturity were recorded, as they are important

productivity considerations. Data indicative of plant yield was collected, and included the number of capitula per plant, the total number of seeds, and the total seed weight. Seed quality was further estimated in the BC1S3 generation with calculations of the average individual seed weight. Leaf margin and leaf shape were measured as they may be of horticultural importance, particularly if increased lobes affects photosynthetic or transpiration rates by increasing surface area. Since leaf spines and the number of branches are defining features of *C. oxyacanthus*, they were also measured (Deshpande, 1952). The traits, considered to reflect yield and the amount of branching, were chosen because previous studies had shown that they exhibit high heritability and significant variation in hybrids with *C. oxyacanthus* (Amini *et. al*., 2008; Kavani *et. al*., 2000; Sabzallian *et. al.*, 2009). Thus, traditional breeding methods should easily be able to improve these traits.

3.2.3. Field Trial Method

The field trial was performed in Warner, Alberta, at the farm of our co-operator (Brian Otto). Safflower has successfully been grown for commercial use on this farm for over 20 years. A year prior to this study, another field trial was performed at the same farm on BC1S1 plants (Bowles, 2010) which were developed in the growth chamber from the backcross population used in linkage mapping (Figure 3.1). The present field trial used the offspring of that BC1S1 generation. Given that safflower reproduces largely by self-fertilization, the seeds planted in this field trial are considered to be BC1S2 seed (Khidir, 1968).

Field Planting

Seeding was done on May 4, 2009. The weather was cool and moist, and the plot had sandy loam soil, which had already been sprayed with herbicides. There were 31 lines of BC1S2 seed used in the 2009 field trial, and 4 parental lines (one line of *C. tinctorius* cv. Centennial, two of *C. tinctorius* cv. S-317, and one *C. oxyacanthus* line). Prior to seeding, preliminary experiments in the laboratory had shown a 50% germination rate of the BC1S2 seed that was to be used in the field trial. For this reason, 20 seeds were sown for each line, with the expectation that about 10 seeds per line would germinate in the field. The entire set of 35 lines was replicated on the field plot. Replicate lines were randomized on one side of the field plot, dividing the plot into two sections. The layout of the field was 10 rows, with 7 lines planted in 7 sections within each row, as shown in Figure 3.2. With the 20 seeds sown for each replicate of each line, a total of 1400 seeds were sown. Seeds were planted directly into the soil moisture, approximately 8cm below surface level. In order to allow for enough space to access individual plants for measuring phenotypes, seeds were sown 15cm apart and rows were 1m apart. The replicates, termed 'A' and 'B' of a given line, were not true seed replicates from the same parental plant, but rather were seeds selected from two different parental plants of the same parental line. This was done so that the variation between plants of a given line could be studied in addition to within-line variation.

It was noted while seeding that there was some apparent variability in soil hardness and colour throughout the plot. Surrounding the 10 rows of plants, a

dense border of the commercial safflower cultivar Saffire was planted to prevent edge effects. To discourage herbivory by deer and pronghorn antelope, a bright orange fence measuring 1.83m in height was put up around the plot, as it had been successful in discouraging large herbivores in the previous year (Bowles, 2010). The total area of the field plot was approximately 23 x 13m. Over the course of the growing season, plants were not fertilized, hand-watered or bagged; however weeding was done as needed. During planting, a new naming system for the lines grown in the field was developed to relate to their position in the field. This system of naming field lines in relation to the name of the BC1 line used in linkage mapping is illustrated in Table 3.1. This naming system was only used in the field trial, and both BC1S2 and BC1S3 generations studied in the growth chamber kept the original naming system as used for the BC1 population.

Row 10	**Row 9**	**Row 8**	**Row 7**	**Row 6**	**Row 5**	**Row 4**	**Row 3**	**Row 2**	**Row 1**	
21B	23B	1B	33B	13B	30A	23A	16A	9A	1A	**1**
17B	27B	7B	34B	10B	31A	24A	17A	10A	2A	**2**
20B	24B	5B	36B	14B	32A	25A	18A	11A	3A	**3**
18B	25B	2B	30B	11B	33A	26A	19A	12A	5A	**4**
22B	28B	6B	32B	9B	34A	27A	20A	13A	6A	**5**
19B	29B	3B	35B	12B	35A	28A	21A	14A	7A	**6**
16B	26B	8B	31B	15B	36A	29A	22A	15A	8A	**7**

Figure 3.2: The field plot layout of the 31 BC1S2 lines and 4 parental lines allocated in 10 rows and 7 sections, with replicates A on the right (white) and B on the left (grey).

Table 3.1: Naming system developed for the field lines showing the corresponding BC1 line. The parental lines *C. tinctorius* cv. S-317 (S-317-7 and S-317-9), *C. tinctorius* cv. Centennial (Cent) and *C. oxyacanthus* (Oxy) are highlighted in grey.

Field Line	BC1 Line	BC1S2 Field Plant ID	
		Replicate A plants	Replicate B plants
1	2.9.12	1.1.1-1.1.10*	8.1.1-8.1.10*
2	1.7.11	1.2.1-1.2.10	8.4.1-8.4.10
3	1.9.10	1.3.1-1.3.10	8.6.1-8.6.10
4	2.9.26	N/A	N/A
5	2.9.4	1.4.1-1.4.10	8.3.1-8.3.10
6	2.9.55	1.5.1-1.5.10	8.5.1-8.5.10
7	1.7.35	1.6.1-1.6.10	8.2.1-8.2.10
8	1.7.16	1.7.1-1.7.10	8.7.1-8.7.10
9	S317-7	2.1.1-2.1.10	6.5.1-6.5.10
10	1.7.32	2.2.1-2.2.10	6.2.1-6.2.10
11	1.7.26	2.3.1-2.3.10	6.4.1-6.4.10
12	2.9.49	2.4.1-2.4.10	6.6.1-6.6.10
13	2.9.16	2.5.1-2.5.10	6.1.1-6.1.10
14	1.7.37	2.6.1-2.6.10	6.3.1-6.3.10
15	1.7.36	2.7.1-2.7.10	6.7.1-6.7.10
16	2.9.33	3.1.1-2.1.10	10.7.1-10.7.10
17	2.9.11	3.2.1-3.2.10	10.2.1-10.2.10
18	2.9.2	3.3.1-3.3.10	10.4.1-10.4.10
19	2.9.54	3.4.1-3.4.10	10.6.1-10.6.10
20	1.7.8	3.5.1-3.5.10	10.3.1-10.3.10
21	2.9.35	3.6.1-3.6.10	10.1.1-10.1.10
22	S317-9	3.7.1-3.7.10	10.5.1-10.5.10
23	1.9.16	4.1.1-4.1.10	9.1.1-9.1.10
24	2.9.47	4.2.1-4.2.10	9.3.1-9.3.10
25	2.9.15	4.3.1-4.3.10	9.4.1-9.4.10
26	1.7.39	4.4.1-4.4.10	9.7.1-9.7.10
27	2.9.44	4.5.1-4.5.10	9.2.1-9.2.10
28	1.9.5	4.6.1-4.6.10	9.5.1-9.5.10
29	1.7.12	4.7.1-4.7.10	9.6.1-9.6.10
30	1.9.11	5.1.1-5.1.10	7.4.1-7.4.10
31	2.9.10	5.2.1-5.2.10	7.7.1-7.7.10
32	Oxy	5.3.1-5.3.10	7.5.1-7.5.10
33	Cent	5.4.1-5.4.10	7.1.1-7.1.10
34	2.9.1	5.5.1-5.5.10	7.2.1-7.2.10
35	1.7.21	5.6.1-5.6.10	7.6.1-7.6.10
36	2.9.57	5.7.1-5.7.10	7.3.1-7.3.10

*The first digit represents the row number (1-10), the second digit represents the section (1-7) in which the seeds of that replicate for each field line were planted, and the third digit represents the individual plant number (1-10, given a 50% germination rate).

Phenotypic Measurements

After seeding, successive trips to the field were done at 35, 49, 78, and 98 days, as well as for the final harvest at 164 days. Phenotyping in the field trial was assisted with the help of Dr. Hans-Henning Mündel, who has been Canada's leading safflower breeder for decades. On day 35, germination rates and the number of pairs of true leaves were counted. Germination rates were compiled from counts of seedlings for replicates of each line in relation to the 20 seeds originally sown. To check for new growth and the survival of seedlings, seedling counts were repeated at day 49. The plant stage for a given line was recorded at 49 days. Lines were graded 'R' if all seedlings remained in the rosette stage, 'E' if all the line's seedlings had begun stem elongation, and 'M' if seedlings of a replicate for a given line had a mixture of plants at the elongated and rosette stage (Mündel, pers. observation).

Between day 49 and 78, herbivory by deer and pronghorn antelope had occurred to such an extent that less than half of all the experimental plants in the field plot were intact and undamaged. This did limit further phenotypic characterization of these plants, however, measurements reflecting plant height, leaf spininess, and leaf shape were taken for the remaining plants on day 78. Height was measured in centimetres to the top of the plant. To estimate the degree of spinescence, the number of leaf spines on a leaf chosen from below the primary capitulum on the main axis was counted for each plant. It was observed at this point that leaf spinescence varied not just in spine number but also in the strength of leaf spines. Relative spine strength of the plants was graded as 1, 2, or

3 based on increasing leaf spine strength (Mündel, pers. observation). Figure 3.3 illustrates this phenotypic characterization. Leaf shape and leaf margin were graded in accordance with the International Board for Plant Genetic Resources (IBPGR) descriptors for safflower (1983), where the leaf shapes are designated as one of ovate (1), oblong (2), lanceloate (3), or linear (4), and leaf margins are either entire (1), serrated (2), or deeply serrated (3). Leaves chosen for measurements of shape were selected off the main stem when possible, and never subtending a branch.

Figure 3.3: Relative spine strength of field plants showing, from left to right, a weakly-spined (1), moderately-spined (2) and strongly-spined (3) leaf.

At day 98, flowering time was estimated. Lines were considered one of early (E), medium (M), medium-late (ML) or late flowering (L), where early flowering plants had 25% of the capitula flowering, medium meant the primary capitulum and numerous other capitula were in bloom, medium-late meant the primary capitulum and one or two others were in bloom, and late flowering plants

had no capitula in bloom by 98 days post seeding (Mündel, pers. observation). More specifically, this measurement estimates four categories whereby 25% of the flower heads were in bloom in under 100 days for early flowering plants, in 102 days for medium flowering plants, in 104 days for medium-late flowering plants, and in 106 or more days for late flowering plants. Finally, plants were harvested and bagged individually on day 164.

3.2.4. Growth Chamber Experiments

Growth chamber experiments were carried out under the same conditions as for the BC1 plants in Chapter 2. That is, a 16-hour day length with 21°C day and 18°C night temperatures was used. The soil used was Sunshine Mix4 soil, and plants were fertilized biweekly with 20-20-20 fertilizer. The average light strength in the growth chamber was 349.3µmol/m^2/s. After germination, plants were transplanted to 6" plastic pots and watered as needed. All plants were bagged prior to flowering in order to prevent out-crossing. Plants were rotated within the growth chamber in order to account for any environmental inconsistencies within the chamber. The BC1S2 plants were grown in the growth chamber first, beginning in May of 2009, followed by the parental plants which were planted on October 20th, 2009, and lastly the BC1S3 plants were seeded on November 16th, 2009.

Parental Plants

There were two and four *C. tinctorius* cv. Centennial and cv. S-317 plants measured, respectively, as well as three *C. oxyacanthus* plants. For the parental

plants, the phenotypic traits recorded during the growth cycle were: height, days to first flowering, the number of flower heads in bloom on day 91, the number of primary branches on day 91 and day 175, the total number of branches at day 175, the leaf margins and branch location on day 112, and the number of days to maturity. Plant height was measured in centimetres to the top of the plant on day 56 and 149. It was also measured just to the top of the main stem on day 91. Leaf margins were measured as in the field experiments. The location of branches along the main stem was recorded as outlined in the IBPGR descriptors (1983), where 0 = no branches, 1 = predominantly basal, 2 = predominantly upper third of the plant, 3 = predominantly upper two thirds of the plant, and 4 = from base to apex. Plants are considered mature when most of the leaves are brown and there is very little green remaining on the bracts of the youngest capitula (Mündel *et. al*., 2004). The number of days to maturity and to first flowering was counted from the date of seeding. The phenotypic measurements taken post-harvest for parental plants were the number of capitula, the number of seeds in the primary capitulum, the total number of seeds per plant, and the average weight per single seed.

BC1S2 Generation

There were 103 BC1S2 plants measured. In this generation, height, leaf margin and relative spine strength was measured the same as it was in the field trial. Also, the number of days to first flowering was recorded, as was the location of branching along the main stem and wilted flower colour. The same categorical classification of branch location used for parental plants was used in

both the BC1S2 and BC1S3 generation. As outlined in the IBPGR descriptors, the phenotypic categories for wilted flower colour are: 1 = grey-white, 2 = pale yellow, 3 = yellow, 4 = light orange, 5 = orange base, 6 = orange, 7 = deep red, and 8 = other (1983). After harvesting, seeds from 48 plants were randomly chosen for use in experiments on the BC1S3 generation.

BC1S3 Generation

The phenotypes characterized for the 48 plants analyzed in growth chamber experiments on the BC1S3 generation were: height, branch location along the main stem, days to first flowering, the number of primary branches at day 64 and 148, as well as the total number of branches at day 148, and the number of days to maturity. Height was measured in centimetres to the top of the plant on days 29 and 122, while it was measured on day 64 just to the top of the main stem. The same post-harvest measurements were taken as for the parental plants.

3.2.5. Introgression Analysis

Once seedlings of the BC1S3 plant population had multiple leaves, tissues were sampled and small-scale DNA extractions were done (DNeasy Plant Mini Kit, Qiagen). The five linkage groups (LG) showing the greatest marker coverage from the BC1 inter-specific *C. oxyacanthus* map (Chapter 2, Figure 2.2) were analyzed for trends in introgression of the BC1S3 generation. That is, linkage groups O1, O3, O4B, O5, and O6 were selected for further study. A subset of microsatellite markers was analyzed from these linkage groups, which were the

same as those used in Bowles' study (2010) and listed in Table 3.2. As done in linkage mapping, a 3-primer system was used for PCR reactions, which were performed in a Gene Amp 9700 thermocycler (Perkin-Elmer, Norwalk, CT). Reactions were performed and analyzed as in Chapter 2. Each locus was genotyped, with 'A' representing *C. tinctorius* alleles, 'B' representing *C. oxyacanthus* alleles, and 'H' indicating a heterozygote. As few markers showed intra-specific variation, alleles of both Centennial and S-317 cultivars of *C. tinctorius* were termed 'A'. Chi-square analysis was then performed for each marker locus to check for deviation from the expected 23:2:7 ratio of A:H:B genotypes. The scoring matrices were then used to comment on the patterns of integration of the *C. oxyacanthus* genome into the *C. tinctorius* genome and to note any genomic regions where selection occurred. Lastly, the pattern of segregation in each of the five linkage groups was examined in individual plants showing phenotypes that were extreme with relation to other plants in the population, in order to highlight any regions of the genome where associated genes may be located. This was done for the four individuals with the greatest and lowest plant height, the fewest days to flowering and maturity, as well as those with the highest values for total number of seeds, number of capitula and the number of seeds per primary capitulum.

Table 3.2: Microsatellite markers analyzed for introgression in the BC1S3 plant population chosen from five linkage groups (LG) of the BC1 inter-specific *C. oxyacanthus* map.

LGO1	LGO3	LGO4B	LGO5	LGO6
ct474	ct549	cm022	ct32	ct139
ct531	ct495	ct408	ct233	ct351
ct783	ct605	ct588	ct185	ct44
ct384	ct410	ct520	ct458	ct381
ct598	ct297	ct619	ct353	ct201
ct405	ct476	ct788	ct266	ct331
ct316	ct535	ct820	ct137	ct590
ct657	ct858		ct642	
	ct390		ct419	

3.3. Results

3.3.1. Field Trial Results

The entirety of the raw field data is given in Appendix A, and is summarized here. An overall image of the field plot at days 49 and 78 is given in Figure 3.4.

Figure 3.4: Images showing the field plot as seen from the south-eastern quadrant at 49 (left) and 78 (right) days post seeding.

Germination Rates & Seedlings

The total rate of germination for plants in the field was 28%. For the BC1S2 field lines, germination ranged from 0% (lines 3 replicate B and 18 replicate B) to 70% (line 17 replicate A). There were 11 BC1S2 lines and 2 parental lines found with greater than 40% germination at 35 days post seeding, as listed in Table 3.3. Six of these lines were also analyzed in the BC1S3 chamber experiments, where they again had greater than 40% germination. These were lines 5, 17, 23, 24, 26, and 31. Survival rates of germinated seedlings to day 49 ranged from 60-100%, with the large majority of lines showing a 100% survival rate. The number of true leaves counted on 35 day old seedlings varied from zero, which it was on seedlings that had only cotyledons, to nine pairs. The field lines that had the mean number of true leaves as 4.0 pairs or greater were lines 15, 20, 25, 28, 30, 34, 35 and the parental *C. tinctorius* cv. S-317 line 22.

Table 3.3: Summary list of BS1S2 and BS1S2 and parental lines and replicates (rep) showing over 40% germination rate in the field.

Line	% Germination
2 rep B	55
6 rep B	50
10 rep A	65
15 rep A	45
16 rep A	55
17 rep A	70
19 rep A	55
20 rep B	65
21 rep A	55
22 (S-317-9) rep A	85
25 rep B	60
33 (Cent) rep B	50
35 rep B	65

Plant Stage at 49 days

There were four lines and one parental line that had all the plants within the line elongated at 49 days as well as having greater than 40% germination. These were lines 15 (replicate B), 16 (replicate A), 25 (replicate B) and 35 (replicate B), and parental *C. tinctorius* cv. S-317 line 21 (replicate B). There were also 16 lines and 2 parental lines which had in-line variation at this time, with some plants being in a rosette stage and some already bolting. These were lines: line 1 (replicate B), line 2 (replicate B), line 5 (replicate B), line 8 (replicate B), line 9 (parental *C. tinctorius* cv. S-317, replicates A and B), line 11 (replicate A), line 14 (replicate B), line 17 (replicate A), line 19 (replicate B), line 21 (replicate A), line 23 (replicate B), line 24 (replicate B), line 26 (replicate B), line 27 (replicate B), line 28 (replicate A), line 31 (replicate A), line 33 (parental *C. tinctorius* cv. Centennial, replicates A and B), and line 35 (replicate A). It was observed that there were four notably vigorous lines of BC1S2 plants in the field plot which were large, bushy, and healthy. These were lines 2, 21, 25, and 35 (Archibald and Mündel, pers. observation).

Height

From field data measured at day 78, *C. oxyacanthus* parental plants were shorter on average (61.17 ± 6.31cm) than the *C. tinctorius* cv. S-317 plants (77.50 ± 3.67 and 71.80 ± 5.89cm). Mean plant height of the lines ranged from 61.71 ± 6.87cm in line 27 to 80.29 ± 4.92cm in line 28. Figure 3.5 shows the average plant heights with standard deviations for each of the lines with a minimum of two surviving plants by day 78. The lines with mean heights less than 65cm were

lines 10, 13, 19, 24, and 27. Lines with a mean plant height greater than 80cm were lines 20, 28, and 35.

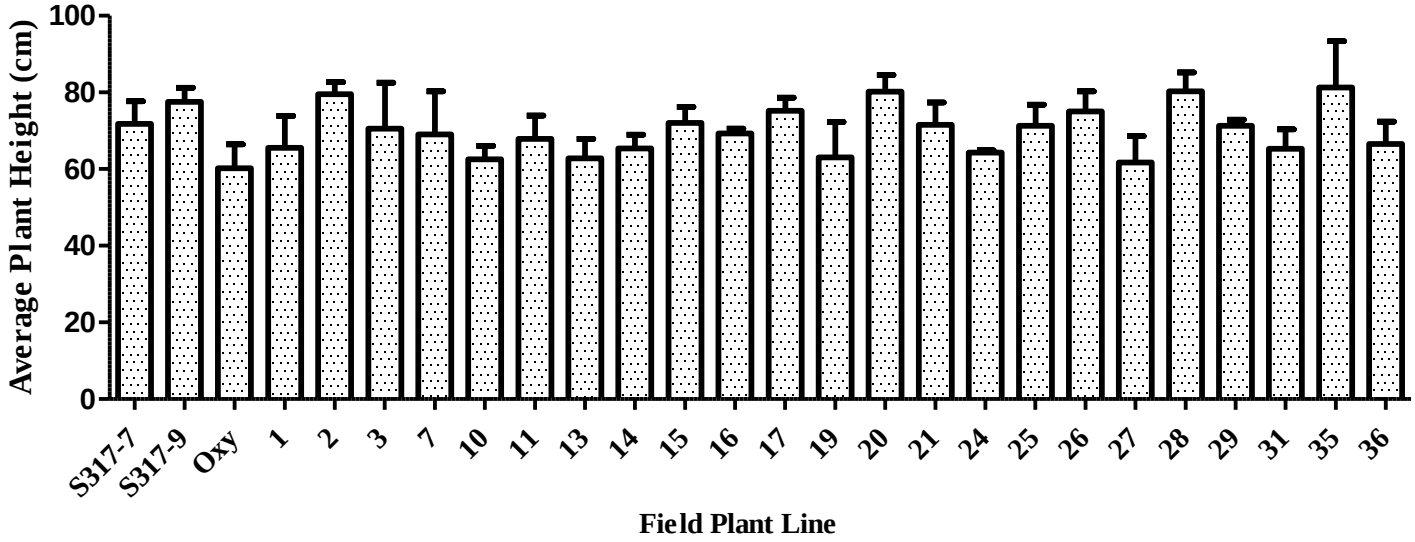

Figure 3.5: Mean plant heights with standard deviation for each BC1S2 field line at day 78. Field plant lines S-317-7, S317-9, and Oxy indicate two parental lines of *C. tinctorius* cv. S-317, and one of *C. oxyacanthus*.

Throughout the field plot, leaves were found in each of ovate, oblong, lanceloate and linear shapes, although linear leaves were rare and only seen in one plant in each of lines 6, 9, 15, and 20, and in two plants of line 29. It is noteworthy that 6 of the 9 plants in line 25 (replicate B) had linear leaves, and these plants were not predated, considered to be highly vigorous, and also found to be early flowering in growth chamber experiments. There were no lines which had only deeply serrated leaves and most lines had a mix of plants with serrated and deeply serrated leaves.

The number of leaf spines counted on a leaf ranged from a mean of 9.7 (line 19) to 16.0 (line 3). Figure 3.6 illustrates an example of the variation in spinescence occurring in the population, showing the appearance of a weakly-spined and more strongly-spined plant. As it was found that *C. oxyacanthus* plants were highly un-predated and had what were considered to be strong leaf spines, while both cultivars of *C. tinctorius* were highly predated and has less strongly-spined leaves (Archibald and Mündel, pers. observation), there was some speculation about the role of leaf spines in preventing herbivory. There were just 2 instances of herbivory on plants with a spine strength graded as 3, while there were 45 instances of herbivory on plants where the spine strength was graded as 1. Chi-square (X^2) analysis was used to test the null hypothesis that relative spine strength has no effect on predation (Table 3.4). The resulting Chi-square value was 41.9. Thus, at α of 0.05, that is, an acceptable type 1 error rate of 5%, the null hypothesis is rejected. This is indicative of a statistically significant relationship between relative spine strength and predation by deer. The nature of

this relationship, illustrated in Figure 3.7, is that predated plants had a higher percentage of plants with spine strength 1, while un-predated plants had a higher percentage of plants with spine strength 3. The percentage of plants with spine strength 2 was only slightly different between the two groups, with a higher percentage for predated plants.

Figure 3.6: An example of variation in spinescence in the BC1S2 field population, showing a strongly-spined (left) and weakly-spined (right) plant.

Table 3.4: Chi-Square analysis for the effect of leaf spine strength on predation.

	Predated	**Unpredated**
N (sample size)	27	127
Observed: Spine strength 1	10	15
Spine strength 2	15	67
Spine strength 3	2	45
Expected (N/3)	9.0	42.3
Calculated X^2	41.881	
Critical $X^2_{(0.05)(2)}$	5.991	

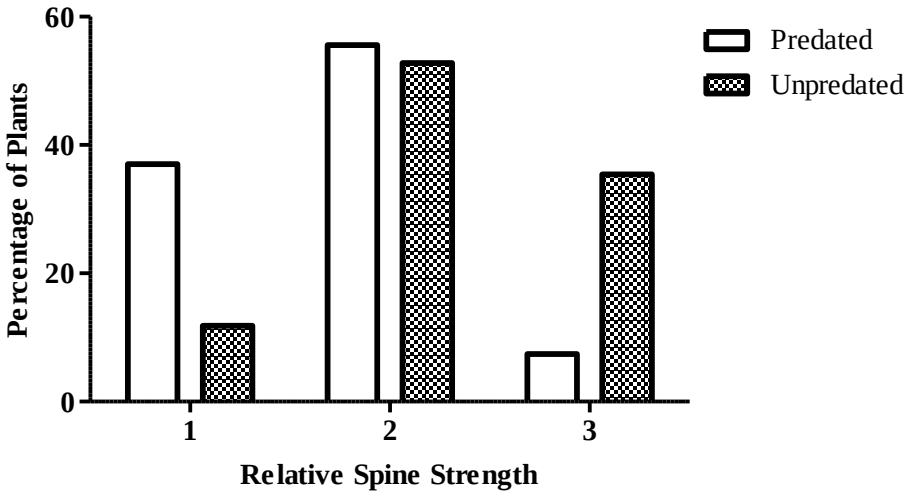

Figure 3.7: A comparison of the percentage of total predated and un-predated plants and their relative spine strengths. Data is grouped by predation.

Flowering time

In relation to the commercial Saffire cultivar of *C. tinctorius*, none of the lines in the field were early flowering. While some plants which had been predated were still scored for earliness of flowering, those measurements are not discussed here due to the potential for predation to influence flowering time. Generally, flowering time was largely homogeneous within a line with the exceptions of lines 14, 16, 17, 21, 24 and 35, which showed some in-line variation, to the greatest degree in line 21. Overall, there appeared to be more variation in flowering time between field lines than within lines. Earliness of flowering in the majority of plants was categorized as medium flowering, followed by medium-late flowering plants. Only three plants were considered late flowering. While none of the field lines were early flowering, there were eight BC1S2 field lines that had at least four of the plants within the line considered to

flower at a 'Medium' speed (Mündel, pers. observation). These were lines 1, 2, 14, 15, 20, 25, 28 and 35. Interestingly, three of the four lines previously said to show noticeable plant vigour (2, 25 and 35) are among those mentioned here.

3.3.2. Growth Chamber Results for Parental Plants

The measurements taken from growth chamber experiments for the parental plants are given in Appendix B, and post-harvest measurements are provided in Appendix C. The first notable difference between the *C. tinctorius* and *C. oxyacanthus* species was the longer rosette period of that of the latter species, which still had not bolted on day 56 (Figure 3.8). Plant height, measured on days 56, 91 and 149, had a combined mean for the parents of 25.2 ± 18.4cm, 40.7 ± 13.0cm, and 54.0 ± 25.2cm, respectively. The variation between the species and cultivar in each stage of measurement is shown in Figure 3.9. While both cultivars of *C. tinctorius* were roughly the same height, the vertical plant growth of *C. oxyacanthus* was delayed, reflective of its longer rosette stage. If not for the unusually tall Oxy-3 plant, *C. oxyacanthus* plant heights would have been comparable with the *C. tinctorius* cultivars at day 149. This large variation in plant heights of *C. oxyacanthus* is likely a reflection of the heterozygosity within this species.

Figure 3.8: Parental plants *C. oxyacanthus* (Oxy-3) in a rosette stage (left) and *C. tinctorius* cv. Centennial (Cent-2) in the stem elongation and branching stage (right) of plant growth on day 55.

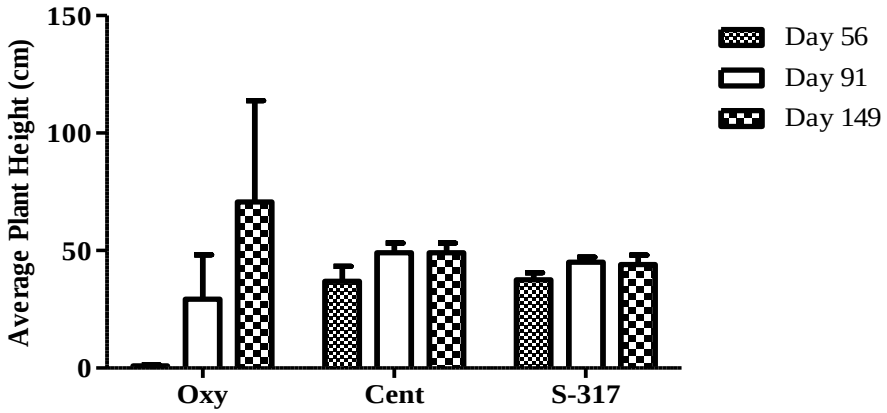

Figure 3.9: Average plant heights with standard deviation of parental plants *C. oxyacanthus* (Oxy), and *C. tinctorius* cultivars Centennial (Cent) and S-317 (S-317) at 56, 91, and 149 days post seeding.

The average days to first flowering combined for all parents was 97.7 ± 24.7 days. It was highly variable among the two species, with mean flowering times of 84.0 ± 1.4 days and 87.0 ± 4.2 days for *C. tinctorius* cultivars S-317 and Centennial, respectively, and 123.0 ± 31.2 days for *C. oxyacanthus*. Not surprisingly, the notably tall *C. oxyacanthus* plant flowered 54 days later than the average first flowering time of the other two *C. oxyacanthus* plants.

The only deeply serrated leaves were found on two *C. oxyacanthus* plants, while the other parental plants had serrated leaves. *C. oxyacanthus* plants also had branching from the base to the apex off the main stem by day 91, with the exception of one plant which was still in the rosette stage, while *C. tinctorius* cultivars had branches located on just the upper 1/3 or 2/3 of the main axis (Figure 3.10). The number of primary branches, measured first at day 91 and again at day 175, showed much more variation between species than between cultivars of *C. tinctorius*. The average number of primary branches at day 91 was 6.0 ± 5.2, 4.0 ± 1.4 and 4.8 ± 0.5 for *C. oxyacanthus* and *C. tinctorius* cultivars Centennial and S-317, respectively. Between the two measurements in time, *C. oxyacanthus* developed many more primary branches than *C. tinctorius* plants, as the mean number of primary branches on day 175 was up to 15.0 ± 1.7 for *C. oxyacanthus*, but was 4.5 ± 0.7 and 5.0 ± 0.8 for *C. tinctorius* cultivars Centennial and S-317, respectively. At maturity, the total number of branches in *C. oxyacanthus* parental plants, including secondary and tertiary branches, was an average of 83.3 ± 30.6 for *C. oxyacanthus*, but just 11.5 ± 5.0 and 19.3 ± 2.2 for the Centennial and S-317 cultivars of *C. tinctorius*, respectively.

The number of capitula in bloom on the parental plants on day 91 ranged from zero for the *C. oxyacanthus* plants to 5 for two of the *C. tinctorius* cv. S-317 plants. Maturation time was highly variable among the two species. The average number of days to maturity for *C. tinctorius* cultivars Centennial and S-317 was 173.5 ± 3.5d and 173.3 ± 1.0d, respectively, while the average days to maturity for the *C. oxyacanthus* plants was 213 ± 11.3d, which indicates that it alone would not be a viable crop in Alberta's short growing season. The post-harvest measurements for the number of flower heads, individual seed weight, total number of seeds, and number of seeds per primary capitulum were variable between plants of a certain cultivar as well as between cultivars. For all plants, seed yield was lower than expected, and one *C. tinctorius* cv. S-317 plant (S-317-2) had no seeds. This may be reflective of unfavourable conditions in the growth chamber during the parental plant generation as the temperature fluctuated as low as 15°C. Not surprisingly, the highest average number of capitula per plant ocurred in *C. oxyacanthus* plants, with 153 ± 113 capitula, far exceeding that of *C. tinctorius* cv. S-317 and cv. Centennial, which were 16 ± 4.2 and 31 ± 3.6, respectively.

Figure 3.10: Parental plants *C. tinctorius* cv. S-317 (left), cv. Centennial (centre), and *C. oxyacanthus* (right) on day 91, showing branching on the upper one third of the main stem in *C. tinctorius* and beginning at the base of the main stem in *C. oxyacanthus*.

3.3.3. Growth Chamber Results for BC1S2 Generation

The entirety of the plant measurements for growth chamber experiments of the BC1S2 plant generation are given in Appendix D, and are summarized here.

Flowering

The mean number of days to first flowering was 70.3 ± 6.6 days. The plants which flowered in less than 63.7 days, that is, more than one standard deviation below the mean time, were plants 1.7.21 (field line 35), 1.7.40, 1.9.17, 2.9.5, 2.9.10 (field line 31), 2.9.15 (field line 25), 2.9.17 , 2.9.19, 2.9.29, 2.9.30, 2.9.34, 2.9.52, and 2.9.59. While all flowers in bloom were yellow, wilted flower colour varied from yellow to a deep orange colour, with the most common colour being light orange.

Leaves

Relative leaf spine strength was highly variable among plants. Of the 103 BC1S2 plants measured, 47 (45.6%) were graded as weakly-spined, 29 (28.2%) had a relative spine strength of 2 and 27 (26.2%) were strongly-spined. Leaf margin was also variable among plants. Just like in the parental plants, the majority of plants had serrated leaves, but not deeply serrated leaves.

Branches

The clear majority of plants had branches located in just the upper 1/3 to 2/3 of the main stem, which is the phenotype of parental *C. tinctorius* plants. The exceptions to this were plants 1-7-35, 2-9-16, and 2-9-18, which had the

phenotype of *C. oxyacanthus*, where branching occurred throughout the main axis. One plant, 1-9-14, had no branching.

Height

The height of full grown plants averaged 62.7 ± 9.6cm. Plants less than 53cm tall that lived to produce healthy flowers were plants 1-7-18, 2-9-13, 2-9-21, 2-9-29, 2-9-30, 2-9-36, 2-9-37, and 2-9-57. Interestingly, there was no correlation between plant height and earliness of first flowering, with a Pearson correlation coefficient of just -0.002.

3.3.4. Growth Chamber Results for BC1S3 Generation

The measurements of growth chamber experiments for the BC1S3 plant generation are given in Appendix E, with post-harvest measurements provided in Appendix F. Results are summarized below.

Germination

The average germination rate was 60.8 ± 31.4%, which is much higher than in the field. The lines with 100% germination were: 1.7.36, 1.7.38, 1.7.40, 2.9.4, 2.9.7, 2.9.8, 2.9.10, 2.9.19, 2.9.38, and 2.9.40. As an indicator of the speed of germination, lines which had 50% or more seedlings emerge as early as Day 4 were lines 1.7.36, 2.9.8, 2.9.10, 2.9.16, and 2.9.38.

Height

The measurements of plant height on day 29 ranged from 0.4cm (plant 1.7.19) to 8.5 cm (plant 2.9.40), with the mean height being 4.6 ± 2.1cm. On day 64, mean plant height was 46.8 ± 7.4cm, and ranged from 32 (plant 2.9.47) to

60cm (plant 1.7.30). The final measurements of plant height on day 122 showed a range from 36cm (plant 2.9.10) to 82cm (plant 1.7.30), and a mean of 56.7 ± 9.6cm, which is comparable with the mean of parental plant heights.

Branches

The number of primary branches was highly variable among plants. On day 64 it ranged from 0 to 12, with the mean as 5.5 ± 2.6 primary branches, and on day 148 it ranged from 0 to 13, with the mean as 7.8 ± 3.5 primary branches. The total number of branches on day 148, including secondary and tertiary branches, ranged from 11 to 52, and the mean total number of branches was 24.2 ± 8.4. The plants with 33 or more branches, which is more than one standard deviation above the mean, were plants 1.7.14, 1.7.19, 1.7.41, 1.9.6, 2.9.15, 2.9.16, and 2.9.36. Most plants had the *C. tinctorius* phenotype of branch allocation, with most branches primarily located on the upper 1/3 to 2/3 of the main axis. Plants with branches located throughout the main axis were plants 1.7.19, 1.9.4, 1.9.6, 1.9.17, 2.9.10, 2.9.16, 2.9.37, and 2.9.47. Plant 2.9.16 also showed this phenotype in the BC1S2 generation.

Flowering & Maturity

Mean days to first flowering was 94.2 ± 5.7 days. Plants which flowered in less than 88.5 days, that is, more than one standard deviation below the mean time, were plants 1.7.13, 1.7.31, 1.7.34, 1.9.3, 2.9.15, 2.9.36, and 2.9.37.3. For the BC1S3 plant population, the average number of days to maturation was 182 ± 23.9d. Early maturing plants, which are plants that matured in 158 days or less, were plants 1.7.24, 1.7.39, 1.9.8, 2.9.11, and 2.9.40.

Post-Harvest Measurements

The number of capitula per plant ranged from 13 to 64, with a mean of 31 ± 11.2 capitula. This was higher than the average parental counts for *C. tinctorius* cv. Centennial, but not higher than the average for the *C. tinctorius* cv. S-317 or *C. oxyacanthus* parents. Largely, there were no seeds found in the primary capitulum of the BC1S3 plants. The highest number of seeds in the primary capitulum was found for plants 1.7.14 and 1.9.18 with 10 and 13 seeds, respectively. The average single seed weight for the plant population was 54 ± 11mg. Plants 1.9.8, 2.9.4, 2.9.15, 2.9.16, 2.9.30, and 2.9.45 had an average individual seed weight of 65mg or higher, which is one standard deviation above the mean or greater. This is likely indicative of greater oil content in seeds from these plants. Lastly, measurements of the total number of seeds per plant showed a range from 0 to 468 in the BC1S3 population, with a mean of 122 ± 118.7 seeds. Plants 1.7.13, 1.7.14, and 1.7.38 had the most seeds, with 354, 468, and 465 seeds, respectively. Interestingly, none of these plants had above average individual seed weights, which suggests there might be a trade off between seed number and seed quality.

Phenotypic Correlations

The correlations between different phenotypic measurements are given in Table 3.5. The traits which showed less than 10% correlation were: height and primary branches, height and total branches, days to flowering and primary branches, days to flowering and total branches, primary branches and number of seeds, days to flowering and days to maturity, number of seeds and days to

flowering, number of seeds and number of capitula, total branches and 1 seed weight, days to maturity and 1 seed weight, and number of capitula and one seed weight. Not surprisingly, the number of primary branches and total branches showed a strong positive correlation (78%). An important finding was that the days to maturity and height showed a strong negative correlation (61%), indicating that the taller plants matured faster than shorter plants. Correlations between total branches and number of capitula and also primary branches and flower heads were 64% and 51%, respectively. The measurements for number of seeds and height had a positive correlation of 40%, suggesting that taller plants yield a greater number of seeds.

Table 3.5: Pearson correlation coefficients (R) for traits measured in growth chamber experiments of the BC1S3 population. Correlations less than ±0.4 were considered non-significant (Ns).

	Primary Braches	**Total Branches**	**DF**	**DM**	**No. Capitula**	**No. Seeds**	**1 Seed weight**
Height	Ns	Ns	Ns	-0.61	Ns	0.40	Ns
Primary Branches		0.78	Ns	Ns	0.51	Ns	Ns
Total Branches			Ns	Ns	0.64	Ns	Ns
DF				Ns	Ns	Ns	Ns
DM					Ns	Ns	Ns
No. Capitula						Ns	Ns
No. Seeds							Ns

3.3.5. Results of Introgression Analysis

Results of introgression and Chi-square analysis (X^2) are given in Figures 3.11a-3.11e. Any 'B' scoring, that is, any locus homozygous for the *C. oxyacanthus* allele, indicates a location where stable integration of the *C. oxyacanthus* genome into the *C. tinctorius* genome has occurred. Based on these results, there are regions of stable genomic integration found on all five linkage groups. The number of individuals retaining some amount of *C. oxyacanthus* DNA was 29 for LGO1, 19 for LGO3, 30 for LGO4B, 22 for LGO5, and 20 for LGO6. The degree to which the *C. oxyacanthus* genome integrates varied among linkage groups. In linkage group O1, there appears to be the least degree of genome mixing, as many individuals have large segments or blocks of one of either *C. tinctorius* or *C. oxyacanthus* alleles. Segments of *C. oxyacanthus* DNA are shorter to some extent in linkage groups O5 and O6, suggesting slightly better genome mixing, and linkage group O4B appears to show the greatest degree of genome mixing, as many areas of *C. oxyacanthus* DNA are found interspersed among *C. tinctorius* genomic regions rather than being found in large blocks. Levels of integration of the *C. oxyacanthus* genome into the *C. tinctorius* genome were low for linkage group O3, as indicated by the high proportion of *C. tinctorius* ('A') alleles. For all linkage groups, precise levels of recombination could not be calculated as recombination is masked by the BC1S3 generation such that recombinant DNA cannot be distinguished from parental DNA with this method. In summary, introgression of the *C. oxyacanthus* genome into the *C. tinctorius* genome was less than expected, as indicated in the scoring matrices by

large genomic blocks of a single species' DNA, which would have otherwise broken apart.

With the exceptions of markers ct619 (LGO4B), ct185 (LGO5), and ct381 (LGO6), there was no distortion of marker segregation on any of the linkage groups besides linkage group O3. All of the markers in linkage group O3 other than ct535 were decidedly distorted based on Chi-square analysis of expected genotypic ratios. Selection on this linkage group appears to be in favour of the *C. tinctorius* ('A') genotype. The most plausible explanation of why ct535 is an exception to this trend is that its genomic location has been misplaced on the map, given there is approximately a 9cM and 7cM gap between marker ct535 and the adjacent markers S30D05 and ct476, respectively. However, if subsequent mapping studies find that marker ct535 is indeed correctly positioned, then this result suggests there is a relative preference for the *C. oxyacanthus* ('B') genotype in comparison with surrounding markers where the *C. tinctorius* genotype is favoured. The segregation distortion of markers ct185 (LGO5) and ct381 (LGO6) is such that there are more *C. tinctorius* ('A') alleles than expected, suggestive of a selective advantage for the *C. tinctorius* genotype at these loci, while the segregation distortion of marker ct619 (LGO4B) is due to an increase in heterozygotes ('H') over the expected proportion of *C. oxyacanthus* homozygotes ('B').

For individuals with extreme phenotypes (i.e. the four plants with the earliest maturity) which were scanned for correlated patterns in marker segregation, there was no notable pattern of segregation consistent among the

individuals scanned in each phenotypic category for any of the linkage groups. Thus, none of the phenotypes examined showed clear correspondence with marker segregation over a certain genomic region.

	Plant																						
Marker	1.7.6	1.7.9	1.7.13	1.7.14	1.7.19	1.7.24	1.7.28	1.7.29	1.7.30	1.7.31	1.7.32	1.7.34	1.7.35	1.7.36	1.7.38	1.7.39	1.7.40	1.7.41	1.7.42	1.9.2	1.9.3	1.9.4	1.9.6	1.9.8
ct474	B	A	A	A	B	B	A	A	A	A	A	A	B	A	-	H	B	A	H	A	A	A	H	A
ct531	B	A	B	A	B	B	A	B	A	A	A	A	B	A	-	B	B	B	B	A	A	B	A	-
ct783	B	A	B	-	H	B	A	H	A	A	A	A	B	A	A	B	B	B	B	A	A	B	A	A
ct384	B	A	B	A	H	B	A	H	A	A	A	A	B	A	A	B	B	B	B	A	A	B	A	A
ct598	H	A	B	A	B	B	A	H	B	A	-	-	B	-	-	B	B	-	A	A	A	B	-	A
ct405	B	A	B	A	H	H	B	H	A	A	A	A	B	A	A	A	A	B	A	A	A	B	A	A
ct316	-	A	A	B	-	-	B	B	-	A	-	-	-	A	-	A	A	-	-	A	A	H	A	-
ct657	B	A	A	B	H	H	B	B	A	A	A	A	-	A	-	A	A	B	A	A	A	A	A	A

	Plant																							
Marker	1.9.12	1.9.15	1.9.16	1.9.17	1.9.18	2.9.1	2.9.2	2.9.4	2.9.7	2.9.8	2.9.10	2.9.11	2.9.15	2.9.16	2.9.19	2.9.21	2.9.26	2.9.30	2.9.36	2.9.37	2.9.38	2.9.40	2.9.45	2.9.47	X^2
ct474	A	B	A	A	-	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	-	A	A	A	1.05
ct531	A	H	A	A	B	A	B	-	A	A	B	A	A	A	A	A	A	B	A	H	B	A	-	A	5.43
ct783	A	H	A	A	B	A	B	A	A	A	B	A	A	A	A	A	A	B	A	H	B	A	A	A	2.41
ct384	A	H	A	A	B	A	B	A	A	A	B	A	A	A	A	A	A	B	A	H	B	A	A	A	2.09
ct598	-	B	A	A	B	-	B	-	A	A	A	A	A	B	A	A	A	B	A	B	-	A	A	A	4.98
ct405	B	A	A	A	B	H	B	A	A	A	A	A	A	B	A	B	A	A	B	B	A	A	A	A	1.28
ct316	-	A	-	A	-	-	B	-	A	A	A	-	A	A	A	-	B	-	B	B	-	A	-	A	0.49
ct657	H	A	A	B	A	A	-	A	A	A	A	A	A	A	A	A	A	A	-	-	A	A	A	A	1.58

Figure 3.11a: Introgression results for linkage group O1 of the BC1S3 plants. *C. oxyacanthus* alleles are highlighted in grey.

	Plant																							
Marker	1.7.6	1.7.9	1.7.13	1.7.14	1.7.19	1.7.24	1.7.28	1.7.29	1.7.30	1.7.31	1.7.32	1.7.34	1.7.35	1.7.36	1.7.38	1.7.39	1.7.40	1.7.41	1.7.42	1.9.2	1.9.3	1.9.4	1.9.6	1.9.8
ct549	A	-	A	A	A	A	A	-	-	A	-	-	A	A	A	A	-	-	-	-	A	A	A	A
ct495	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
ct605	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
ct410	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	B	A	A	A	A	A	A
ct297	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A
ct476	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	A	A	A	A	A
ct535	A	H	B	A	H	H	A	A	A	A	H	-	-	A	-	A	A	B	A	A	A	A	A	A
ct858	A	A	A	A	A	A	A	A	H	A	A	B	A	A	-	A	A	A	A	H	A	A	A	A
ct390	A	A	A	A	A	A	A	A	H	A	A	B	A	A	A	A	A	A	A	H	A	A	A	A

	Plant																								
Marker	1.9.12	1.9.15	1.9.16	1.9.17	1.9.18	2.9.1	2.9.2	2.9.4	2.9.7	2.9.8	2.9.10	2.9.11	2.9.15	2.9.16	2.9.19	2.9.21	2.9.26	2.9.30	2.9.36	2.9.37	2.9.38	2.9.40	2.9.45	2.9.47	X^2
ct549	A	A	A	B	-	-	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	A	A	11.38
ct495	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	16.36
ct605	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	16.36
ct410	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	-	A	A	A	A	A	A	A	A	13
ct297	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	14.67
ct476	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	11.04
ct535	A	B	B	B	B	A	A	B	A	A	B	A	A	A	A	-	B	A	A	A	A	-	A	A	0.69
ct858	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	B	A	B	A	-	-	A	A	6.48
ct390	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	B	A	B	-	A	A	A	A	7.13

Figure 3.11b: Introgression results for linkage group O3 of the BC1S3 plants. *C. oxyacanthus* alleles are highlighted in grey and markers with segregation distortion are bolded.

	Plant																							
Marker	1.7.6	1.7.9	1.7.13	1.7.14	1.7.19	1.7.24	1.7.28	1.7.29	1.7.30	1.7.31	1.7.32	1.7.34	1.7.35	1.7.36	1.7.38	1.7.39	1.7.40	1.7.41	1.7.42	1.9.2	1.9.3	1.9.4	1.9.6	1.9.8
cm022	A	A	A	B	A	B	A	A	A	A	A	A	A	A	-	B	A	A	H	B	A	H	A	A
ct408	A	A	A	B	-	-	-	-	A	A	-	-	-	A	-	H	A	-	-	B	A	H	A	A
ct588	A	A	A	H	A	B	A	A	A	A	-	A	A	A	B	H	A	-	H	B	A	H	A	A
ct520	A	A	-	A	A	A	B	A	A	B	A	A	A	A	-	A	A	A	A	A	B	A	A	B
ct619	A	A	A	B	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H
ct788	A	-	A	B	A	A	H	A	A	A	A	B	A	B	-	A	A	A	B	-	A	A	A	B
ct820	A	A	A	B	A	A	H	A	A	A	A	B	B	B	-	A	A	A	B	A	A	B	A	B

	Plant																								
Marker	1.9.12	1.9.15	1.9.16	1.9.17	1.9.18	2.9.1	2.9.2	2.9.4	2.9.7	2.9.8	2.9.10	2.9.11	2.9.15	2.9.16	2.9.19	2.9.21	2.9.26	2.9.30	2.9.36	2.9.37	2.9.38	2.9.40	2.9.45	2.9.47	X^2
cm022	-	H	-	A	A	A	A	A	A	-	-	B	-	A	A	A	-	-	-	A	A	A	A	A	3.1
ct408	A	-	-	A	-	A	-	-	A	-	-	-	H	A	A	A	A	-	A	A	-	A	-	A	3.95
ct588	A	B	H	A	A	A	A	A	A	A	A	B	B	A	A	-	A	A	A	A	A	A	A	A	3.29
ct520	B	A	B	A	-	A	A	A	A	B	B	A	A	A	A	B	B	A	B	A	A	A	A	A	3.03
ct619	B	H	A	A	A	A	H	A	H	H	A	A	A	A	A	A	H	B	B	A	A	A	A	A	9.54
ct788	B	B	A	A	A	A	H	A	H	H	A	A	H	B	A	A	H	B	A	A	A	A	A	A	3.85
ct820	A	B	A	A	A	A	B	A	H	A	A	A	A	B	A	A	H	-	A	B	A	A	A	H	0.59

Figure 3.11c: Introgression results for linkage group O4B of the BC1S3 plants. *C. oxyacanthus* alleles are highlighted in grey and markers with segregation distortion are bolded.

	Plant																							
Marker	1.7.6	1.7.9	1.7.13	1.7.14	1.7.19	1.7.24	1.7.28	1.7.29	1.7.30	1.7.31	1.7.32	1.7.34	1.7.35	1.7.36	1.7.38	1.7.39	1.7.40	1.7.41	1.7.42	1.9.2	1.9.3	1.9.4	1.9.6	1.9.8
ct032	A	A	A	A	A	B	B	A	A	H	A	A	A	A	A	A	A	B	B	A	B	A	B	A
ct233	A	A	-	A	A	B	B	A	A	A	A	A	A	A	A	A	A	B	B	A	B	-	B	A
ct185	A	A	A	A	A	A	B	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A
ct458	A	A	A	A	A	B	B	A	A	A	A	A	A	A	A	A	A	H	A	A	-	A	-	A
ct353	A	A	A	A	A	B	B	A	A	A	A	A	A	A	A	A	A	B	B	A	B	A	B	A
ct266	A	A	A	A	A	A	A	A	A	A	B	A	A	A	A	A	A	B	B	A	B	A	B	A
ct137	A	A	A	A	A	H	A	A	A	A	B	A	A	A	A	A	B	B	B	A	B	A	B	A
ct642	A	A	B	A	A	A	A	A	A	A	B	A	A	A	-	A	B	B	B	A	B	A	B	A
ct419	-	A	A	A	A	A	A	-	A	A	B	-	-	A	-	A	B	-	B	A	B	A	B	A

	Plant																								
Marker	1.9.12	1.9.15	1.9.16	1.9.17	1.9.18	2.9.1	2.9.2	2.9.4	2.9.7	2.9.8	2.9.10	2.9.11	2.9.15	2.9.16	2.9.19	2.9.21	2.9.26	2.9.30	2.9.36	2.9.37	2.9.38	2.9.40	2.9.45	2.9.47	X^2
ct032	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	-	A	A	A	A	A	A	A	A	3.93
ct233	A	A	A	A	A	A	A	A	A	A	A	-	B	H	A	A	B	B	H	B	A	B	A	H	0.2
ct185	A	A	A	A	A	H	A	A	A	A	A	H	A	A	A	A	H	A	A	A	A	A	A	H	11.56
ct458	A	A	A	A	A	A	A	A	A	A	-	B	B	A	A	A	B	A	A	B	A	A	A	H	2.41
ct353	A	A	A	A	A	-	A	A	A	A	A	B	B	H	A	A	B	B	H	B	A	B	A	H	0.38
ct266	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	A	B	B	A	A	3.38
ct137	A	A	A	A	A	-	A	A	A	A	A	-	A	A	A	A	B	A	B	H	B	B	A	A	0.29
ct642	-	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	B	A	B	H	B	B	A	A	1.32
ct419	A	A	-	A	-	A	-	-	A	A	A	-	A	-	A	A	B	-	B	H	-	B	-	A	0.64

Figure 3.11d: Introgression results for linkage group O5 of the BC1S3 plants. *C. oxyacanthus* alleles are highlighted in grey and markers with segregation distortion are bolded.

	Plant																							
Marker	1.7.6	1.7.9	1.7.13	1.7.14	1.7.19	1.7.24	1.7.28	1.7.29	1.7.30	1.7.31	1.7.32	1.7.34	1.7.35	1.7.36	1.7.38	1.7.39	1.7.40	1.7.41	1.7.42	1.9.2	1.9.3	1.9.4	1.9.6	1.9.8
ct139	A	A	A	A	A	B	A	A	A	A	A	A	A	H	A	A	A	A	A	B	A	B	A	A
ct351	A	A	A	A	A	B	A	A	A	A	A	A	A	H	A	A	A	H	A	B	A	B	B	A
ct044	A	A	A	A	A	B	A	A	A	A	A	A	A	H	-	A	A	B	A	A	A	B	B	A
ct381	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A	B	A	A	A	B	B	A
ct201	H	A	B	A	A	H	A	A	A	A	A	A	H	A	A	A	A	B	A	A	A	H	B	A
ct331	H	A	B	A	A	H	A	A	A	A	A	A	H	A	-	A	A	B	A	A	A	A	B	A
ct590	H	A	A	A	B	A	A	A	H	A	A	A	B	-	B	A	-	A	A	A	A	A	B	-

	Plant																								
Marker	1.9.12	1.9.15	1.9.16	1.9.17	1.9.18	2.9.1	2.9.2	2.9.4	2.9.7	2.9.8	2.9.10	2.9.11	2.9.15	2.9.16	2.9.19	2.9.21	2.9.26	2.9.30	2.9.36	2.9.37	2.9.38	2.9.40	2.9.45	2.9.47	X^2
ct139	A	A	A	A	A	A	B	A	A	B	A	A	A	A	A	A	A	A	A	A	B	A	A	A	4.49
ct351	A	A	A	A	A	A	B	A	A	B	A	A	A	A	A	A	H	A	A	A	B	A	A	A	1.52
ct044	A	A	A	A	A	A	B	A	A	H	A	A	H	A	A	A	H	A	A	A	B	A	-	A	2.34
ct381	A	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A	A	A	B	A	A	A	6.56
ct201	B	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A	A	B	B	A	A	A	0.73
ct331	B	A	A	A	A	A	B	A	A	A	A	A	A	A	A	A	A	A	A	B	B	H	A	A	1.58
ct590	A	A	A	A	A	-	B	-	A	A	A	A	B	A	A	A	A	A	A	A	-	B	-	A	0.77

Figure 3.11e: Introgression results for linkage group O6 of the BC1S3 plants. *C. oxyacanthus* alleles are highlighted in grey and markers with segregation distortion are bolded.

3.4. Discussion

The experimental work in this chapter has characterized the phenotypes of an inter-specific cross between commercial safflower and its wild relative, *C. oxyacanthus*. To my knowledge, this marks the first in-depth research on phenotypic diversity and genome introgression of a *C. tinctorius* x *C. oxyacanthus* hybrid population after two and three further generations of self-fertilization. Thus, this is the first time insight can be offered regarding the stability of previously noted hybrid vigour and novel phenotypes of a *C. tinctorius* x *C. oxyacanthus* cross (Deshpande, 1952; Sabzalian *et. al*., 2009).

The first area of phenotypic diversity noted was in the large variation in germination rates for BC1S2 plants grown in the field experiment. Overall, germination rates in the field were far lower than expected, likely due to the combination of dry weather in early spring and poor seed quality, given that the seeds sown were obtained from the previous year's field trial. That germination rates were higher in the BC1S3 generation of growth chamber experiments than in the field trial reflects the lack of stress the plants grown in the controlled conditions were under at that time. Only line 1.7.36 (field line 15) had an above average germination rate in both field and growth chamber experiments.

Field measurements of the stage of plant growth at 49 days showed variation both within and among lines. Much of this variation, however, is likely due to environmental factors because even *C. tinctorius* cv. Centennial plants, which are known to be homozygous, showed variation in their level of growth. While four lines were identified as having both an elongated plant stage and at

least a 40% germination rate, this cannot be attributed to hybrid vigour or transgressive segregation in these lines, as the same phenotypes were seen in *C. tinctorius* cv. S-317 parental plants.

Plant height, a known complex quantitative trait, showed substantial phenotypic diversity between field plant lines. However, certain lines were more uniform than others, which could be due to either heterozygosity or environmental factors, or simply the unequal sample sizes of each plant line. Regardless, the five shortest lines and the three tallest lines identified may be useful in breeding programs interested in improving traits that are correlated with height. Values for plant height, as recorded both in the field on day 78 for the BC1S2 population and in the growth chamber for BC1S3 plants at maturity, were comparable with the parental plants, so again there is no indication of trangressive segregation for plant height.

The effect of predation on the field trial was such that further phenotypic characterization was limited by the reduced population size and individual plant health. Through studying the effect of spinescence on predation, the results suggest that strong leaf spines may have the potential to defer predation by deer, and this finding should be further researched with appropriate quantitative measures. Earlier studies have demonstrated that even when all the leaves are preyed upon in mature safflower, yield is not decreased by more than 25% (Urie *et. al.*, 1968). This suggests that while herbivory is a concern as it was in the present field trial, especially for young plants with fewer leaves to depend on for photosynthesis, it is not overly concerning in mature safflower.

The observation that none of the plants in the field trial were early flowering with respect to the commercial Saffire cultivar of safflower is as expected, since *C. oxyacanthus* maturity is delayed due to its prolonged rosette stage and increased branching (Dajue and Mündel, 1996). Not surprisingly, late flowering of *C. oxyacanthus* also occurred for parental plants in the growth chamber experiments. This delay in plant growth would be an important consideration for plant breeders using *C. oxyacanthus* to widen the safflower gene pool.

For phenotypic characterization of the BC1S2 and BC1S3 populations in the growth chamber experiments, the distribution of branches along the main axis was one of the simplest traits to measure. As *C. oxyacanthus* is bushier, later flowering and maturing, and has more flower heads than the *C. tinctorius* parental plants, it is appropriate that its branches were distributed throughout the main plant axis, rather than just on the upper two-thirds. The number and distribution of branches was highly variable in the BC1S3 population. There were fewer plants in the BC1S2 generation that had the branching pattern that is characteristic of *C. oxyacanthus* than in the BC1S3 generation, which may be because plants were spaced very close together in that generation rather than a preference for the typical branching distribution of *C. tinctorius*.

There were an unusually low number of seeds produced by the parental plants in growth chamber experiments. This is probably because of an unexpected temperature drop to 15°C that occurred in the growth chamber for a period of time. Additionally, the low light intensity in the growth chamber

relative to the field is known to result in lower seed set in many species. As safflower is a warm weather crop suitable for arid and semi-arid conditions, low temperature and light intensity will impact both the time to flowering and maturation as well as the overall plant yield. Since the BC1S3 plant generation overlapped with the growth of the parental plants species, they may also have been subject to a temperature drop near the start of their growth cycle. This would explain why the days to flowering was much higher on average for the BC1S3 population than for the BC1S2 population of plants grown in the growth chamber. Nonetheless, the number of seeds produced in the BC1S3 generation was highly variable between plants.

In relation to other BC1S3 plants, five plants were significantly early maturing. Of all the measurements taken, it is most interesting from an agronomic perspective to see if any of the early maturing plants also showed an increase in phenotypic characteristics related to plant yield. Individual plant 1.9.8 had a high value for average individual seed weight in addition to a very early date of maturity. The three plants that produced the most seeds also matured earlier than the average BC1S3 plant, which is a highly desirable phenotypic combination that needs only to be improved for single seed weight.

The results of phenotypic correlations given for the BC1S3 plant population can offer insight into the potential trade-offs that selection for certain beneficial traits in crop improvement programs may have on other plant phenotypes. In both BC1S2 and BC1S3 generations, there was no correlation between plant height and days to first flowering. The lack of correlation between

the number of days to first flowering and days to maturity suggests that initial flowering time may not be a suitable indicator of earliness, an important consideration for plant breeders. Instead, the strong negative correlation between plant height and date of maturity indicates that safflower breeders interested in reducing the generation time may be able to do so by selecting for taller plants. The small positive correlation found between plant height and total seed number is inconsistent with earlier findings where they were not correlated (Smith, 1996). The correlation analysis also reveals pairs of traits with little correlation, showing phenotypes that can be selected for in safflower breeding with minimal impact or trade-off effect on another trait. As an example, the results of this experiment suggest that breeding for plants with a high single seed weight should not greatly affect the number of capitula on the plant. That the seed number and weight were not strongly correlated is also consistent with earlier findings from a study of the BC1S1 population developed from the same *C. tinctorius* and *C. oxyacanthus* cross (Bowles, 2010). However, that study also found a positive correlation between the total number of seeds and the number of capitula per plant which did not occur here.

The introgression analysis performed revealed less genome mixing than you would expect given that the two species readily cross. The number of individuals retaining *C. oxyacanthus* DNA varied among linkage groups, as did the amount of DNA retained and the level of genome mixing. There were four areas of selection or marker distortion found, all of which favoured the *C. tinctorius* genotype. In comparison to the introgression analysis performed

previously on the BC1S1 population (Bowles, 2010), there were some differences found in the present study. Firstly, in none of the linkage groups where segregation distortion was observed in Bowles' analysis were distorted markers grouped together, as seen in this study for linkage group O3. Furthermore, Bowles found an absence of *C. oxyacanthus* alleles in marker ct657, but with the larger population analyzed here, one individual was found to be heterozygous. In both studies, marker ct139 had no heterozygotes, as here all *C. oxyacanthus* genotypes were homozygous. Marker ct266 had one heterozygous individual in this study which was homozygous for the *C. tinctorius* genotype in Bowles' study, possibly because of weak amplification in the earlier generation. In general, there was less *C. oxyacanthus* DNA in the BC1S3 generation as some of the heterozygosity reported earlier had been lost. This is also representative of increased genomic stability, as occurs with the production of recombinant inbred lines. Additional comparison between the two studies is limited as there were only 13 individual plants used in common.

Based on the results of this introgression analysis, safflower breeding programs employing *C. oxyacanthus* may have difficulty incorporating traits of interest into the *C. tinctorius* genome. However, these results are promising for scientists using transgenic safflower for plant-made pharmaceuticals, and for studying the potential for introgression of transgenic safflower DNA into wild cultivars of safflower, which may be less likely than previously suspected. This idea is also supported in earlier work, where transgenes were selectively lost 21% of the time in certain crosses with safflower (Mayerhofer *et. al*., submitted).

In conclusion, the phenotypic characterization studies performed have identified several lines of interest. Field analysis revealed four vigorous lines (1.7.11, 1.7.21, 2.9.15, 2.9.35). In subsequent growth chamber experiments, two of these four lines were significantly early flowering in one or both plant generations (1.7.21 in BC1S2 generation, 2.9.15 in BC1S2 and BC1S3 generations), and one had a single seed weight that was markedly above average (2.9.15). Results of the growth chamber experiments have identified three plant lines that both matured earlier than average in the plant population and produced a significantly high number of seeds (1.7.13, 1.7.14, 1.7.38). In addition, one line was identified from growth chamber experiments as being remarkably early maturing and having a significantly high average single seed weight (1.9.8). These lines may provide highly valuable germplasm for use in numerous plant breeding experiments and can be further propagated to produce recombinant inbred lines.

3.5. References

Abel, G.H. and Driscoll, M.F. 1976. Sequential trait development and breeding for high yield in safflower. *Crop Sci.* **16**: 213-216.

Acquaah, G. 2007. Principles of plant genetics and breeding. Oxford: Blackwell Publishing Ltd. pp. 282-312.

Amini, F., Saeidi, G., and Arzani, A. 2008. Study of genetic diversity in safflower genotypes using agro-morphological traits and RAPD markers. *Euphytica* **163**(1): 21-30.

Ashri, A., and Efron, Y. 1964. Inheritance studies with fertile inter-specific hybrids of three *Carthamus* L. species. *Crop Sci* **4**(5): 510-514.

Ashri, A. and Knowles, P.F. 1960. Cytogenetics of safflower (*Carthamus* L.) species and their hybrids. *Agron. J.* **52**(1): 11-17.

Bergman, J.W. and Flynn, C. R. 2008. Evaluation of safflower and other oilseed crops grown in the United States Northern Plains Regions for biofuels/biobased products. *In* Proceedings of the 7th International Safflower Conference, Wagga Wagga, NSW, Australia, 3-6 November 2008.

Bowles, V.G. 2010. Relationships and introgression within *Carthamus* (Asteraceae), with an emphasis on safflower (*Carthamus tinctorius*). Masters Thesis, University of Alberta, Edmonton. pp. 66-114.

Chungru, B., Verma, N., Mohanty, A., Pradhan, A., and Shivanna, K.R. 1999. Production and characterization of interspecific hybrids between *Brassica maurorum* and crop brassicas. *Theor. Appl. Genet*. **98**(3): 608-613.

Dajue, L., and Mündel, H. 1996. Safflower. *Cathamus tinctorius* L. Promoting the conservation of underutilized and neglected crops. Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany and International Plant Genetic Resources Institute, Rome, Italy.

Deshpande, R. B. 1952. Wild safflower (*Carthamus oxyacantha* Bieb.) – A possible oilseed crop for the desert and arid regions. *Ind. J. Genet. Pl. Br.* **12**(1): 10-14.

Doebly, J.F., Gaut, B.S., and Smith, B.D. 2006. The molecular genetics of crop

domestication. *Cell* **127**(7): 1309-1321.

International Board of Plant Genetic Resources (1983). Safflower Descriptors. Retrieved from http://www2.bioversityinternational.org/publications/Web_version/237/ch2.htm#DESCRIPTOR%20LIST%20FOR%20SAFFLOWER. [Accessed Aug 2010].

Kavani, R.H., Shukla, P.T., and Madariya, R.B. 2000. Analysis of variability for seed yield and related characters in safflower (*Carthamus tinctorius* L.). *Madras Agric J.* **87**: 449-452.

Khidir, M.O. 1969. Evolution of the genetic system of safflower. *Genetica* **40**(1): 84-88.

Mayerhofer, M., Mayerhofer, R., Topinka, D., Christianson, J., and Good, A.G. (submitted). Introgression potential between cultivated safflower (*Carthamus inctorius*) and wild relatives within the *Carthamus* family.

Mayerhofer, R., Archibald, C., Bowles, V., and Good, A. 2010. Development of molecular markers and linkage maps for the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*. *Genome* **53**(4): 266-276.

Mündel, H., Blackshaw, R.E., Byers, J.R., Huang, H.C., Johnson, D.L., Keon, R., Kubik, J., McKenzie, R., Otto, B., Roth, B., and Stanford, K. 2004. Safflower Production on the Canadian Prairies: revisited in 2004. Agriculture and Agri-Food Canada, Lethbridge Research Centre, Lethbridge, AB, 36 pp.

Poehlman, J.M. and Sleper, D.A. 1995. Breeding field crops. Ames: Iowa State

University Press.

Rashid, A.,Rakow, G., and Downey, R.K. 1994. Development of yellow seeded *Brassica napus* through interspecific crosses. *Plant Breeding* **112**(2): 127-134.

Ravikumar, R.L., Roopa, V.K., Soregaon, C.D., and Satish, D. 2008. Molecular diversity in *Carthamus* species and development of inter-specific mapping population for the first molecular map in safflower. *In* 7th International Safflower Conference, Wagga Wagga, Australia, 3-6 November 2008.

Roy, NN. 1978. A study of disease variation in the populations of an interspecific cross of *Brassica juncea* L x *B. napus* L. *Euphytica* **27**(1): 145-149.

Sabzalian, M., Saeidi, G., and Mirlohi, A. 2008. Oil content and fatty acid composition in seeds of three safflower species. *J. Am. Oil. Chem. Soc.* **85**(8): 717-721.

Sabzalian, M. R., Mirlohi, A., Saeidi, G., and Rabbani, M. 2009. Genetic variation among populations of wild safflower, *Carthamus oxyacanthus* analyzed by agro-morphological traits and ISSR markers. *Genet. Resour. Crop. Evol.* **56**(8): 1057-1064.

Sabzalian, M.R., Saeidi, G., Mirlohi, A., and Hatami, B. 2010. Wild safflower species (*Carthamus oxyacanthus*): A potential source of resistance to the safflower fly (*Acanthiophilus helianthi*). *Crop. Prot.* **29**(6): 550-555.

Sehgal, D., Rajpal, V.R., and Raina, S.N 2008. Chloroplast DNA diversity

reveals the contribution of two wild species to the origin and evolution of diploid safflower (*Carthamus tinctorius* L.) *Genome* **51**(8): 638-643.

Smith, J.R. 1996. Safflower. AOCS Press, USA, pp. 40-45.

Thurling, N. 2003. Application of the ideotype concept in breeding for higher yield in the oilseed brassicas. *Field Crops Res.* **26**(2): 201-219.

Weiss, E.A. 2000. Oilseed crops, 2nd ed. Oxford: Blackwell Science Ltd.

Urie, A.L., Leininger, L.N., and Zimmer, D.E. 1968. Effects of degree and time of defoliation on yield and related attributes of safflower. *Crop. Sci.* **8**: 747-750.

4. General Conclusions

4.1. Research Contributions and Future Study

The goals of this thesis project were to improve upon the genetic resources available for *Carthamus tinctorius*, to screen inter-specific (*C. tinctorius* x *C. oxyacanthus*) backcross populations for phenotypes of agricultural and horticultural interest, and to gain understanding of the introgression of *C. oxyacanthus'* DNA into the *C. tinctorius* genome through microsatellite marker analysis. The contributions of this research toward developing genetic resources for *C. tinctorius* include detailed linkage analysis and the development of a linkage map from an inter-specific (*C. tinctorius* x *C. oxyacanthus*) backcross population. Microsatellite marker polymorphism was identified from parental screening of over 1100 SSR markers. The low levels of polymorphism found in microsatellite marker screening and the low proportion of mappable markers were obstacles in creating a saturated inter-specific linkage map. Methods to improve genomic coverage by molecular markers would do best to draw upon hitherto unused marker systems which can be developed in target genomic regions where marker saturation is low.

In the short-term, the best approach to improving the marker saturation of the present composite map may be to exploit anonymous marker systems such as AFLPs which were not used in the present study. While anonymous markers are not optimal because of their low transferability among laboratories, they may prove useful in filling in gaps in marker saturation and linking fragmented linkage

groups. In addition, further analysis with RFLPs could improve marker density given the high levels of polymorphism they exhibited in this study. While generating more polymorphic mapping populations would aid the mapping process in the long-run, this task is complicated by two matters. The first consideration is that generating more polymorphic mapping populations may require the use of more divergent plant species, which are less likely to produce fertile F1 hybrid plants. The second matter is that, with the exception of elite cultivars, pedigrees of most safflower accessions are not readily available, so new mapping parents need to be selected based on morphological differences and trial and error. An alternative long-term strategy to improving the saturation of the linkage maps would be to employ the use of powerful single nucleotide polymorphism (SNP) marker systems, though SNP detection and marker development will require a substantial input of resources comparable to the cost of SSR markers.

As it is, the development and publication of the composite linkage map provides a framework for further comparative mapping and studies of genome evolution of the *Carthamus* species (Mayerhofer *et. al.*, 2010). It also supplies a large number of mapped markers able to be used in candidate gene discovery and marker assisted selection. The role of translocation and inversion events in speciation or speciation reinforcement has been well documented (Lagercrantz, 1998; Navarro and Barton, 2003; Noor *et. al.*, 2001; Rieseberg, 2001; Yogeeswaren *et. al.*, 2005). Thus, subsequent research on the comparative genomics of several *Carthamus* species, including *C. tinctorius* and *C.*

oxyacanthus, may shed light on the role that the translocation event found in this study played in their evolutionary divergence. Further examination of the genomic regions found to show marker clustering may also provide insight into the mechanism of evolutionary divergence between *C. tinctorius* and *C. oxyacanthus*, as sequence divergence has been known to repress recombination in many plant species, including *Arabidopsis* and maize (Opperman *et. al*., 2004; Schnable *et. al.*, 1998). While some of the marker clustering found on the inter-specific map in the present study may be due to the smaller mapping population size used in creating the inter-specific map in relation to that of the intra-specific map, software analysis of the intra-specific F2 population of *C. tinctorius* using a reduced population size did not result in extensive marker clustering. This supports the hypothesis that marker clustering is a consequence of low homology and suppressed recombination in certain genomic regions. These results are similar to the findings of Truco *et. al*. (2007), where linkage mapping of a backcross population derived from the inter-specific cross between the wild and cultivated lettuce species *Lactuca saligna* and *L. sativa* revealed areas of suppressed recombination which were attributed to genome divergence.

Several important findings came from the phenotypic characterization performed in both field and growth chamber experiments on two inter-specific backcross generations (BC1S2 and BC1S3) derived from the original cross between *C. tinctorius* and *C. oxyacanthus*. These include four lines which showed vigorous plant growth in the field trial, and several lines which were relatively high-yielding and early-maturing in growth chamber analyses. These

lines may be highly valuable for use in numerous plant breeding experiments if further study shows firstly that the observed phenotypes are consistent in subsequent generations, and secondly that these beneficial traits are not closely linked with detrimental secondary characters. Supplementary plant breeding studies using these lines may wish to attempt the introgression of specific characters typical of *C. oxyacanthus,* such as increased branching or number of seeds, into the *C. tinctorius* genome through the production of recombinant inbred lines from a segregating backcross population.

In the future, field trials with *C. tinctorius* in southern Alberta or the Canadian Prairie regions will have to ensure the border fence is sufficiently tall to discourage herbivory by deer and pronghorn antelope. The use of a smaller plot area should help with this, as it did in a safflower field trial in Warner, Alberta in 2007 (Bowles, 2010). The traits studied which showed the most variation overall were plant height, seed numbers or yield, seed weight, and maturation time. These are all important breeding considerations because of their impact on production efficiency and indication of plant health (Poehlman and Sleper, 1995; Weiss, 2000). Also, since they were polymorphic in the inter-specific population studied, they represent ideal traits for which to begin QTL mapping. The addition of QTLs to the inter-specific map produced in this study would be best done after further marker saturation of the linkage map, and it will require a sufficient population size to detect low levels of recombination. Hamdan *et. al.* (2008) found monogenic inheritance in safflower for oleic acid content. As this trait is of increasing interest because its importance in biodiesel (Bergman and Flynn,

2008), it should also be mapped. This can be done by first creating a mapping population from the hybridization of safflower cultivars with highly divergent oil profiles (i.e. high and low oleic acid producing cultivars), genotyping the mapping population at marker loci, and comparing their fatty acid profiles. Other traits to consider mapping include wilted flower colour, germination rates, number of capitula, and number of branches.

The results of the preliminary introgression analysis indicate that while the introgression of *C. oxyacanthus* DNA into the *C. tinctorius* genome is relatively stable, breeding strategies that incorporate this wild safflower species may find the genomes do not recombine well in certain chromosomal regions. Identifying the four marker loci which showed segregation distortion or selection in the introgression analysis was important because distortion can hinder gene flow between species and reduce the introgression of key agricultural alleles into the target plant species' genome (Truco *et. al*., 2007). It is recommended that additional introgression analysis of the remaining linkage groups be performed once they have been further saturated with molecular markers. Safflower breeding efforts that use *C. oxyacanthus* to increase genetic diversity will be required to focus on balancing the effects of the additional characteristic traits of *C. oxyacanthus* that may improve crop yield, including increased branching and number of capitula, with those that delay flowering and maturity.

In summary, the contributions of this thesis have fulfilled the initial project objectives, with the provision of a *C. tinctorius* linkage map, the discovery of several lines showing phenotypes of agronomic interest, and preliminary

genome introgression analysis of a population derived from a *C. tinctorius* x *C. oxyacanthus* cross. The findings provide important information for further development of safflower as an oilseed crop, and the genetic materials necessary to begin QTL mapping, marker assisted selection, and additional genome introgression studies between *C. tinctorius* and other *Carthamus* species.

4.2. References

Bergman, J.W. and Flynn, C. R. 2008. Evaluation of safflower and other oilseed crops grown in the United States Northern Plains Regions for biofuels/biobased products. *In* Proceedings of the 7th International Safflower Conference, Wagga Wagga, NSW, Australia, 3-6 November 2008.

Bowles, V.G. 2010. Relationships and introgression within *Carthamus* (Asteraceae), with an emphasis on safflower (*Carthamus tinctorius*). Masters Thesis, University of Alberta, Edmonton. pp. 66-114.

Hamden, Y.A.S., Pérez-Vich, B., Fernández-Martínez, J.M., and Velasco, L. 2008. Inheritance of very high oleic acid content and its relationship with several morphological and physiological traits. *In* Proceedings of the 7th International Safflower Conference, Wagga Wagga, NSW, Australia, 3-6 November 2008.

Lagercrantz, Ulf. 1998. Comparative mapping between *Arabidopsis thaliana* and

Brassica nigra indicates that Brassica genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150**: 1217-1228.

Mayerhofer, R., Archibald, C., Bowles, V., and Good, A. 2010. Development of molecular markers and linkage maps for the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*. *Genome* **53**(4): 266-276.

Navarro, A. and Barton, N.H. 2003. Chromosomal speciation and molecular divergence: Accelerated evolution in rearranged chromosomes. *Science* **300**: 321-324.

Noor, M.A.K., Grams, K.L., Bertucci, L.A., and Reiland, J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci.USA* **98**: 12084-12088.

Opperman, R., Emmanuel, E., and Levy, A.A. 2004. The effect of sequence divergence on recombination between direct repeats in Arabidopsis. *Genetics* **168**: 2207-2215.

Poehlman, J.M. and Sleper, D.A. 1995. Breeding field crops. Ames: Iowa State University Press.

Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**(7): 351-358.

Schnable, P.S., Hsia, A.P., and Nikolau, B.J. 1998. Genetic recombination in plants. *Curr. Opin. Plant Biol.* **1**: 123-129.

Truco, M.J., Antonise, R., Lavelle, D., Ochoa, O., Kozik, A., Witsenboer, H.,

Fort, S.B., Jeuken, M.J.W., Kesseli, R.V., Lindhout, P., Michelmore, R.W., and Peleman, J. 2007. A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). *Theor. Appl. Genet.* **115**: 735-746.

Weiss, E.A. 2000. Oilseed crops, 2nd ed. Oxford: Blackwell Science Ltd.

Yogeeswaren, K., Frary, A., York, T.L., Amenta, A., Lesser, A.H., Nasrallah, J.B., Tanksley, S.D., and Nasrallah, M.E. 2005. Comparative analysis of *Arabidopsis* spp.: Inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Res.* **15**: 505-515.

5. Appendices

5.1. Appendix A: Results of the safflower field trial in Warner, Alberta, 2009 on the BC1S2 population.

					Leaf Spine Measurements		Leaf Measurements			
BC1 ID; Line[1]	Field ID[2]	True Leaves[3]	Plant Stage[4]	Predated[5]	Spine Strength[6]	# Spines[7]	Leaf Shape[8]	Leaf Margin[9]	Height[10]	Flowering[11]
2.9.12; 1	Replicate A: 25% germinated by 35 days, 80% survived to 49 days.									
	1.1.1	3	3E, 1R	N	3	11	1	1	61	M
	1.1.2	3		N	3	12	1	1	72	M
	1.1.3	3		N	3	14	1	1	73	M
	1.1.4	4		N	3	13	3	2	56	M
	Replicate B: 20% germinated by 35 days, 100% survival to 49 days.									
	8.1.1	2	M	Y						L
	8.1.2	4		Y						L
	8.1.3	3		Y	1		3	3		L
	8.1.4	2		Y						L
	Mean	**3**				**12.5**			**65.5**	
	SD	**0.756**				**1.291**			**8.347**	
1.7.11; 2	Replicate A: 25% germinated by 35 days, 60% survived to 49 days.									
	1.2.1	3	R	Y						L
	1.2.2	3		Y						L
	1.2.3	1		Y						L
	1.2.4	3		Y						L
	Replicate B: 55% germinated by 35 days, 100% survived to 49 days after thinning to 10 seedlings.									

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	8.4.2	4		N	2	14	1	2	80	M
	8.4.3	4		N	2	8	3	3	74	M
	8.4.4	4		N	2	14	3	3	77	M
	8.4.5	4		Y						M
	8.4.6	4		Y						L
	8.4.7	4		N	2	16	3	3	81	M
	8.4.8	4		Y						L
	8.4.9	4		N	3	13	1	2	82	M
	8.4.10	4		N	3	10	1	1	83	M
	Mean	**3.571**				**12.143**			**79.571**	
	SD	**0.852**				**2.854**			**3.101**	
1.9.10; 3	Replicate A: 30% germinated by 35 days, 67% survived to 49 days.									
	1.3.1	1		N	2	17	1	2	62	
	1.3.2	3		Y						
	1.3.3	2	R	D						
	1.3.4	1		D						
	1.3.5	3		N	2	15	1	1	79	M
	Mean	**2**				**16**			**70.5**	
	SD	**1**				**1.414**			**12.021**	
2.9.4; 5	Replicate A: 20% germinated by 35 days, 75% survived to 49 days.									
	1.4.1	0		Y						
	1.4.2	2	R	Y						
	1.4.3	4		Y						

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	Replicate B: 35% germinated by 35 days, 86% survived to 49 days.									
	8.3.1	7		D						
	8.3.2	1		N	3	16	2	2	71	ML
	8.3.3	1	M	Y						
	8.3.4	0		Y						
	8.3.5	5		Y						
	8.3.6	2		y						
	Mean	**2.667**								
	SD	**2.271**								
2.9.55; 6	Replicate A: 5% germinated by 35 days, 100% survived to 49 days.									
	1.5.1	3	R	N	2	14	3	1	61	ML
	Replicate B: 50% germinated by 35 days, 100% survived to 49 days.									
	8.5.1	4		Y						
	8.5.2	3		Y						
	8.5.3	0		Y						
	8.5.4	5		Y						
	8.5.5	4	R	Y						
	8.5.6	4		Y						L
	8.5.7	4		Y						
	8.5.8	4		Y	2		4	3		
	8.5.9	5		Y						
	8.5.10	4		Y						
	Mean	**3.636**								
	SD	**1.362**								

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
1.7.35; 7	Replicate A: 15% germindated by 35 days, 100% survived to 49 days.									
	1.6.1	2		Y						
	1.6.2	3	1E, 2R	N	3	15	3	2	61	ML
	1.6.3	4		N	3	12	3	2	77	
	Replicate B: 5% germinated by 35 days, 100% survived to 49 days.									
	8.2.1	5	E	Y						L
	Mean	**3.5**				**13.5**			**69**	
	SD	**1.291**				**2.121**			**11.314**	
1.7.16; 8	Replicate A: 5% germinated by 35 days, 100% survived to 49 days.									
	1.7.1	4	E	Y						
	Replicate B: 20% germinated by 35 days, 100% survived to 49 days.									
	8.7.1	4		Y						
	8.7.2	5	M	Y						
	8.7.3	4		Y						
	8.7.4	2		Y						
	Mean	**3.8**								
	SD	**1.095**								
S317-7; 9	Replicate A: 40% germinated by 35 days, 63% survived to 49 days.									
	2.1.1	4		Y	2	16	4	2	74	ML
	2.1.2	4		Y	2	10	4	2	63	ML
	2.1.3	4	M	Y						
	2.1.4	1		D						
	2.1.5	1		D						

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	2.1.7	3		Y						
	2.1.8	4		Y						
	Replicate B: 25% germinated by 35 days, 80% survived to 49 days.									
	6.5.1	4		Y		10			79	M
	6.5.2	3	M	N	2	8	3	2	70	M
	6.5.3	3		Y						
	6.5.4	4		N	2	12	3	2	73	M
	Mean	**2.917**				**11.2**			**71.8**	
	SD	**1.443**				**3.033**			**5.891**	
1.7.32; 10	Replicate A: 65% germinated by 35 days, 100% survived to 49 days after thinning to 10 seedlings.									
	2.2.1	4		Y						
	2.2.2	3		Y						L
	2.2.3	4		Y						L
	2.2.4	4		Y						L
	2.2.5	3	R	Y	3	15	1	1	65	M
	2.2.6	3		Y						L
	2.2.7	4		Y						L
	2.2.8	4		Y						L
	2.2.9	4		Y						
	2.2.10	4		Y						
	Replicate B: 25% germinated by 35 days, 100% survived to 49 days.									
	6.2.1	4		Y		13			60	ML
	6.2.2	3	R	Y						

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	6.2.4	3		Y						
	6.2.5	4		Y						
	Mean	**3.6**				**14**			**62.5**	
	SD	**0.507**				**1.414**			**3.536**	
1.7.26; 11	Replicate A: 35% germinated by 35 days, 86% survived to 49 days.									
	2.3.1	4		Y	1	14	3	3	57	ML
	2.3.2	3		Y						ML
	2.3.3	4		Y		15			66	ML
	2.3.4	3	M	Y	1	13	1	2	65	ML
	2.3.5	3		Y	1	14	1	2	71	ML
	2.3.6	3		Y	1	14	1	2	68	ML
	2.3.7	2		D						
	Replicate B: 35% germinated by 35 days, 86% survived to 49 days.									
	6.4.1	5		Y		10			72	
	6.4.2	1		Y						
	6.4.3	2		Y						
	6.4.4	1	E	Y						
	6.4.5	1		D						
	6.4.6	4		N	2	12	1	2	76	M
	6.4.7	2		Y						
	Mean	**2.714**				**13.143**			**67.857**	
	SD	**1.267**				**1.676**			**6.094**	
2.9.49; 12	Replicate A: 25% germinated by 35 days, 60% survived to 49 days.									

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	2.4.2	4		Y						
	2.4.3	1		Y						
	Replicate B: 5% germinated by 35 days, 100% survived to 49 days.									
	6.6.1	1	R	N	3	13	3	3	54	L
	Mean	**1.5**								
	SD	**1.732**								
2.9.16; 13	Replicate A: 35% germinated by 35 days, 100% survived to 49 days.									
	2.5.1	4		Y						
	2.5.2	0		Y						
	2.5.3	1	R	Y						
	2.5.4	0		Y						
	2.5.5	4		Y	2	11	3	2	56	
	2.5.6	1		Y						
	Replicate B: 35% germinated by 35 days, 86% survived to 49 days.									
	6.1.1	3		D						
	6.1.2	2		N	3	17	1	2	68	ML
	6.1.3	3		Y						
	6.1.4	2	E	Y						
	6.1.5	4		Y						
	6.1.6	4		N	3	11	1	1	62	ML
	6.1.7	3		Y		13			65	ML
	Mean	**2.385**				**13**			**62.75**	
	SD	**1.502**				**2.828**			**5.123**	

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
1.7.37; 14	Replicate A: 10% germinated by 35 days, 100% survived to 49 days.									
	2.6.1	4	R	N	3	23	1	2	64	M
	2.6.2	6		D						
	Replicate B: 40% germinated by 35 days, 100% survived to 49 days.									
	6.3.1	3		Y						M
	6.3.2	4		N	3	10	1	2	63	M
	6.3.3	4		N	2	16	1	2	68	M
	6.3.4	4	M	N	3	16	1	2	67	M
	6.3.5	3		N	2	11	1	2	66	ML
	6.3.6	3		N	2	14	1	2	61	M
	6.3.7	3		N	2	13	1	2	62	M
	6.3.8	4		N	2	12	1	2	72	M
	Mean	**3.8**				**14.375**			**65.375**	
	SD	**0.919**				**4.104**			**3.623**	
1.7.36; 15	Replicate A: 45% germinated by 35 days, 100% survived to 49 days.									
	2.7.1	3		Y						
	2.7.2	4		Y						
	2.7.3	4		Y						
	2.7.4	4		Y						
	2.7.5	2	R	Y						
	2.7.6	9		Y						
	2.7.7	5		Y						

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	2.7.9	4		Y						
	Replicate B: 40% germinated by 35 days, 75% survived to 49 days.									
	6.7.1	9		N	2	13	4	3	71	M
	6.7.2	3		Y	2	12	3	2	71	ML
	6.7.3	5	E	N	2	16	3	2	76	M
	6.7.4	6		Y	1	10	3	2	76	M
	6.7.5	6		Y						L
	6.7.6	5		N	2	12	3	2	66	M
	Mean	**4.867**				**12.6**			**72**	
	SD	**1.995**				**2.191**			**4.183**	
2.9.33; 16	Replicate A: 55% germinated by 35 days, 90% survived to 49 days after thinning to 10 seedlings.									
	3.1.1	4		N	3	14	3	2	67	M
	3.1.2	3		Y	3	15	3	2	69	ML
	3.1.3	3		N	2	11	1	2	71	M
	3.1.4	4		N	2	16	3	2	70	ML
	3.1.5	2	E	N	2	16	1	2	69	ML
	3.1.6	4		N	2	13	3	2	70	ML
	3.1.7	4		N	2	19	1	2	69	M
	3.1.8	2		Y						
	3.1.9	2		Y						
	3.1.10	4		D						
	Replicate B: 10% germinated by 35 days, 100% survived to 49 days.									
	10.7.1	6	R	Y						

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	Mean	**3.417**				**14.857**			**69.286**	
	SD	**1.165**				**2.545**			**1.254**	
2.9.11; 17	Replicate A: 70% germinated by 35 days, 100% survived to 49 days after thinning to 10 seedlings.									
	3.2.1	4		Y						
	3.2.2	3		Y						
	3.2.3	4		N	3	13	3	2	78	M
	3.2.4	4		N	2	17	3	2	78	ML
	3.2.5	3	M	N	2	10	3	2	77	ML
	3.2.6	3		N	2	11	1	2	72	ML
	3.2.7	2		Y						
	3.2.8	4		Y	2	13	3	2	71	ML
	3.2.9	2		Y						
	3.2.10	4		Y						
	Replicate B: 35% germinated by 35 days, 86% survived to 49 days.									
	10.2.1	5		Y						
	10.2.2	5		Y						
	10.2.3	4		Y						
	10.2.4	5	5E, 2R	Y						
	10.2.5	4		Y						
	10.2.6	4		Y						
	10.2.7	4		Y						L
	Mean	**3.765**				**12.8**			**75.2**	
	SD	**0.903**				**2.683**			**3.421**	

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
2.9.2; 18	Replicate A: 15% germinated by 35 days, 67% survived to 49 days.									
	3.3.1	2		Y						
	3.3.2	1	R	Y						
	3.3.3	0		Y						
	Mean	**1**								
	SD	**1**								
2.9.54; 19	Replicate A: 55% germinated by 35 days, 100% survived to 49 days after thinning to 10 seedlings.									
	3.4.1	3		Y	2	11	3	2	61	ML
	3.4.2	3		Y						
	3.4.3	4		Y						
	3.4.4	4		N	2	8	3	3	55	ML
	3.4.5	4	R	N	2	10	3	2	58	ML
	3.4.6	4		N	2	10	3	2	58	ML
	3.4.7	4		N	2	6	3	2	56	ML
	3.4.8	3		N	2	9	3	3	55	ML
	3.4.9	3		Y						
	3.4.10	3		Y						
	Replicate B: 35% germinated by 35 days, 100% survived to 49 days.									
	10.6.1	3		Y						ML
	10.6.2	5		N	2	10	3	3	80	ML
	10.6.3	4	M	N	1	14	3	2	70	ML
	10.6.4	4		Y						
	10.6.5	4		N	3	9	3	3	74	ML

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	10.6.7	5		Y						
	Mean	**3.765**				**9.667**			**63.0**	
	SD	**0.664**				**2.179**			**9.287**	
1.7.8; 20	Replicate A: 5% germinated by 35 days, 100% survived to 49 days.									
	3.5.1	3	R	Y						
	Replicate B: 65% germinated by 35 days, 100% survived to 49 days after thinning to 10 seedlings.									
	10.3.1	5		Y						
	10.3.2	5		N	3	13	1	3	87	M
	10.3.3	4		Y	2	8	3	2	77	ML
	10.3.4	4		Y						
	10.3.5	5	R	Y	2		1	1		
	10.3.6	4		N	2	12	3	2	80	M
	10.3.7	5		N	2	9	3	2	81	M
	10.3.8	5		N	2	12	4	2	76	M
	10.3.9	5		Y	2		3	3		
	10.3.10	6		Y						
	Mean	**4.636**				**10.8**			**80.2**	
	SD	**0.809**				**2.168**			**4.324**	
2.9.35; 21	Replicate A: 55% germinated by 35 days, 100% survived to 49 days after thinning to 10 seedlings.									
	3.6.1	3		N	3	17	1	2	73	M
	3.6.2	4		N	3	17	1	2	76	ML
	3.6.3	4	M	N	3	21	3	2	67	ML
	3.6.4	3		N	3	13	3	3	68	ML

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	3.6.6	3		N	3	15	3	2	73	ML
	3.6.7	3		N	3	13	3	3	65	L
	3.6.8	2		N	2	18	3	3	57	M
	3.6.9	3		N	2	13	3	3	69	M
	3.6.10	4		N	3	15	1	2	70	M
	Replicate B: 35% germinated by 35 days, 86% survived to 49 days.									
	10.1.1	5		N	3	17	1	2	77	M
	10.1.2	5		N	2	13	3	2	81	ML
	10.1.3	5		N	3	8	2	2	81	M
	10.1.4	6	E	N	3	13	1	3	72	ML
	10.1.5	5		N	3	14	1	3	74	ML
	10.1.6	1		Y		9			70	ML
	10.1.7	5		N	3	12	1	2	73	ML
	Mean	**3.824**				**14.176**			**71.529**	
	SD	**1.286**				**3.206**			**5.821**	
S317-9; 22	Replicate A: 85% germinated by 35 days, 100% survived after thinning to 10 seedlings.									
	3.7.1	4		Y	2	16	3	3	73	ML
	3.7.2	4		Y						ML
	3.7.3	5		N	1	12	3	2	80	ML
	3.7.4	4		Y						L
	3.7.5	4	E	Y						
	3.7.6	5		Y		11			76	ML
	3.7.7	4		Y						
	3.7.8	5		N	1	10	3	3	80	ML

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	3.7.9	4		Y						L
	3.7.10	4		Y						L
	Replicate B: 25% germinated by 35 days, 80% survived to 49 days.									
	10.5.1	4		Y	2	10	3	2	74	ML
	10.5.2	4	E	Y						
	10.5.3	4		N	2	10	3	3	82	M
	10.5.4	4		Y						
	Mean	**4.214**				**11.5**			**77.5**	
	SD	**0.426**				**2.345**			**3.674**	
1.9.16; 23	Replicate A: 30% germinated by 35 days, 100% survived to 49 days.									
	4.1.1	1		Y						L
	4.1.2	2		Y						
	4.1.3	3	R	Y						L
	4.1.4	0		Y						L
	4.1.5	3		Y						L
	4.1.6	3		Y						L
	Replicate B: 15% germinated by 35 days, 100% survived to 49 days.									
	9.1.1	6		Y						L
	9.1.2	5	M	Y						L
	9.1.3	5		Y						L
	Mean	**3.111**								
	SD	**1.965**								
2.9.47; 24	Replicate A: 30% germinated by 35 days, 50% survived to 49 days.									

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	4.2.2	6		D						L
	4.2.3	4		Y						L
	4.2.4	4		Y						
	4.2.5	4		Y						L
	Replicate B: 30% germinated by 35 days, 83% survived to 49 days.									
	9.3.1	5		N	2	14	3	2	65	M
	9.3.2	4		N	1	12	3	2	64	ML
	9.3.3	4	M	N	1	17	3	3	64	ML
	9.3.4	3		D						
	9.3.5	2		Y						
	9.3.6	4		Y						
	Mean	**3.909**				**14.333**			**64.333**	
	SD	**1.044**				**2.517**			**0.577**	
2.9.15; 25	Replicate A: 25% germinated by 35 days, 100% survived to 49 days.									
	4.3.1	0		Y						
	4.3.2	0		Y						
	4.3.3	1	R	Y						
	4.3.4	3		Y						
	4.3.5	3		Y						
	Replicate B: 60% germinated by 35 days, 100% survived to 49 days after thinning to 10 seedlings.									
	9.4.1	7		N	3	12	4	3	71	M
	9.4.2	5		N	1	7	4	2	77	M
	9.4.3	5	E	N	1	9	4	2	77	M

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	9.4.5	7		N	2	12	4	3	70	M
	9.4.6	5		N	2	13	4	3	67	M
	9.4.7	6		Y						M
	9.4.8	5		N	1	10	3	3	72	M
	9.4.9	5		N	1	13	3	2	66	M
	9.4.10	6		N	2	5	3	3	63	M
	Mean	**4.4**				**10.333**			**71.333**	
	SD	**2.501**				**2.828**			**5.500**	
1.7.39; 26	Replicate A: 15% germinated by 35 days, 67% survived to 49 days.									
	4.4.1	2	R	Y						
	4.4.2	3		Y	2	11	3	2	73	ML
	Replicate B: 10% germinated by 35 days, 100% survived to 49 days.									
	9.7.1	2	M	N	1	12	2	2	81	ML
	9.7.2	4		N	2	13	3	2	71	ML
	Mean	**2.75**				**12.000**			**75.000**	
	SD	**0.957**				**1.000**			**5.292**	
2.9.44; 27	Replicate A: 35% germinated by 35 days, 100% survived to 49 days.									
	4.5.1	4		N	2	6	3	2	59	ML
	4.5.2	2		Y		9			52	ML
	4.5.3	3		N	2	12	4	3	67	ML
	4.5.4	3	R	N	2	12	4	3	69	ML
	4.5.5	3		N	2	13	4	3	63	ML
	4.5.6	5		N	2	11	4	3	54	ML

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	Replicate B: 15% germinated by 35 days, 100% survived to 49 days.									
	9.2.1	3		Y						L
	9.2.2	5	M	Y						L
	9.2.3	5		Y	1	14	3	2	68	ML
	Mean	**3.5**				**11**			**61.714**	
	SD	**1.179**				**2.708**			**6.873**	
1.9.5; 28	Replicate A: 25% germinated by 35 days, 100% survived to 49 days.									
	4.6.1	1		Y						
	4.6.2	3		N	3	11	2	3	80	M
	4.6.3	4	4E, 1R	N	3	15	3	3	77	M
	4.6.4	3		N	1	8	2	3	89	M
	4.6.5	4		N	3	10	3	3	74	M
	Replicate B: 20% germinated by 35 days, 100% survived to 49 days.									
	9.5.1	5		N	2	13	2	3	84	M
	9.5.2	5	E	N	2	9	1	2	80	M
	9.5.3	5		N	2	16	3	2	78	M
	9.5.4	6		Y						
	Mean	**4**				**11.714**			**80.286**	
	SD	**1.5**				**3.039**			**4.923**	
1.7.12; 29	Replicate A: 30% germinated by 35 days, 100% survived to 49 days.									
	4.7.1	2		Y						
	4.7.2	3		Y						
	4.7.3	5	M	Y						ML
	4.7.4	4		N	1	14	3	3	70	ML

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	4.7.5	4		Y	1		3	3		L
	4.7.6	0		Y						
	Replicate B: 25% germinated by 35 days, 100% survived to 49 days.									
	9.6.1	8		N	2	11	3	2	73	ML
	9.6.2	2		Y						
	9.6.3	4	E	Y						
	9.6.4	5		Y	1		4	2		ML
	9.6.5	4		N	1	10	4	2	71	L
	Mean	**3.727**				**11.667**			**71.333**	
	SD	**2.054**				**2.082**			**1.528**	
1.9.11; 30	Replicate A: 10% germinated by 35 days, 100% survived to 49 days.									
	5.1.1	3	R	D						L
	5.1.2	4		Y	2		1	2		L
	Replicate B: 10% germinated by 35 days, 100% survived to 49 days.									
	7.4.1	4	E	Y						
	7.4.2	6		Y						
	Mean	**4.25**								
	SD	**1.258**								
2.9.10; 31	Replicate A: 20% germinated by 35 days, 100% survived to 49 days.									
	5.2.1	3		Y						L
	5.2.2	4		N	2	10	3	3	70	ML
	5.2.3	0	3E, 1R	Y						
	5.2.4	4		N	3	9	1	2	66	ML

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	Replicate B: 10% germinated by 35 days, 100% survived to 49 days.									
	7.7.1	1	R	Y						L
	7.7.2	2		Y		15			60	L
	Mean	**2.333**				**11.333**			**65.333**	
	SD	**1.633**				**3.215**			**5.033**	
Oxy; 32	Replicate A: 25% germinated by 35 days, 100% survived to 49 days.									
	5.3.1	3	R	N	3	13	2	2	51	ML
	5.3.2	2		N	3	18	1	2	62	ML
	5.3.3	3		N	3	13	3	3	62	ML
	5.3.4	3		N	3	21	3	3	57	ML
	5.3.5	2		N	3	11	3	3	59	ML
	Replicate B: 5% germinated by 35 days, 100% survived to 49 days.									
	7.5.1	4	E	N	3	13	3	3	70	ML
	7.5.2	1		Y						
	Mean	**2.571**				**14.833**			**60.167**	
	SD	**0.976**				**3.817**			**6.306**	
Cent; 33	Replicate A: 15% germinated by 35 days, 100% survived to 49 days.									
	5.4.1	1	M	Y						
	5.4.2	5		Y						
	5.4.3	4		Y						
	Replicate B: 50% germinated by 35 days, 90% survived to 49 days.									
	7.1.1	3	6E, 4R	Y						
	7.1.2	3		Y						

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	7.1.4	4		Y						
	7.1.5	2		Y						L
	7.1.6	3		Y	2	12	3	2	67	ML
	7.1.7	4		Y						L
	7.1.8	4		Y						
	7.1.9	2		D						
	7.1.10	2		Y						
	Mean	**3.077**								
	SD	**1.115**								
2.9.1; 34	Replicate A: 5% germinated by 35 days, 100% survived to 49 days.									
	5.5.1	6	R	Y						
	Replicate B: 5% germinated by 35 days, 100% survived to 49 days.									
	7.2.1	2	R	Y						
	Mean	**4**								
	SD	**2.828**								
1.7.21; 35	Replicate A: 20% germinated by 35 days, 100% survived to 49 days.									
	5.6.1	3		N	3	8	2	1	71	M
	5.6.2	1	M	Y		18			47	L
	5.6.3	3		Y						
	5.6.4	4		Y						
	Replicate B: 65% germinated by 35 days, 1005 survived to 49 days after thinning to 10 seedlings.									
	7.6.1	6	E	N	2	14	3	3	88	M
	7.6.2	6		N	2	11	3	3	84	M

BC1 ID; Line	Field ID	True Leaves	Plant Stage	Predated	Spine Strength	# Spines	Leaf Shape	Leaf Margin	Height	Flowering
	7.6.4	4		N	2	13	1	2	90	M
	7.6.5	4		N	2	11	1	3	89	M
	7.6.6	6		N	2	11	1	3	78	ML
	7.6.7	5		N	1	14	3	3	84	M
	7.6.8	5		N	2	9	3	3	83	M
	7.6.9	6		N	2	12	3	2	90	M
	7.6.10	5		N	2	10	3	3	87	M
	Mean	**4.571**				**11.917**			**81.250**	
	SD	**1.505**				**2.644**			**12.084**	
2.9.57; 36	Replicate A: 40% germinated by 35 days, 100% survived to 49 days.									
	5.7.1	3		Y						
	5.7.2	3		Y						
	5.7.3	4		Y						
	5.7.4	4	R	Y						
	5.7.5	4		Y						L
	5.7.6	4		Y	1	14	3	2	72	ML
	5.7.7	3		Y	2	9	1	2	71	ML
	5.7.8	4		Y						L
	Replicate B: 15% germinated by 35 days, 100% survived to 49 days.									
	7.3.1	4		N	3	10	1	2	63	ML
	7.3.2	4	R	N	2	12	1	2	60	ML
	7.3.3	4		Y						L
	Mean	**3.727**				**11.25**			**66.5**	
	SD	**0.467**				**2.217**			**5.916**	

Note:

1. BC1 ID is the ID used for individuals in mapping the BC1 populations, while 'Line' refers to the field ID given to offspring of that BC1 individual.
2. This is the naming system for each plant in the field, where the first digit refers to the Row # (1-10) where the seed was planted, the second digit is the section (1-7) within a Row that the field was planted, and the third digit is the individual plant number. Plant 8.4.3 then, would be Row 8, section 4, plant 3, and the adjacent left-hand column shows this plant is Line 2, the offspring of BC1 plant 1.7.11. As this is the BC1S2 population, ten seeds from each of two BC1S1 parental plants (called Replicate A and Replicate B) from each line were planted in the field.
   Germination and survival rates at 35 and 49 days are given at the start of plant data for each Replicate within a line.
   Note that Replicate B for both Line 3 (Row8.6) and 18 (Row10.4) had 0% germination.
   Note also that the lines most obviously vigorous were line 21 Replicate A (Row3.6), line 35 Replicate B (Row7.6), line 2 Replicate B (Row8.4), and line 25 Replicate B (Row9.4).
3. This is count data of the number of pairs of true leaves on a plant at 35 days post seeding.
4. This measurement is an estimate of the growth stage at 49 days for a given Replicate (up to ten plants) for a Line, where E=elongated, R=rossette stage, M=mix of elongated and rossette.
5. Many of the plants were predated by 78 days post seeding. Predation of each plant is described here as Y=yes, predated, N=not predated, D=dead plant. Predation limited the ability to record further measurements.
6. Leaf spine strength was estimated relatively among plants on day 78, where 1=weak spine strength, 2=moderate spine strength, 3=strong spine strength.
7. The number of leaf spines were counted from a leaf subtending the primary capitulum on day 78.
8. Leaf shape was recorded on day 78 as per the IBPGR descriptors, where 1=ovate, 2=oblong, 3=lanceloate, and 4=linear.
9. Leaf margins were graded on day 78 as in IBPGR descriptors where 1=entire, 2=serrated, and 3=deeply serrated.
10. Plant height was measured on day 78 in cm to the highest point of the plant.
11. Flowering was estimated at 98 days post seeding by counting the number of capitula in bloom, such that Early (E)=25% in bloom;. Medium (M)=primary capitulum and several other capitula in bloom; Medium-Late (ML)= primary capitulum and one or two others in bloom; Late (l)= none in bloom. Flowering time estimates on plants which were unpredated are highlighted in grey.

5.2. Appendix B: Growth chamber data for parental plants *C. oxyacanthus* (Oxy) and *C. tinctorius* cv. Centennial (Cent) and cv. S-317 (S317).[1]

	Height						Branches				
Plant ID[2]	Day 56[3]	Day 91[4]	Day 149[3]	DF[5]	In Bloom[6]	Leaf Margin[7]	# Primary[8]	Location[9]	# Primary[10]	Total #[10]	DM[11]
Oxy-1	1	44	52	106	0	3	9	4	17	90	204
Oxy-2	0.5	36	40	104	0	2	9	4	14	110	209
Oxy-3	1.4	8	120	159	0	3	0	0	14	50	226
Cent-1	41.4	52	52	84	4	2	5	2	5	15	176
Cent-2	32.5	46	46	90	1	2	3	2	4	8	171
S317-1	35	45	43	86	3	2	4	3	4	18	174
S317-2	42	48	50	84	5	2	5	2	6	22	173
S317-3	36	44	42	83	4	2	5	2	5	17	172
S317-4	37	43	41	83	5	2	5	3	5	20	174
Mean:	25.20	40.67	54.00	97.67	2.44		5.00		8.22	38.89	186.56
SD:	18.41	12.97	25.18	24.67	2.19		2.78		5.19	36.87	20.70
Oxy Mean:	0.97	29.33	70.67	123.00	0.00		6.00		15.00	83.33	213.00
Oxy SD:	0.45	18.90	43.14	31.19	0.00		5.20		1.73	30.55	11.53
Cent Mean:	36.95	49.00	49.00	87.00	2.50		4.00		4.50	11.50	173.50
Cent SD:	6.29	4.24	4.24	4.24	2.12		1.41		0.71	4.95	3.54
S317 Mean:	37.50	45.00	44.00	84.00	4.25		4.75		5.00	19.25	173.25
S317 SD:	3.11	2.16	4.08	1.41	0.96		0.50		0.82	2.22	0.96

Note:

1. Shown are plants that germinated upon seeding 8 *C. oxyacanthus* plants and 6 of each cultivar of *C. tinctorius* on Oct. 20, 2009.
2. Numbering of parents was arbitrarily done to distinguish replicates within species.
3. Height in cm measured to top of the main stem. *C. oxyacanthus* plants were still in a rosette stage by Day 56.

Note Continued:

4. Height in cm measured to top of the main stem. Oxy-1 plant had 2 stems; Oxy-3 plant was still in rosette stage at this time.
5. Days for first flowering counted from seeding date.
6. These are counts of the number of flower heads in bloom on Day 91.
7. Leaf margins were recorded on Day 112, where 1=entire, 2=serrated, 3=deeply serrated.
8. These are counts of the number of primary branches on Day 91. Oxy-2 had 6 branches on one stem and 3 on the other.
9. Location of branches was measured as 0=no branches, 1=predominantly basal, 2=predominantly upper third of the plant, 3=predominantly upper two thirds of the plant, 4=from base to apex.
10. These are counts of the number of primary and total branches on day 175.
11. Days to maturity counted from seeding date.

5.3. Appendix C: Post-Harvest data for parental plants *C. oxyacanthus* (Oxy), and *C. tinctorius* cultivars Centennial (Cent) and S-317.

Plant ID	# Flower Heads	# Seeds/Capitulum[1]	1 Seed Weight[2]	# Seeds
Oxy-1	175	0	0.018	9
Oxy-2	254	0	0.01	18
Oxy-3	30	21	0.037	263
Cent-1	19	2	0.055	127
Cent-2	13	3	0.045	22
S-317-1	26	0	0	0
S-317-2	32	0	0.062	16
S-317-3	33	0	0.068	11
S-317-4	34	2	0.071	26
Mean	68.44	3.11	0.04	54.67
SD	85.41	6.81	0.03	86.85
Oxy Mean	153.00	7.00	0.02	96.67
Oxy SD	113.61	12.12	0.01	144.12
Cent Mean	16.00	2.50	0.05	74.50
Cent SD	4.24	0.71	0.01	74.25
S-317 Mean	31.25	0.50	0.05	13.25
S-317 SD	3.59	1.00	0.03	10.81

Note:

1. This was counted from the primary capitulum during harvest.
2. Measurements are in grams.

5.4. Appendix D: Growth chamber results for the BC1S2 plant generation.

Plant ID	DF[1]	Spine Strength[2,3]	Leaf Margin[2,4]	Location Branches[2,5]	Flower Colour[6]	Height[7]
1-7-4-2	66	1	2	2	3	66
1-7-5-2	90	1	2	2	4	56
1-7-6-2	78	2	1	3	3	70
1-7-7-2	-	1	2	2	3	61
1-7-8-2	70	1	3	2	3	78
1-7-9-2	76	1	2	2	3	69
1-7-10-2	67	1	3	2	3	61
1-7-11-2	76	1	3	3	4	57
1-7-12-2	83	1	2	2	4	81
1-7-13-2	70	2	2	2	6	78
1-7-14-2	73	1	3	2	3	86
1-7-15-2	80	1	3	2	4	69
1-7-16-2	65	2	2	2	-	68
1-7-17-2	69	1	2	2	3	65
1-7-18-2	67	2	2	2	3	51
1-7-19-2	76	2	1	3	4	80
1-7-20-2	69	1	2	2	6	78
1-7-21-2	65	1	2	3	3	62
1-7-22-2	-	1	3	2	5	51
1-7-24-2	74	3	2	3	5	64
1-7-25-2	82	1	2	2	6	63
1-7-26-2	70	1	3	2	3	65
1-7-27-2	70	1	3	3	4	72
1-7-28-2	72	2	2	2	6	68
1-7-29-2	72	2	2	2	6	70
1-7-30-2	72	3	1	2	4	70
1-7-31-2	73	2	2	2	4	55
1-7-32-2	67	1	2	3	4	56
1-7-33-2	66	3	2	2	6	67
1-7-34-2	68	2	2	3	5	60
1-7-35-2	68	3	2	4	3	54
1-7-36-2	70	3	2	3	4	47
1-7-37-2	76	2	1	2	4	67
1-7-38-2	73	3	2	3	4	57
1-7-39-2	67	1	2	3	3	73
1-7-40-2	62	3	2	2	5	45
1-7-41-2	72	2	2	2	5	65
1-7-42-2	76	1	3	2	4	63
1-9-2-2	63	3	3	2	5	71

Plant ID	DF	Spine Strength	Leaf Margin	Location Branches	Flower Colour	Height
1-9-3-2	68	1	3	3	6	81
1-9-4-2	73	2	1	2	4	56
1-9-5-2	65	1	2	2	4	73
1-9-6-2	72	1	3	2	4	73
1-9-8-2	64	2	2	2	5	65
1-9-9-2	60	2	3	2	6	64
1-9-11-2	68	1	2	2	4	59
1-9-12-2	66	1	2	2	4	83
1-9-13-2	64	1	3	2	3	67
1-9-14-2	69	1	1	0	4	79
1-9-15-2	72	3	2	3	4	63
1-9-16-2	76	3	2	2	4	62
1-9-17-2	62	2	3	2	5	61
1-9-18-2	76	2	2	3	3	48
2-9-1-2	65	1	2	2	4	65
2-9-2-2	66	2	2	2	6	54
2-9-4-2	64	2	2	2	4	66
2-9-5-2	63	1	1	2	3	73
2-9-6-2	76	1	3	2	4	53
2-9-7-2	-	1	2	2	4	64
2-9-8-2	66	1	2	2	4	69
2-9-10-2	62	1	2	2	3	55
2-9-11-2	87	1	2	3	6	55
2-9-13-2	80	2	2	2	4	51
2-9-14-2	72	1	1	2	3	64
2-9-15-2	62	3	3	2	4	64
2-9-16-2	66	3	2	4	3	56
2-9-17-2	63	1	3	2	3	65
2-9-18-2	66	3	2	4	4	56
2-9-19-2	61	1	1	2	3	64
2-9-20-2	67	3	2	3	4	46
2-9-21-2	72	3	2	2	3	51
2-9-22-2	70	2	2	3	6	57
2-9-24-2	65	3	2	2	3	53
2-9-25-2	81	3	1	3	4	62
2-9-26-2	66	1	2	2	3	88
2-9-28-2	77	3	2	2	4	69
2-9-29-2	63	3	2	2	3	51
2-9-30-2	62	1	2	2	6	52
2-9-33-2	73	1	2	2	3	77

Plant ID	DF	Spine Strength	Leaf Margin	Location Branches	Flower Colour	Height
2-9-34-2	61	3	3	2	4	66
2-9-36-2	78	2	2	3	3	52
2-9-37-2	79	2	2	2	4	44
2-9-38-2	70	3	2	3	3	54
2-9-39-2	72	1	1	2	3	62
2-9-40-2	66	3	2	3	3	75
2-9-41-2	66	3	3	2	3	59
2-9-44-2	87	1	3	2	4	59
2-9-45-2	69	2	2	3	3	63
2-9-47-2	62	1	2	2	4	57
2-9-48-2	79	3	2	2	3	67
2-9-49-2	70	1	2	2	3	63
2-9-50-2	66	3	3	2	5	56
2-9-51-2	68	2	2	2	4	65
2-9-52-2	59	2	2	2	4	62
2-9-53-2	76	3	1	3	3	55
2-9-54-2	77	1	2	3	4	56
2-9-55-2	78	2	2	3	3	62
2-9-56-2	76	1	2	2	3	64
2-9-57-2	66	3	1	3	3	43
2-9-58-2	-	1	2	2	3	43
2-9-59-2	63	2	2	2	3	57
2-9-60-2	66	2	2	3	6	50
2-9-61-2	84	2	2	2	4	69
Mean:	70.313					62.680
SD:	6.569					9.642

Note:

1. The number of days until first flowering was counted from the seeding date for all plants.
2. Spine strength, leaf margin and location of branches was measured on Aug 17,
   except for plants 2-9-58-2 a and b which were measured on Sep 10,
   and plants 2-9-37-2-b, 2-9-44-2-b, 1-7-22-2, 1-7-7-2, and 2-9-7-2, which were measured on Aug 25.
3. Leaf spine strength was evaluated relatively where 1=weak spine strength, 2=moderate spine strength,
   3=strong spine strength.
4. Leaf margins were graded as 1=entire, 2=serrated, 3=deeply serrated.
5. Predominant location of branches along the main axis was graded as 0=none,
   1=basal, 2=upper 1/3 of the plant, 3=upper 2/3 of the plant, 4=from base to apex.
6. Measurements are of wilted flower colour (for 100-141 day old plants), where 1=grey-white, 2=pale yellow,
   3=yellow, 4=light orange, 5=orange base, 6=orange, 7=deep red, 8=other.
7. Measured in cm from the top of the pot to the tallest point when plants were between
   92-125 days post seeding.

5.5. Appendix E: Results of growth chamber experiments for the BC1S3 plant population[1].

	Seedling Emergence[2]:				Height				Branches:				
Plant ID	Day 4	Day 7	Day 9	% by Day 9	Day 29[3]	Day 64[4]	Day 122[3]	DF[5]	Location[6]	# Primary[7]	# Primary[8]	Total[8]	DM[9]
1-7-6-3	0	3	5	50.00	3.6	40	56	99	3	4	4	27	181
1-7-9-3	0	4	4	66.67	2.2	45	45	86	3	5	13	22	231
1-7-13-3	0	4	5	66.67	1.9	53	53	88	3	5	7	19	175
1-7-14-3	0	3	4	50.00	3.6	66	77	101	2	10	11	34	165
1-7-19-3	0	3	4	50.00	0.4	42	48	91	4	6	19	52	233
1-7-24-3	0	1	4	16.67	2.3	36	62	108	3	3	8	18	140
1-7-28-3	0	3	4	50.00	6.5	58	58	97	3	4	6	24	182
1-7-29-3	0	4	4	66.67	5	46	46	93	3	4	5	19	191
1-7-30-3	0	0	4	0.00	10.6	69	82	90	2	9	9	19	172
1-7-31-3	0	4	4	66.67	6.8	48	50	86	3	4	4	14	177
1-7-32-3	0	4	4	66.67	2.3	48	51	92	3	4	4	17	161
1-7-34-3	0	4	4	66.67	4.8	57	60	86	3	6	10	31	179
1-7-35-3	0	0	4	0.00	3	44	49	99	4	4	7	22	233
1-7-36-3	3	6	6	100.00	4.5	44	45	90	2	5	7	17	179
1-7-38-3	0	6	6	100.00	5.2	43	53	91	3	6	6	17	172
1-7-39-3	0	4	5	66.67	4.7	52	66	102	2	5	6	19	158
1-7-40-3	1	6	6	100.00	5.6	45	46	94	3	4	4	14	231
1-7-41-3	0	2	4	33.33	4	52	59	94	3	10	11	39	182
1-7-42-3	0	1	4	16.67	0.7	44	56	105	3	0	13	21	175
1-9-2-3	0	4	4	66.67	4.3	46	67	94	3	4	6	24	175
1-9-3-3	0	5	5	83.33	6.5	50	58	88	3	3	4	18	170

	Seedling Emergence:				Height:				Branches:				
Plant ID	Day 4	Day 7	Day 9	% by Day 9	Day 29[3]	Day 64[4]	Day 122[3]	DF[5]	Location[6]	# Primary[7]	# Primary[8]	Total[8]	DM[9]
1-9-4-3	0	1	2	16.67	2.4	29	38	105	4	12	12	32	233
1-9-6-3	0	5	5	83.33	0.4	49	62	95	4	4	17	43	177
1-9-8-3	0	5	6	83.33	4.5	47	49	97	2	5	5	14	158
1-9-12-3	0	1	1	16.67	5.6	54	62	91	2	4	4	16	172
1-9-15-3	0	5	6	83.33	5.7	52	59	93	3	6	8	27	165
1-9-16-3	0	5	5	83.33	7	52	61	90	3	6	10	32	163
1-9-17-3	0	2	2	33.33	8.4	42	50	100	4	8	8	21	175
1-9-18-3	0	3	3	50.00	1.8	46	47	94	3	6	6	18	193
2-9-1-3	0	3	3	50.00	5.6	48	63	95	2	3	9	26	176
2-9-2-3	0	5	5	83.33	5.2	42	55	99	3	5	5	28	191
2-9-4-3	0	6	6	100.00	6.3	54	54	92	2	4	4	18	162
2-9-7-3	0	6	6	100.00	4.7	45	59	99	3	3	6	23	175
2-9-8-3	5	6	6	100.00	4.8	49	60	94	3	3	3	13	161
2-9-10-3	3	6	6	100.00	4.3	33	36	101	4	9	9	25	233
2-9-11-3	0	3	3	50.00	4	44	49	97	3	7	8	20	156
2-9-15-3	0	3	3	50.00	3.4	48	52	84	2	6	10	34	175
2-9-16-3	4	5	5	83.33	6	49	47	91	4	11	12	34	175
2-9-19-3	0	6	6	100.00	6.8	45	53	95	2	3	3	11	191
2-9-21-3	0	0	3	0.00	2.8	54	58	90	3	7	8	21	179
2-9-26-3	0	3	6	50.00	2	42	75	106	2	0	9	22	157
2-9-30-3	0	5	5	83.33	5.3	41	45	94	3	6	7	24	191
2-9-36-3	0	0	3	0.00	5.2	46	60	86	3	10	9	37	175
2-9-37-3	0	2	3	33.33	4.8	48	52	88	4	11	11	27	184
2-9-38-3	4	6	6	100.00	6	42	-	-	-	4	-	-	-

	Seedling Emergence:				Height:				Branches:				
Plant ID	Day 4	Day 7	Day 9	% by Day 9	Day 29[3]	Day 64[4]	Day 122[3]	DF[5]	Location[6]	# Primary[7]	# Primary[8]	Total[8]	DM[9]
2-9-40-3	0	6	6	100.00	8.5	45	57	91	3	5	5	23	158
2-9-45-3	0	2	2	33.33	7	42	43	92	3	4	7	32	179
2-9-47-3	2	4	3	66.67	3.8	32	37	94	4	5	8	28	234
Mean:	0.458	3.646	4.375	60.764	4.6	46.833	54.681	94.191		5.458	7.809	24.170	181.915
SD:	1.22	1.885	1.331	31.415	2.120	7.367	9.599	5.701		2.649	3.468	8.399	23.925

Note:

1. Seeding was on Nov 16, 2009, and the dates of measurements taken were counted from the number of days since the seeding date.
2. Emergence rates were counts of the number of seeds emerged by day 4, 7 and 9, and the total % emerged (of 6 planted) by day 9.
3. Height was measured in cm to the top of the plant.
4. Height was measured in cm to the top of the main stem.
5. This is the number of days from seeding to first flowering.
6. Location of branches measured as 0=no branches, 1=predominantly basal,
   2=predominantly upper third of the plant, 3=predominantly upper two thirds of the plant, 4=from base to apex.
7. Data are counts of the number of primary branches present on day 64.
8. Data are counts of the number of primary branches and total number of branches present on day 148.
9. This is the number of days from seeding to maturity.

5.6. Appendix F: Post-harvest measurements of the BC1S3 plant population.

Plant ID	# Flower Heads	# Seeds/Capitulum[1]	1 Seed Weight[2]	# Seeds
1-7-6-3	27	0	0.053	93
1-7-9-3	32	0	0.062	28
1-7-13-3	23	1	0.0541	354
1-7-14-3	40	10	0.038	468
1-7-19-3	31	0	0.052	99
1-7-24-3	13	0	0.027	68
1-7-28-3	32	0	0.059	232
1-7-29-3	26	0	0.072	97
1-7-30-3	31	0	-	0
1-7-31-3	16	0	0.063	34
1-7-32-3	19	0	0.054	17
1-7-34-3	53	0	0.039	105
1-7-35-3	29	0	0.043	76
1-7-36-3	21	0	0.056	18
1-7-38-3	19	6	0.046	465
1-7-39-3	20	0	0.07	31
1-7-40-3	25	0	0.054	89
1-7-41-3	48	0	0.045	90
1-7-42-3	43	0	0.046	41
1-9-2-3	28	0	0.047	224
1-9-3-3	18	0	0.06	131
1-9-4-3	41	0	0.029	22
1-9-6-3	42	0	0.053	212
1-9-8-3	19	0	0.065	98
1-9-12-3	20	1	0.035	220
1-9-15-3	35	0	0.052	371
1-9-16-3	41	1	0.058	229
1-9-17-3	30	0	0.061	1
1-9-18-3	26	13	0.049	152
2-9-1-3	22	0	0.055	238
2-9-2-3	35	0	0.057	119
2-9-4-3	22	0	0.072	80
2-9-7-3	28	0	0.063	98
2-9-8-3	25	0	-	0
2-9-10-3	48	0	0.064	5
2-9-11-3	23	0	0.03	2
2-9-15-3	48	0	0.067	57
2-9-16-3	40	0	0.068	89

Plant ID	# Flower Heads	# Seeds/Capitulum[1]	1 Seed Weight[2]	# Seeds
2-9-19-3	14	0	0.054	47
2-9-21-3	31	0	0.062	124
2-9-26-3	20	7	0.055	222
2-9-30-3	64	0	0.068	11
2-9-36-3	48	0	0.05	136
2-9-37-3	30	0	0.064	54
2-9-38-3	-	-	-	-
2-9-40-3	32	0	0.053	111
2-9-45-3	36	0	0.068	124
2-9-47-3	36	0	0.06	53
Mean	30.85	0.83	0.054	119.89
SD	11.223	2.657	0.011	115.17

Note:

1. This was counted from the primary capitulum.
2. Measurements are in grams.