University of Alberta

Relationships and introgression within *Carthamus* (Asteraceae), with an emphasis on safflower (*Carthamus tinctorius*).

by

Victoria G. Bowles

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

Biological Sciences

Fall 2010
Edmonton, Alberta

Examining Committee

Allen G. Good, Department of Biology

Jocelyn C. Hall, Department of Biology

James Cahill, Department of Biology

Andreas Hamann, Department of Renewable Resources

Dedication

"They say a little knowledge is a dangerous thing, but it's not one half so bad as a lot of ignorance." Terry Pratchett

This thesis is for everyone who encouraged my pursuit of knowledge.

Abstract

Carthamus (Asteraceae) contains both crop species (*C. tinctorius*, safflower) and weedy species, increasing the need for a better understanding of the genus. Despite previous studies, many outstanding questions remain regarding the phylogenetic relationships of safflower, especially with regards to the weedy species. Investigation of the relationships in *Carthamus* was done using sequence data. The closest relative to *C. tinctorius* was studied using microsatellite data. Microsatellite data was also utilized to track the introgression of *C. oxyacanthus* DNA into the *C. tinctorius* genome in an interspecific cross. Sequence data supports the division of the genus into two sections, *Carthamus* and *Atractylis*. Both sequence and microsatellite data reveal that most traditionally recognized species are not monophyletic. Microsatellite data indicates that *C. palaestinus* is the closest relative of cultivated safflower. Microsatellites also indicate that *C. oxyacanthus* DNA is able to move into the *C. tinctorius* genome, showing potential for breeding programs and raising concerns for potential transgenic crops.

Acknowledgements

I would like to thank both my supervisors for their help and for allowing me to work in their labs. To all my lab mates – thank you for help with both my research and editing. Also thank you for eating all the baking that has resulted from thesis writing. To my family and friends, thank you for your support and encouragement, especially on days when there was no end in sight. Thank you to Izabella for her help with illustrations and for all the specially made drinks to keep me going.

Table of Contents

1. General Introduction and Objectives……..1

    General Introduction……..2

    History of safflower……..3

    Genus *Carthamus*……..4

    Species Distributions……5

    Morphology and Biology……6

    Taxonomy……7

    Introgression and Crossing……8

    Crop Development……..10

    Microsatellites……..11

    Purpose of the thesis……..12

    Literature Cited……..13

2. Chapter 2: A phylogenetic investigation of *Carthamus* combining sequence and microsatellite data……24

    Introduction……25

    Materials and Methods……28

        Taxon Sampling……28

        Extractions, amplification and sequencing……29

        Phylogenetic analysis of sequence data……30

        Microsatellites……31

    Results……33

        Sequence analysis……33

Microsatellite analysis of sect. *Carthamus*……..34

Discussion……35

Sectional relationships within *Carthamus*……36

Relationships within sect. *Atractylis*……38

Relationships within sect. *Carthamus*……39

Species within *Carthamus* are not monophyletic……39

Progenitor of *Carthamus tinctorius*……41

Conclusions……43

Literature Cited……44

3. Chapter 3: Introgression of *Carthamus oxyacanthus* into the *C. tinctorius* genome, utilizing a BC1S1 generation of an interspecific cross and existing genetic maps……66

Introduction……67

Materials and Methods……70

Population……70

Field Season 2007……70

Field Season 2008……71

Marker Selection……72

Primer Amplification……73

Marker Analysis……73

Results……75

Field Season 2007……75

Field Season 2008……75

Marker analysis……76

Discussion……79

Morphological variation……79

Potential for crop development……80

Genetic variation and introgression……81

Conclusions……85

Literature Cited……87

4. General Conclusions……115

Future Work……120

Final Conclusions……124

Literature Cited……126

List of Tables

Table 1.1 Classification systems for *Carthamus*……20

Table 2.1 Classification systems for *Carthamus*……55

Table 2.2 Samples used in this study……57

Table 2.3 Microsatellite loci and Primers……62

Table 2.4 Phylogenetic information on nuclear and chloroplast regions utilized in this study……65

Table 3.1 Raw measurements of BC1S1 lines from field season 2008……108

Table 3.2 Raw measurements of cypselae for the 31 randomly chosen BC1S1 plants……111

List of Figures

Figure 1.1 Native range of *Carthamus*……21

Figure 1.2 Range of *Carthamus* species in North America……22

Figure 1.3 Capitulum of *Carthamus oxyacanthus*……23

Figure 2.1 Strict consensus of 8500 most parsimonious trees from analysis of combined ITS, *trn*T-*trn*L and *trn*L-*trn*F……51

Figure 2.2 Neighbour joining tree using Nei's genetic distance measure of scored microsatellite data……53

Figure 3.1 Crossing Scheme and naming system……91

Figure 3.2 Cypselae measurement diagrams……92

Figure 3.3 Linkage group maps and scaffolds for *C. oxyacanthus*……93

Figure 3.4 Scoring data for linkage groups of the BC1 and BC1S1 generations of the *C. oxyacanthus* x *C. tinctorius* interspecific cross……96

Figure 3.5 Morphology of parental genotypes from field grow out, summer 2007.102

Figure 3.6 Morphology of BC1S1, summer 2008 ……103

Figure 3.7 Phenotypic measurements of BC1S1 plants……104

Figure 3.8 Masked recombination events in the BC1S1 generation of the *C. oxyacanthus* x *C. tinctorius* interspecific cross……107

Chapter 1: Introduction and Objectives

Genetic resources, such as marker systems and maps, are extremely useful and valuable for plant breeding and molecular studies. Marker systems can be used for numerous applications including genetic diversity studies, creation of genetic maps and phylogenetic studies (Weising et al. 2005). A map of the species' genome leads to valuable knowledge of where markers and closely linked traits are located. Associating markers with specific traits aids in plant breeding programs through marker assisted selection of plants for specific traits (Lande and Thompson 1990, Collard et al. 2005). Information on the location of markers in relation to traits of interest may also aid in the determination of candidate genes responsible for focal traits using map based cloning (Jander et al. 2002).

Because genetic knowledge is so advantageous, a number of species are being developed as model systems (see Kramer 2009). Currently model species (e.g., *Arabidopsis* – Bell and Ecker 1993) and major crop species (e.g., *Zea mays* – Lee et al. 2002, *Oryza sativa*– Kurata et al. 1994, *Glycine max* – Cregan et al. 1999, *Brassica* – Piquemal et al. 2005; *Lycopersicon*– Tanksley et al. 1992) have the most well-developed genetic tools. Because *Carthamus tinctorius* (safflower) is a small acreage crop, it lacks many of the resources available for major crop species. Safflower has gained interest in recent years both as an oilseed crop and for biotechnical applications. Genetically modified lines have been developed for the production of modified oil (Arcadia Biosciences, Davis CA, USA) and pharmaceutical products (SemBioSys Genetics, Calgary AB, Canada). The increased interest in *C. tinctorius* as an alternative crop necessitates the development of genetic resources

for this species and a better understanding of the species, its relatives and their relationships. The relatives of a crop species contain valuable traits for crop development, but may also be a liability if crossing in nature creates weeds.

The work presented in this thesis is part of a project to develop genetic tools for safflower and to address regulatory concerns about the release of transgenic safflower crops. This thesis aimed to complete two parts of the project: (1) to clarify relationships within *Carthamus* and determine the closest relative to safflower, and (2) to study the introgression of genes from *C. oxyacanthus* genome into the *C. tinctorius* genome.

History of Safflower

Safflower is one of the world's oldest crops (Johnson et al. 2002). Evidence of seeds and dyes from *Carthamus* have been found in Egyptian tombs, along with pictographic representations (Smith 1996). Historically, the crop was grown for the flowers that contained cardimin, an agent that was used to dye cloth. The flowers were also used in food preparation and Chinese herbal medicines. It was not until the 1800s that safflower was grown as an oil crop (Mundel et al. 2004).

Safflower has traditionally been grown in small plots for personal use and remains a minor crop in the world market. Currently, safflower is grown in more than 60 countries, with area of growth totaling 691,436 Ha worldwide (FAO 2008) with greatest production in India, Mexico and the USA. India is the largest producer of safflower seed in the world with most of the oil used for the domestic food market. Safflower is still grown in China for herbal medicines derived from the floral tissue (Dajue and Mundel 1996). In North America, safflower is grown in the

USA, Canada, and Mexico for both oil and birdseed. In 1985, Canadian varieties of safflower were released and enabled greater production in these northern climates (Mundel et al. 2004). Safflower is now grown in the southern regions of the Canadian prairies, with the acreage ranging from 810 to 2025 Ha a year (Mundel et al. 2004) and approximately 17,920 tonnes being produced in 2008 (FAOstat http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor).

Genus *Carthamus*

Safflower is the commercial species in the genus *Carthamus*, which belongs in the sunflower family (Asteraceae or Compositae). *Carthamus* is a relatively small genus of 14 to 25 species (Keil 2006; López-González 1989; Vilatersana et al. 2005, 2007; Sasanuma 2008). Despite low species numbers, there is a fascinating diversity across the genus. For example, distributions of species vary widely. At least three species (*C. lanatus, C. leucocaulos* and *C. oxyacanthus*) are classified as noxious weeds and have spread well outside their native range, whereas other species have limited geographic distribution. Furthermore, variation in chromosome number across the genus (n=10 to n=32) is the focus of cytological studies (e.g., Vilatersana et al. 2000b, Ashri and Knowles 1960, Knowles and Schank 1964, Estilai and Knowles 1976). Two species (*C. creticus* and *C. turkestanicus*) have been shown to have originated from hybrids of the allopolyploid *C. lanatus* with other members of the genus (Vilatersana et al. 2007). Safflower (*C. tinctorius*) is the most economically important member of the genus, although some weedy species also have economic impact.

Species distributions

Carthamus is native to the eastern Mediterranean (Vilatersana et al. 2000a; Fig. 1.1), although the distributions of individual species vary within the region. *Carthamus lanatus* has a large range including Northern Africa (Egypt), Western Asia (Turkey) and parts of Europe (Albania, Bulgaria, Czechoslovakia, France, Greece, Hungary, Portugal, Romania, Switzerland and Ukraine) (Flora Europaea). In contrast, other species have more restricted ranges. *Carthamus oxyacanthus* is native to the Asia-temperate and Asia- tropical regions within Afghanistan, Iran, Iraq and Pakistan whereas *C. palaestinus* is only found in Western Asia (Israel).

Several species can be found outside their native range. *Carthamus oxyacanthus* is a weed in Pakistan and India (Kiel 2006) that has spread to North America where it is classified as a noxious weed. It is important to note that the range of *C. oxyacanthus* overlaps with the cultivation of *C. tinctorius* in the USA, mainly in California (Fig. 1.2). As such, eradication of *C. oxyacanthus* is proposed, and thus this species has limited range in North America (Keil 2006). Another potentially problematic species is *C. lanatus*, which is now found on four continents (Australia, Africa, Europe, North America). In fact, *C. lanatus* is a noxious weed in Australia and North America (Ash et al. 2002) and is known from five states in the USA (Fig. 1.2; Keil 2006). The only non-native species distributed in Canada are *C. tinctorius* and *C. creticus,* which are found in the USA along the western coast and in South Carolina (Fig. 1.2 – Keil 2006). Like *C. lanatus* and *C. oxyacanthus*, *C. creticus*

has been classified as a noxious weeds. Finally, *C. turkestanicus* is not present in North America, but is a noxious weed in the Mediterranean (Vilatersana et al. 2007).

Morphology and Biology

All species of *Carthamus* are annual thistles with spines present on the leaves and bracts to varying degrees, with the exception of a spineless variety of *C. tinctorius*. The highly branched plants have a large range in height across the genus (30 to 180 cm; Keil 2006) and within species (30 to 150cm for *C. tinctorius*; Dajue and Mundel 1996). After germination, safflower remains in a rosette stage and produce leaves near ground level before the stem elongates. The rosette stage is tolerant of cold weather, but due to the time spent in this stage, the plants can be vulnerable to competition with faster growing weeds (Dajue and Mundel 1996).

The members of the genus are addapted to the warm and dry climates of the Middle East and Mediterranean areas. The ability of *C. tinctorius* to grow in dry climates is partly attributed to the presence of large taproots that can grow 2 to 3 meters in length (Dajue and Mundel 1996) and allow the plants to access nutrients and moisture deep within the soil. In *Carthamus,* taproots are established during the rosette stage, before stem elongation occurs. The number of spines on the leaves increases from the lower to upper leaves, with the exception of spineless varieties of safflower (Smith 1996). Depending on the species, leaf margins can range between smooth and serrated (Smith 1996).

As with all members of the Asteraceae family, flowers of *Carthamus* are arranged in heads or capitula. Spiny bracts, called phyllaries, surround heads of *Carthamus* (Fig. 1.3). Long tubular flowers vary in colour throughout the genus

from white to red or purple, and become darker after blooming. Blooming time can be as long as four weeks from the time the primary capitulum blooms, until the last of the tertiary capitulum blooms (Dajue and Mundel 1996).

The fruits are specialized achenes, called cypselae. Occasionally, the cypsela is referred to as a seed in the safflower literature (e.g., Deshpande 1952). The outer fruit wall, or hull, can range in color from white to brown, and may or may not have stripes. Pappi (narrow overlapping scales at the top of the achene) are present in some species (*C. lanatus* and *C. tinctorius*; Fig. 1.3).

Taxonomy

Depending on the authority, the number of species within *Carthamus* ranges from 14 to 25 (Keil 2006; López-González 1989; Vilatersana et al. 2005, 2007; Sasanuma 2008). Variation in species number is reflected in different classification systems, highlighting taxonomic difficulties in the genus. Some researchers recognize broadly circumscribed species as being divided into subspecies (Hanelt 1963) whereas others elevate subspecies to species (Lopez-Gonzalez 1989). There is no consensus in the literature, but the three most commonly used classifications are Hanelt (1965), Estali and Knowles (1976) and Lopez-Gonzalez (1989). In addition, Vilatersana et al. (2000) recently proposed a classification system based on molecular data (Table 1.1). Inconsistencies between systems are also evident in the number of sections, which range from two (Vilatersana, 2000a) to five (Hanelt 1963) (Table 1.1).

Despite the many systems, there are commonalities among them. For example, all classifications (Vilatersana et al. 2000a, Lopez-Gonzalez 1989, Estali and Knowles 1976, Hanelt 1963) propose that all species with a chromosome number of 2n=24, except *C. nitudus*, are closely related. Species with this chromosome number include *C. curdicus, C. gypsicola, C. oxyacanthus, C. palaestinus,* and *C. tinctorius.* All classifications recognized that *C. nitidus*, although 2n=24, is unrelated; however, its exact relationships are still unclear. In addition, *C. creticus* and *C. turkestanicus* are often placed in the same section along with *C. lanatus* (Vilatersana et al 2000, Lopez-Gonzalez 1989, Estali and Knowles 1976), or are classified as subspecies of *C. lanatus* (Hanelt 1963).

Introgression and Crossing

The ability of plant species to hybridize can result in gene movement from one species to another (introgression). While this often occurs in nature, introgression is of particular concern when it occurs between crops and their wild relatives. Hybridization and introgression can have effects on speciation and the ability of plants to move into new environments (Martinsen et al. 2001). If the two species produce a viable hybrid, there is a risk that the hybrid will be able to move into territories that the either one or both of the parent species had previously been unable to occupy (Campbell et al. 2006). When a noxious weed is a close relative to a crop species, introgression potentially becomes a major concern due to increased risk of creating a new species or population (Ellestrand et al. 1999) with the invasiveness capabilities of the weedy parent.

Carthamus tinctorius is able to cross with several species in the genus, including those that differ in chromosome number; however, the resulting offspring are not always fertile (Mayerhofer et al. submitted, reviewed in McPherson et al. 2004). *Carthamus oxyacanthus, C. palaestinus,* and *C. tinctorius* are able to cross with each other (McPherson et al. 2004) and produce fertile offspring. Hybrids of *C. tinctorius* and either *C. oxyacanthus* or *C. palaestinus* have been documented in Israel (*C. palaestinus*) and in Pakistan and Iran (*C. oxyacanthus*) (McPherson et al. 2004). Crosses of *C. tinctorius* (n=12) with *C. leucocaulos* (n=10) and *C. tinctorius* with *C. lanatus* (n=22) both resulted in F1 plants that were infertile (Mayerhofer et al. submitted, McPherson et al. 2004). Crosses of *C. tinctorius* with *C. creticus* and *C. turkestanicus* (n=32) failed to produce viable F1 plants as well (Mayerhofer et al. submitted).

The ability of the species within the genus to cross and produce viable hybrids must be thoroughly investigated before the release of transgenic crop lines. Increased interest in safflower as a platform for genetic modification necessitates a better understanding of the genus. Safflower has been targeted for genetic modification, with Arcadia producing a modified oil line and SemBioSys creating lines that will produce various pharmaceuticals. In some crosses of safflower with its relatives, the transgene is selectively lost 21% of the time (Mayerhofer et al. submitted), but the escape of the transgene into wild populations through hybridization and introgression is still an important matter. Depending on the transgenic line, the end result of introgression may be wild plants producing pharmaceutical products or weeds being herbicide resistant.

Introgression of genomes is also beneficial to crops and plays a crucial role in development and breeding programs. Wild relatives are an important source of genetic variation for improving domesticated crop species as they may have traits that would benefit crops such as disease resistance or increased stress tolerance (Fernie et al. 2006, Tanksley and McCouch 1997, McCouch 2004). The lack of development for *Carthamus* over the last 20 years has left a large gap in crop improvement, restricting it to a minor crop (Mundel et al. 2004).

Crop Development

Sembiosys, a Calgary based company, is developing safflower lines that produce a variety of pharmaceutical compounds. Currently in development are safflower lines that produce insulin for diabetics and ApoAI, a drug for cardiovascular disease treatment (Sembiosys). This technology could take the production of safflower in a new direction and open new markets. The oil from some varieties of *C. tinctorius* is high in linoleum acid, making it suitable for the production of biodiesel (Meka et al. 2007). Meka et al. (2007) produced biodiesel fuel from *C. tinctorius* oil, with a 96.8% yield of useable product from the oil. With decreasing amounts of fossil fuels available, demand for alternative fuels will increase in the future. The potential for *Carthamus* is promising.

Microsatellites

Previous phylogenetic studies of *Carthamus* revealed low genetic variation between species using a number of marker systems (random amplified polymorphic

DNA [RAPD]; Vilatersana et al. 2005, single nucleotide polymorphism [SNPs] – Chapman and Burke 2007, variation in nucleotide sequences – Vilatersana et. al. 2000, Bowles et al. 2010). In light of these findings, it is clear that new marker systems are necessary to elucidate relationships among species. Microsatellites (also known as short tandem repeats – STRs or simple sequence repeats - SSRs) are a good candidate marker system for this genus for a number of reasons. First, microsatellites are co-dominant with relatively high levels of polymorphism (Ellegren 2004), making them highly useful for a genus with low levels of variation. Second, they have been shown to be useful for other problematic genera (finches - Petren et al. 1999; Western Canary Island lizard – Richard and Thorpe 2001; bovine – Ritz et al. 2000; *Lycopersicon* – Alvarez et al. 2001; sinojacki – Yao et al. 2008). Microsatellites are found throughout the genome of eukaryotes and consist of long strings of small repeats in the sequence (e.g., dinucleotide repeats ACACACACACA or trinucleotide repeat ACTACTACTACT). The nature of microsatellites is such that the mutation rate is higher than that of the surrounding genome (Jarne and Lagoda 1996). It has also been proposed that these areas may be hotspots for recombination in the genome (Treco and Arnheim 1986, Wahls et al. 1990), allowing the plants to recover lost variation or adjust rapidly to evolutionary changes (Foster and Trimarchi 1994, Rosenberg 1994 – GET out of Oliveira 2006).

Purpose of the Thesis

There are outstanding issues in resolving relationships within *Carthamus* and the identification of the closest relative to safflower. To these ends, Chapter 2

presents a phylogenetic analysis of the genus based on sequence and microsatellite data. This study, based on increased taxonomic sampling, represents the first explicit test of recent classification changes of Vilatersana et al. (2000). Relationships within sect. *Carthamus* are of particular interest because this group contains cultivated safflower. In addition, microsatellite data was used in an attempt to increase the resolution and determine the closest relative of the cultivated *C. tinctorius.* Although sectional relationships were clarified, species boundaries remain problematic. The closest relative to safflower also remains unclear in recent work (Chapman and Burke 2007, Chapman et al. 2010).

The unclear boundaries for the species within sect. *Carthamus* observed in Chapter 2, and the apparent ability for species in the section to cross with one another, raises concerns about the transfer of DNA between species. Chapter 3 focuses on the ability of *C. tinctorius* to exchange genetic material with a relative, *C oxyacanthus,* and examines what happens to the genome of *C. oxyacanthus* when it is combined with the genome of *C. tinctorius*. Using the newly developed microsatellite markers and the recently published *Carthamus* map (Mayerhofer et al. 2010), an interspecific cross population (*C. tinctorius* x *C. oxyacanthus*) has been utilized to examine the exchange of genetic material and combination of traits from the two parental species.

Literature Cited

Alvarez AE, van de Wiel CCM, Smulders MJM, Vosman B (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theoretical Applied Genetics 103:1283-1292.

Ash GJ, Raman R, Crump NS (2003) An investigation of genetic variation in *Carthamus lanatus* in New South Wales, Australia, using intersimple sequence repeats (ISSR) analysis. Weed Research 43:208-213.

Ashri A, Knowles PF (1960) Cytogenetics of Safflower (*Carthamus* L.) species and their hybrids. Agronomy Journal 52: 11-17.

Bell CJ and Ecker JR (1993) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137-144.

Campbel LG, Snow AA, Ridley CE (2006) Weed evolution after crop gene introgression: greater survival and fecundity of hybrids in a new environment. Ecology Letters 9:1198-1209.

Chapman MA, Burke JM (2007) DNA sequence diversity and the origin of cultivated safflower (*Carthamus tinctorius* L.; Asteraceae). Bmc Plant Biology 7: 60.

Chapman MA, Hvala J, Strever J, Burke JM (2010) Population genetic analysis of safflower (*Carthamus tinctorius*; Asteraceae) reveals a near eastern origin and five centers of diversity. American Journal of Botany 97:831-840.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-

196.

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kayer N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Science 39:1464-1490.

Dajue L and Mundel HH (1996) Promoting the conservation and use of underutilized and neglected crops 7 Safflower: *Carthamus tinctorius* L. International Plant Genetic Resources Institute. Rome, Italy.

Deshpande RB (1952) Wild safflower (*Carthamus oxyacantha* Bieb.) - a possible oilseed crop for the desert and arid regions. Indian Journal of Genetic Plant Breeding 12: 10-14.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539-563.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nature Reviews 5:435-445.

Estilai A, Knowles PF (1976) Cytogenetic studies of *Carthamus divaricatus* with 11 pairs of chromosomes and its relationship to other *Carthamus* species (Compositae). American Journal of Botany 63: 771-782.

Fernie AR, Tadmor Y, Zamir D (2006) Natural genetic variation for improving crop quality. Current Opinion in Plant Biology 9:196-202.

Foster L and Trimarch JM (1994) Adaptive reversion of frameshift mutation in *Escherichia coli* by simple base deletions in homopolymeric runs. Science 265:209-215.

Hanelt P (1963) Monographische Ubersicht der Gattung *Carthamus* L. (Compositae). Feddes Repertorium 67: 41-180.

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) Arabidopsis map based cloning in the post-genome era. Plant Physiology 129:440-450.

Jarne P and Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution 11:424-429.

Johnston AM, Tanaka DL, Miller PR, Brandt SA, Mielsen DC, Lafond GP, Riveland NR (2002) Oilseed crops for semiarid cropping systems in the Northern Great Plains. Agronomy Journal 94:231-240.

Keil DJ (2006) *Carthamus*. In: Flora of North America. Edited by Barkley TM, Brouillet L, Jeude H, Strother JL, Gandhi K, Kiger RW, Yatskievych K, Sarucchi JL. New York, New York, Vol.19 pp178-181.

Knowles PF, Schank SC (1964) Artificial hybrids of *Carthamus nitidus* Boiss. and *C. tinctorius* L. (Compositae). Crop Science 4:596-599.

Kramer EM (2009) New model systems for the study of developmental evolution in plants. Current Topics in Developmental Biology 86:67-105

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Shong HS, Tamua Y, Wang ZX, Momma T, Umerhara Y, Yano M, Sasaki T, Minobe Y (1994) A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genetics 8:365-372.

Land R and Thompson R (1990) Efficiency of marker-assisted selection in the

improvement of quantitative traits. Genetics Society of America 124:745-756.

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Molecular Biology 48:453-461.

López-González G (1989) Acerca de la Clasificación natural del género *Carthamus* L., s. l. Anales del Jardín Botánico de Madrid 47: 11-34.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. Evolution 55:1325-1335.

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

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

McCouch S (2004) Diversifying selection in plant breeding. Plos Biology 2:1507-1512.

McPherson MA, Good AG, Topinka AKC, Hall LM (2004) Theoretical hybridization potential of transgenic safflower (*Carthamus tinctorius* L.) with weedy relatives in the New World. Canadian Journal of Plant Science 84: 923-934.

Meka PK, Tripathi V, Singh RP (2007) Synthesis of biodiesel fuel from safflower

oil using various reaction parameters. Journal of Oleo Science 56:9-12.

Mündel HH, Blacksaw RE, Byers JR, Huang HC, Johnson DL, Keon R, Kubik J, McKenzie R, Otto B, Roth B, Stanford K (2004) Safflower production on the Canadian prairies: revisited in 2004. Agriculture and Agri-Food Canada, Lethbridge Research Center, Alberta. Pg 1.

Petren K, Grant BR, Grant PR (1999) A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proceedings of the Royal Society of London Series B-Biological Sciences 266: 321-329.

Piquemal J, Cinquin E, Couton F, Rondeau C, Seignoret E, doucet I, Perret D, Villeger MJ, Vincourt P, Blanchard P (2005) Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theoretical Applied Genetics 111:1514-1523.

Richard M, Thorpe RS (2001) Can microsatellites be used to infer phylogenies? Evidence from population affinities of the Western Canary Island lizard (*Gallotia galloti*). Molecular Phylogenetics and Evolution 20: 351-360.

Ritz LR, Glowatzki-Mullis ML, MacHugh DE, Gaillard C (2000) Phylogenetic analysis of the tribe Bovini using microsatellites. Animal Genetics 31: 178-185.

Rosenberg SM, Longerich S, Gee P, Harris RS (1994) Adaptive mutation by deletions in small mononucleotide repeats. Science 265:405-407.

Sasanuma T, Sehgal D, Sasakuma T, Raina SN (2008) Phylogenetic analysis of *Carthamus* species based on the nucleotide sequence of the nuclear SACPD gene and chloroplast *trn*L-*trn*F IGS region. Genome 51: 721-727.

Smith JR (1996) Safflower. AOCS Press, United States of America. Pg. 2-4

Tanksley SD, Ganal MW, Prince JP, de Vincente MC, Bonierbale MW, Brown P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics Society of America 132:1142-1160

Tanksley SD and McCouch SR (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063-1066.

Treco D and Arnheim N (1986) The evolutionarily conserved repetitive sequence d(TG-AC)$_n$ promotes reciprocal exchange and generates unusual recombinant tretrads during yeast meiosis. Molecular and Cellular Biology 6:3934-3947.

Vilatersana R, Brysting AK, Brochmann C (2007) Molecular evidence for hybrid origins of the invasive polyploids *Carthamus creticus* and *C. turkestanicus* (Cardueae, Asteraceae). Molecular Phylogenetics and Evolution 44: 610-621.

Vilatersana R, Garnatje T, Susanna A, Garcia-Jacas N (2005) Taxonomic problems in *Carthamus* (Asteraceae): RAPD markers and sectional classification. Botanical Journal of the Linnean Society 147: 375-383.

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

Vilatersana R, Susanna A, Garcia-Jacas N, Garnatje T (2000b) Karyology, generic delineation and dysploidy in the genera *Carduncellus, Carthamus* and *Phonus* (Asteraceae). Botanical Journal of the Linnean Society 134:425-438.

Weising KN, Nybom H, Wolff H, Kahl G (2005) DNA Fingerprinting in Plants: Principles, Methods and Applications, 2nd edition, pp. 41-45. Taylor & Francis Group, Boca Raton, FL.

Whals WP, Wallace LJ, Moore PD (1990) The Z-DNA motif $d(TG)_{30}$ promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture. Molecular and Cellular Biology 10:785-793.

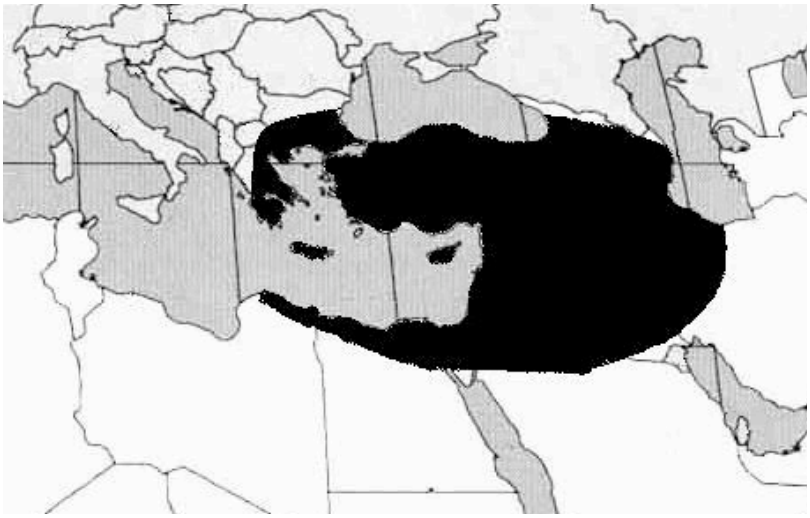
Yao XH, YE QG, Fritsch PW, Crux BC, Huang HW (2008) Phylogeny of *Sinojackia* (Styracaceae) based on DNA sequene and microsatellite data: Implications for taxonomy and conservation. Annals of Botany 101: 651-659.

Table 1.1: Classification systems for *Carthamus*

Hanelt (1964)	**Estali and Knowles (1976)**	**Lopez-Gonzales (1989)**	**Vilatersana (2000)**
Sect. *Carthamus* n=12	Sect. I n=12	Sect. *Carthamus* n=12	Sect. *Carthamus* n=12
Sect. *Odonthagnathis* n=10	Sect. II n=10	Sect. *Odonthagnathis* n= 10, 11	Sect. *Atractylis* n= 10, 11, 12*, 22, 32
Sect. *Atractylis* n=22, 32	Sect. III n=22	Sect. *Atractylis* n=22, 32	
Sect. *Thamnacanthus*	Sect. IV n= 11, 32	Removed from Genus: Sect. *Thamnacanthus*	
Sect. *Lepidopappus* n=10, 12*	Other n=12*		

* = *Carthamus nitidus*

Figure 1.1: Native range of *Carthamus*.



Shaded area represents the approximate native range of *Carthamus*. Figure has been adapted from map previously published by Vilatersana et al. (2000).

Figure 1.2: Range of *Carthamus* species in North America

Modified from maps found in Flora of North America (Keil 2006). Species found in North America are *C. creticus* (C), *C. lanatus* (La), *C. leucocaulos* (Le), *C. oxyacanthus* (O) and *C. tinctorius* (T). Letters denote the species that are found in each state or province.

Figure 1.3: Captilum of *Carthamus oxyacanthus*

Multiple flowers are contained in a single capitulum or head, enclosed by spiny bracts. Photo courtesy of the Good Lab.

Chapter 2: A phylogenetic investigation of *Carthamus* combining sequence and microsatellite data.

Victoria Bowles

A version of this paper has been published in Plant Systematics and Evolution

Bowles VG, Mayerhofer R, Davis C, Good AG, Hall JC (2010) A phylogenetic investigation of *Carthamus* combining sequence and microsatellite data. Plant Systematics and Evolution 287:85-97

Introduction

Carthamus (Asteraceae) is a small genus comprised of 14 to 25 species (Keil 2006; López-González 1989; Vilatersana et al. 2005, 2007; Sasanuma 2008) native to the Mediterranean region. An economically important member is *Carthamus tinctorius* (safflower), a commercial crop grown for oilseed, birdseed, spices, dyes and herbal medicines (Mündel et al. 2004). One of the oldest known crops (Johnston et al. 2002), safflower is now grown in more than 60 countries worldwide with cultivation in the New World and Australia being relatively recent. Three other species have also spread globally, but are classified as noxious weeds in North America and Australia (*C. lanatus*, *C. leucocaulos* and *C. oxyacanthus*). In addition, some species are polyploids (e.g., *C. creticus* and *C. turkestanicus*; Vilatersana et al. 2007) and there is potential for further hybridization (reviewed in McPherson et al. 2004). Hybridization is of critical interest because safflower is currently being genetically engineered for novel traits, possibly exposing related species to the introgression of transgenes (McPherson et al. 2004). Thus, *Carthamus* is a genus containing both cultivated plants and noxious weeds, which necessitates a thorough understanding of species relationships, species boundaries and the origin of cultivated safflower.

Taxonomy of *Carthamus* has been problematic, as demonstrated by differences among classifications and disagreement of generic boundaries. Hanelt (1963) divided the genus into five sections based on morphology and chromosome number: *Atractylis, Carthamus, Lepidopappus, Odonthagnathius,* and *Thamnacanthus* (Table 2.1). Based on cytological information, Estilai and Knowles (1976) also

recognized five sections, though with different species composition compared to Hanelt (1963): Sections I, II, III, IV and other (Table 2.1). Later, López-González (1989) made two changes to Hanelt's (1963) classification. First, sect. *Thamnacanthus* was removed from the genus and placed in *Phonus*, a modification supported by molecular data (Vilatersana et al. 2000a). Second, López-González (1989) subsumed sect. *Lepidopappus* into sect. *Odonthagnathius*. All classification systems imply a relationship between *C. oxyacanthus, C. palaestinus, C. persicus,* and *C. tinctorius* (Ashri and Efron 1964; Imrie and Knowles 1970) as well as *C. curdicus* and C. *gypsicola* (Hanelt 1963; López-González 1989). Recent phylogenetic analyses of ITS (internal transcribed spacer) sequence variation reveal two major clades requiring a reduction in the number of traditionally circumscribed sections (Vilatersana et al. 2000a). The first clade corresponds to Hanelt's (1963) sect. *Carthamus* based on sampling three of six species: *C. gypsicola, C. oxyacanthus* and *C. tinctorius* (Vilatersana et al. 2000a). The second clade, referred to as sect. *Atractylis* (Vilatersana et al. 2000a, 2005), includes Hanelt's (1963) sects *Atractylis, Lepidopappus,* and *Odonthagnathius,* and contains the weedy species *C. lanatus* and *C. leucocaulos*.

Although sectional boundaries have been clarified with limited sampling (Vilatersana et al. 2000a, 2005), outstanding questions regarding relationships within *Carthamus* remain. First, species relationships within the sect. *Carthamus* remain unresolved. After the establishment of the two major clades in the genus, Vilatersana and colleagues focused their efforts on resolving relationships within sect. *Atractylis* (Vilatersana et al. 2005, 2007). Because relationships within sect.

Carthamus are poorly understood, the identification of the closest wild relative to cultivated safflower remains elusive. Safflower has either been not included (e.g., Vilatersana et al. 2005, 2007) or, when included, results are difficult to interpret because the presented trees are unrooted (Chapman and Burke 2007). Both *C. oxyacanthus* (Deshpande 1952; Bassiri 1977) and *C. palaestinus* (Ashri and Knowles 1960; Ashri and Efron 1964; Chapman and Burke 2007) have been implicated as the wild progenitor of *C. tinctorius*. Furthermore, difficulties in elucidating species relationships and origins of safflower are compounded by unclear species boundaries in the genus. Whereas RAPD (Random Amplified Polymorphic DNA) data indicate species of sect. *Atractylis* are monophyletic (Vilatersana et al. 2005), unrooted phylogenetic trees of sect. *Carthamus* suggest species may not be natural groups (Chapman and Burke 2007) despite morphological differences.

A number of analyses reveal low levels of molecular variation across *Carthamus* (Vilatersana et al. 2005; Chapman and Burke 2007). Although DNA sequences (e.g., Vilatersana et al. 2000a) and other markers (SNPs (single nucleotide polymorphisms), Chapman and Burke 2007; RAPD, Vilatersana et al. 2005) have been used to address phylogenetic relationships, they were of limited success in resolving species relationships. For example, ITS data are unable to resolve species relationships within sections (Vilatersana et al. 2000a). Given the lack of variation found in previous analyses, additional markers are clearly needed. Microsatellites are highly polymorphic due to the high mutation rates (Goldstein et al. 1995). Although traditionally used for population studies, microsatellites are potentially useful for resolving species relationships (Goldstein and Pollock 1997) especially

when sequence data reveal low levels of variation (Yao et al. 2008). Moreover, microsatellite data has been useful in determining relationships between closely related species (Weising et al. 2005) for a range of taxa (finches - Petren et al. 1999; western Canary island lizard – Richard and Thorpe 2001; Bovine – Ritz et al. 2000; *Lycopersicon* – Alvarez et al. 2001; Sinojacki – Yao et al. 2008). As such, microsatellites are a good candidate for determining the relationships of the closely related species of *Carthamus.*

The primary focus of this study is to improve phylogenetic resolution within *Carthamus* using a tiered approach. First, sectional relationships were addressed by increasing previous taxonomic and character sampling of *Carthamus*. Second, microsatellites were used to investigate species relationships and boundaries in sect. *Carthamus*. Specifically, the goals of this study are (1) to evaluate species relationships within *Carthamus* using sequence variation, with the purpose of testing the sections proposed by Vilatersana et al. (2000a); (2) to examine relationships within sect. *Carthamus*, focusing on the identification of the closest relative to cultivated safflower; and (3) to address whether species in sect. *Carthamus* are monophyletic.

Materials and Methods

Taxon Sampling

Broad taxon sampling across the genus included four representatives from sect. *Carthamus* and twelve representatives from sect. *Atractylis* (Table 2.2). Species were chosen based on geography and taxonomy. We were unable to obtain material

of two geographically restricted species (*C. curdicus* and *C. gypsicola*). A total of 39 individuals were sampled, 37 from 16 species of *Carthamus,* and two outgroups (*Centaurea cyanus* and *Centaurea montana*) based on previous studies (Vilatersana et al. 2000a). To increase geographic coverage more than one individual was sampled from eight species: *C. alexandrinus* (2), *C. glaucus* (5), *C. lanatus* (6), *C. leucocaulos* (2), *C. oxyacanthus* (5), *C. palaestinus* (2), *C. tenuis* (2), *C. tinctorius* (7), and *C. turkestanicus* (2). All 39 representatives were examined in the broad, sequenced-based analyses. A reduced set of exemplars from sect. *Carthamus* were included for microsatellite analyses (see below), along with one accession of *C. lanatus*.

Extractions, amplification and sequencing

Genomic DNA was extracted from herbarium specimens and freshly grown material using either DNeasy Plant mini kits (Qiagen inc, Mississauga, Ont) or a CTAB method (Doyle and Doyle 1990). One nuclear ribosomal DNA and eleven chloroplast regions were initially screened on eight taxa to assess sequence variation for phylogenetic analysis: ITS (White et al. 1990), *trn*T - *trn*L (Taberlet et al. 1991), *trn*L-*trn*F (Taberlet et al. 1991), *atp*I-*atp*H (Shaw et al. 2007), *ndh*F-*rpl*32 (Shaw et al. 2007), *psb*J-*pet*A (Shaw et al. 2007), *ndh*J-*trn*F (Shaw et al. 2007), *rpl132-trn*L (Shaw et al. 2007), *3'rps16-5'trn*K (Shaw et al. 2007), *trn*Q-5'*rps16* (Shaw et al. 2007), *trn*H -*psb*A (Shaw et al. 2005) and *trn*S-*trn*G (Shaw et al. 2007). Only three regions showed potentially useful variation (e.g., > 2% of characters were parsimony informative) and were subsequently sequenced for all taxa: (1) ITS, (2) *trn*T-*trn*L, including the intergenic spacer between *trn*T and the *trn*L exon (Taberlet

et al. 1991), and (3) *trn*L-*trn*F region, including the *trn*L intron, the 3'*trnL* exon, and the intergenic spacer between the 3'*trn*L exon and the *trn*F gene (Taberlet et al. 1991). PCR products were purified (Qiagen PCR Purification Kits) then sequenced using Big Dye V3.1 and amplification primers. Cycle sequencing reactions were conducted on both strands and cleaned using Performa DTR V3 96-well short plates (Edge Biosystems, Gathersburg, MD). Extension products were resolved on an ABI3730 DNA analyzer (Applied Biosystems) and aligned using Sequencher v4.8 (Gene Codes Corporation, Ann Arbon, Mich.). Sequences were then adjusted manually using MacClade v.4.08 (Maddison and Maddison 2005).

Phylogenetic analyses of sequence data

To infer phylogenetic relationships, we performed maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) on the sequence data. All analyses were run on both individual and combined data sets. Given the low number of parsimony informative characters, permutation tail probability (Hillis and Huelsenbeck 1992) tests were run on all individual data sets in PAUP4.0* (Swofford 2002) to assess whether the data are non-random. All characters were equally weighted and unordered (Fitch 1971) in MP analyses. The MP parameters for the combined and individual data in PAUP 4.0* (Swofford 2002) were as follows: 100 replicates of random addition sequence and tree bisection and reconnection (TBR) branch swapping with the search being limited to 100 trees per replicate. One thousand bootstrap replicates were conducted with simple addition of taxon, TBR branch swapping saving 1000 trees per replicate for combined data and 100 trees in the individual analyses. Individual data MP bootstrap trees were visually

compared for conflicting branches of a bootstrap of >70%. The absence of such branches indicated that the data sets are congruent (Mason Gamer and Kellogg 1996) and could be combined.

For ML and BI, the optimal model of sequence evolution was chosen using the Akaike Information Criterion (AIC) in MrModeltest v.2.2 (Nylander 2004) for both individual and combined data sets (Table 2.4). Parameters of the model were estimated when conducting the ML search in Garli v.0.951 (Zwickl 2006) under default searching. Nonparametric bootstrap values were determined by running 100 replicates in Garli. Two independent Markov Chain Monte Carlo (MCMC) were run in Mr. Bayes V.3.2.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) under default parameters, each with one cold and three heated chains. The analysis was run for two million generations with a reduced temperature of 0.1 to increase mixing of the chains. The resulting average deviation of split frequencies was 0.0061. Trees from each of the MCMC runs were sampled every 100 generations. Excluding a burnin of 25% (5000 trees), posterior probabilities were calculated from the remaining trees.

Microsatellites

Given the low level of variation observed in the DNA sequences, alternative nuclear markers were explored to assess relationships within sect. *Carthamus*. Microsatellite loci were developed from a microsatellite enriched genomic library and from a publically available EST library (Compositae Genome Project). The enriched genomic library was created from DNA extracted from leaf tissue of *C. tinctorius* (cv. S317, Seedtec Inc., Woodland, CA) using the CTAB method (Doyle and

Doyle 1990). Genomic DNA was digested using RsaI, AluI and NheI before being treated with mung bean nuclease and antarctic phosphatase. SNX linkers (SNX-F/SNX-R) were ligated to the ends of genomic fragments and enriched for microsatellites following the protocol of Hamilton et al. (1999). Biotinylated dinucleotide probes ($(CT)_{14}$ and $(GT)_{14}$) were hybridized to linker ligated genomic DNA and captured on Steptavidin coated magnetic beads (Invitrogen, Carlsbad, CA), eluted and made double stranded by PCR using SNX-F. Products of the PCR reactions were digested with NheI. Enriched fragments were ligated into pBSII SK+ plasmid cut with XbaI and used to transform XL10 Gold competent cells. Clones were sequenced using T3 and T7 primers on an ABI3730 sequencer (Applied Biosystems, Foster City, CA) and manually screened for the presence of microsatellite loci. Primers were designed for suitable clones using Primer3 (Rozen and Skaletsky 2000) with the default settings. Microsatellites were also identified using a publicly available EST library from the Compositae Genome Project (http://cgpdb.ucdavis.edu/asteraceae_assembly/). EST sequences were screened for the presence of microsatellite loci using PrimerPro (http://www.cs.ualberta.ca/~yiferng/primerpro/), which also uses Primer3 to design primers that flank the microsatellite loci.

Microsatellite regions were amplified from four species in sect. *Carthamus* (*C. oxyacanthus, C. palaestinus, C. persicus, C. tinctorius*) and one in sect. *Atractylis.* Due to the risk of lost alleles reappearing (Nauta and Weissing 1996) microsatellite analyses was conducted only on sect. *Carthamus*, which contains closely related species.

Twenty-three loci (Table 2.3) were amplified using a protocol for fluorescent labeling adapted from Schuelke (2000). Labelled PCR products were run on an ABI 3730 (Applied Biosystems) with Gene Scan 600 LIZ (Applied Biosystems) size standard. Products were viewed and sized using Gene Mapper V4.0 (ABI).
A minimum of two replicates were made for every marker sampled. Peaks were then scored as the number of basepairs present. Distances for the microsatellite data were generated using Nei's genetic distance in MICROSAT V 1.5 (http://hpgl.stanford.edu/projects/microsat/), as suggested by Goldstein and Pollock (1997). Distance trees were inferred in PAUP *4.0 (Swofford 2002) using neighbour joining, which were rooted using Mesquite (Maddison and Maddison, 2009).

Results

Sequence analyses

Because there was little difference in phylogenetic relationships and support for individual analyses (data not shown), only combined data results are presented here. The MP search resulted in 8500 trees of length 345. The ML analysis resulted in a single tree with a score of -6012.6202. Bayesian inference resulted in 20000 trees with an average best state likelihood for the two MCMC chains of -6004.22.

All analyses result in two major clades that correspond to sect. *Carthamus* and sect. *Atractylis* (Fig. 2.1) with limited resolution within each section. Section *Carthamus* has strong support in all analyses (Posterior Probability [PP] 1.0; MP BS 98%; ML BS 97%) whereas the monophyly of sect. *Atractylis* is supported only by BI

(PP 0.99). For species in which multiple accessions were sampled, none are monophyletic in sect. *Carthamus* (*C. oxyacanthus, C. palaestinus, C. tinctorius*). Two representatives of *C. oxyacanthus* (244786 and 244792) are sister to all other members of the section (PP1.0, ML BS 97%, MP BS 95%). Most species within sect. *Atractylis* are not monophyletic with two exceptions of limited support: C. *leucocaulos* (PP 0.82), and *C turkestanicus* (PP 0.87). *Carthamus nitidus* is moderately supported as sister to all remaining members of sect. *Atractylis* (PP 0.99) as is the sister relationship between *C. alexandrinus* and *C. leucocaulos* (ML BS 77%; MP BS 77%). *Carthamus lanatus* is tentatively grouped with *C. creticus* (PP 0.76, *C. lanatus* 668-3, 24885, 34452, 244776, 244778) and *C. turkestanicus* (PP 0.71, *C. lanatus* 364-1). *Carthamus creticus, C. lanatus* (all members), and *C. turkestanicus* were grouped with *C. divaricatus* (PP 0.7).

Microsatellite analysis of section Carthamus

Given the strong support for monophyly of sect. *Carthamus* (Fig. 2.1), *C. lanatus* from sect. *Atractylis* was used as an outgroup for microsatellite analysis. All 23 loci (Table 2.3) amplified in four species of sect. *Carthamus* and the outgroup except for VL071, which did not amplify in *C. lanatus*. In addition, reliable amplification of primers VL004, ct171 and ct127 was problematic in *C. persicus,* likely due to low quality template DNA in this sample. The number of alleles for each locus ranged from three to seven across species, with many of the alleles shared between species. All loci were polymorphic within at least one species (Table 2.3) and 12 loci displayed unique alleles in four species (Table 2.3). A large number of

the species - specific alleles were found in *C. oxyacanthus.* The final dataset included 15 individuals and 23 markers.

The dendrogram (Fig. 2.2) based on microsatellite data reveals two major groupings of sect. *Carthamus*. This first group corresponds to all five individuals sampled from *C. oxyacanthus*, which is the only monophyletic species based on current sampling. The other distance-based group includes individuals of *C. palaestinus*, *C. persicus,* and *C. tinctorius*. Whereas the monophyly of *C. persicus* was not tested because only one individual was included, neither *C. palaestinus* nor *C. tinctorius* are monophyletic. Two representatives of *C. tinctorius* (Centennial and 244787) are nested with *C. persicus*. The remaining individuals of *C. tinctorius* are nested with a sample of *C. palaestinus* (663-3).

Discussion

Increased taxon sampling and analysis of both sequence and microsatellite data has led to a better understanding of relationships within *Carthamus*. First, monophyly of the genus and sect. *Carthamus* are strongly supported, results which are consistent with previous studies (Vilatersana et al. 2000a, 2005). Second, most species are not monophyletic, suggesting all species in the genus should be examined in more detail. Third, pertinent relationships are observed within sects. *Carthamus* and *Atractylis* despite limited resolution and lack of monophyletic species. Specifically, *C. oxyacanthus* appears to be the earliest diverging species in sect. *Carthamus*, and there is a close relationship between *C. lanatus* and both *C. creticus* and *C. turkestanicus*. Finally, *C. palaestinus* is likely the most closely related species to cultivated safflower.

Sectional relationships within Carthamus

Our analyses continue to support the division of *Carthamus* into two monophyletic clades that correspond to sections *Carthamus* and *Atractylis* (Fig. 2.1). Vilatersana et al. (2000a) first reported this generic partition based on sampling nine species and made subsequent classification changes without additional sampling of sect. *Carthamus* (Vilatersana et al. 2005). The analysis presented here increases taxon sampling by five species (one from sect. *Carthamus* and four from sect. *Atractylis*) and confirms the split of the genus. The congruence between the phylogeny presented here and that in Vilatersana et al. (2000a) is unsurprising as the ITS region was used in both analyses. However, the inclusion of the chloroplast data provides additional support for the monophyly of both sections from new markers. All analyses show much stronger support for the monophyly of sect. *Carthamus* than sect. *Atractylis* (Fig. 2.1). The current study represents 81% of species in the genus (based on López-González's 1989 classification). The following three species have yet to be sampled and should be included in additional molecular studies: *C. boissieri, C. curdicus* and *C. gypsicola*.

All members of sect. *Carthamus* have a chromosome number of n=12 (Hanelt 1963; López-González 1989; Vilatersana et al. 2000a). As this section is strongly supported by sequence data presented here and elsewhere (Vilatersana et al. 2000a), this chromosome number represents a putative synapomorphy for the section. *Carthamus nitidus* is the only species with this chromosome number found outside the section (Fig. 2.1, Hanelt 1963; Vilaterana et al. 2000a). However this species is clearly distinct from species of sect. *Carthamus*, which is reflected in

earlier classifications based on morphology and cytology (Hanelt 1963; Estilai and Knowles 1976; López-González 1989). The major basis for the separation of *C. nitidus* is artificial crossing studies (Knowles and Schank 1964). Most species within sect *Carthamus*, except *C. persicus*, are able to cross with one another (Ashri and Knowles 1960). However, *C. nitidus* is unable to cross with *C. tinctorius* and produce fertile offspring despite having the same chromosome number (Knowles and Schank 1964; Estilai and Knowles 1976).

Unlike sect. *Carthamus,* sect. *Atractylis* has no obvious putative morphological or cytological synapomorphies. Instead it appears to include species that do not fit into sect. *Carthamus.* In fact, the current circumscription includes other sections (*Odonthagnathius, Thamnacanthus, Atractylis* and *Lepidopappus*, Hanelt 1963; sects. II, III, IV, Estilai and Knowles 1976; Table 2.1). Included in this section are putative alloploid species such as *C. creticus*, C*. lanatus* and *C. turkestanicus.*

There is some conflict with the placement and chromosome number of *C. glaucus*, which in this study and others (Vilatersana 2000a; Hanelt 1963; Ashri and Knowles 1960; López-González 1989) is reported as n=10 and placed in the section with the same chromosome number. However recent work (Sehgal et al. 2009) included *C. glaucus* in sect. *Carthamus*, classifying it with a chromosome number of n=12. Concern with identification of *C. glaucus* from the USDA (used by Sehgal et al. 2009) has been expressed by Mayerhofer et al. (unpublished), which may explain the different placement of *C. glaucus*. Representatives of *C. glaucus* (Fig. 2.1),

collected from Iran, Turkey, Syria and Jerusalem (Table 2.2) are all strongly supported within sect. *Atractylis*.

Relationships within section Atractylis

Although most relationships in sect. *Atractylis* are unresolved, *C. nitidus* is strongly supported as sister to all remaining species (PP 1.00, Fig. 2.1). This finding is consistent with that of Vilatersana et al. (2000a), also based on ITS. Moreover, *C. nitidus* has features that are divergent from other members of the section, including a different chromosome number (n=12 versus n=10 or n=11). The placement of *C. nitidus* supports the hypothesis that the dysploidy series in the section (n = 10, 11 and 12) is the result of descending dysploidy from a base chromosome number of n=12 (Estilai and Knowles 1976; Vilatersana et al. 2000b). In addition, the pappus morphology of *C. nitidus* is different from other members of the section (Hanelt 1963). However, these differences were not deemed significant enough to warrant placing *C. nitidus* in its own section (Vilatersana et al. 2005). Relationships surrounding *C. nitidus* should be further investigated, as it may represent a bridge between the two sections.

Although the relationships within the sections are relatively unresolved, some patterns are consistent with previous classification systems. The clade including *C. creticus* (= *C. baeticus), C. divaricatus, C. lanatus* and *C. turkestanicus* is moderately supported (Fig. 1; PP 0.7). All these species correspond to sect. *Atractylis* as defined by Hanelt (1963, Table 1) except for *C. divaricatus* (which was not included in the classification). Estilai and Knowles (1976) also grouped *C.*

divaricatus with *C. creticus* and *C. turkestanicus* (Table 2.1), which is consistent with sequence data presented here (Fig. 2.1). Thus, the clade of *C. creticus, C. divaricatus, C. lanatus* and *C. turkestanicus* corresponds to a combination of section III and IV of Estilai and Knowles (1976), whose sectional affinities are based on morphological data and chromosome number. The close relationship of *C. lanatus* and *C. creticus*, is also seen using RAPD data (Vilatersana et al. 2005), however in that analysis *C. turkestanicus* is grouped with *C. glaucus,* not with *C. lanatus* and *C. creticus.* Consistencies between sequence analysis and previous classification systems reveal promise of increased investigation within the sect. *Atractylis*.

Relationships within section Carthamus

Both the sequence and microsatellite data indicate *C. oxyacanthus*, at least in part, is sister to all other representatives of sect. *Carthamus*. Whereas microsatellite data suggests this species is monophyletic, sequence data indicate that only some individuals of *C. oxyacanthus* (244786, 244792) are sister to all remaining samples including two individuals of *C. oxyacanthus* (185-4, 428-4). As such, the monophyly of this species needs further study. *Carthamus oxyacanthus* is morphologically distinct from remaining members of the section. For example, its seed morphology is divergent from the other species (Ashri and Knowles 1960).

Species within Carthamus *are not monophyletic*

The monophyly of morphologically based species in sect. *Atractylis* are not supported by sequence information presented here. Because only one individual of three species (*C. creticus, C. divaricatus, C. nitidus*) was sampled, monophyly could not be tested. Of the species that had more than one individual sampled (*C. glaucus,*

C. lanatus, C. leucocaulos, C. turkestanicus, C. tenuis) only *C. leucocaulos* and *C. turkestanicus* are monophyletic based on sequence data (PP 0.82 and 0.87 respectively). Individuals of *C. lanatus* are grouped with both *C. creticus* and *C. turkestanicus*, which is consistent with the hypothesis that *C. lanatus* is the male progenitor to both of these species through hybridization with other diploids (Khidir and Knowles 1970, Garnatje et al. 2006, Vilatersana et al. 2007). Resolution among the other species (*C. glaucus, C. tenuis*) is not sufficient to determine whether species are monophyletic or not. The observation that some of the species are not monophyletic contradicts the RAPD results presented by Vilatersana et al. (2005). However samples for the RAPD study (Vilatersana et al. 2005) were taken from single populations, or from populations that were relatively close geographically. In contrast, the samples presented here were taken from a broad geographical range (Table 2.2). This discrepancy reveals that broad geographic sampling is necessary when addressing species boundaries within *Carthamus*.

The boundaries between morphologically defined species in sect. *Carthamus* are also unsupported by either sequence or microsatellite data presented. A possible exception is *C. oxyacanthus* (Fig. 2.2), but monophyly was contradicted by sequence data (Fig. 2.1). Individuals of *C. palaestinus* are mixed with the members of *C. tinctorius.* This pattern is consistent with unrooted trees presented in Chapman and Burke (2007) where members of *C. palaestinus* and *C. tinctorius* are clustered together and mixed or unresolved. Also *C. oxyacanthus* is grouped with *C. persicus* in both trees presented here and by Chapman and Burke (2007).

A number of biological explanations explain the lack of congruence between morphologically based species and molecular analyses including introgression and incomplete sorting. Within *Carthamus*, introgression may explain the lack of distinct species boundaries. Introgression from domesticated plants to wild relatives is common (Ellstrand et al. 1999) and is likely in *Carthamus* as crossing has been documented between *C. palaestinus* and *C. tinctorius*, and *C. oxyacanthus* and *C. tinctorius* in natural settings (Ashri & Rudich 1965, Ashri & Knowles 1960). *Carthamus oxyacanthus, C. palaestinus* and *C. persicus* overlap within their natural ranges, with *C. tinctorius* likely being cultivated in the regions (McPherson et al. 2004), and *C. tinctorius* and *C. oxyacanthus* overlap in areas of North America (USDA), allowing the species to come into contact with each other in several locations. It has been proposed that these species are races of a single species (Imrie and Knowles 1970; Ashri and Efron 1964), which is a consistent with data presented here. However more information is needed before formal taxonomic changes are recommended. Subsuming all taxa into a single broad species potentially obscures valuable morphological differences between species.

Progenitor of Carthamus tinctorius

With the lack of defined species boundaries and low levels of resolution within the section, determining the closest relative to *C. tinctorius* is challenging. Previous studies hypothesized that either *C. palaestinus* or *C. oxyacanthus* is the progenitor to the cultivated *C. tinctorius* (Ashri and Efron 1964; Ashri and Knowles 1960; Deshpande 1952; Chapman and Burke 2007; Bassiri 1977; Hanelt 1963). The distribution of both species in the Near East is consistent with safflower originating

in this region (Knowles 1976, Smith 1996). However, *C. oxyacanthus* is morphologically distinct. In addition to seed differences mentioned previously (Ashri and Knowles 1960), *C. oxyacanthus* has more spines and smaller flower heads than *C. tinctorius* (Deshpande 1952). In contrast, the shape of the achenes and pappus indicates *C. palaestinus* is the closest relative to *C. tinctorius* (Ashri and Efron 1964). Most recent findings based on sequence data suggest that both are progenitors of different variations of *C. tinctorius* (Sehgal et al. 2008). Chloroplast data (Sehgal et al. 2008) showed a distinct difference in the sequence of *C. tinctorius* corresponding to the presence or absence of spines (*C. tinctorius* var*. tinctorius* and *C. tinctorius* var*. inermis* respectively). The sequence of *C. tinctorius* var. *tinctorius* and *C. tinctorius* var. *inermis* was virtually indistinguishable from that of *C. oxyacanthus* and *C. palaestinus* respectively (Sehgal et al. 2008).

Both the sequence and microsatellite data show that *C. oxyacanthus* is an earlier diverging lineage than either *C. palaestinus* or *C. tinctorius*. In fact, accessions of *C. palaestinus* and *C. tinctorius* are intermixed. One notable exception is the placement of *C. oxyacanthus* 428-4 from the sequence data (Fig. 2.1), which nested with *C. tinctorius* and *C. palaestinus* (Fig. 2.1). This relationship is not consistent with microsatellite data (Fig. 2.2). The close, indistinguishable, relationship between *C. tinctorius* and *C. palaestinus* (Figs. 2.1-2.2) indicates that *C. palaestinus* is most closely related to *C. tinctorius.* The microsatellite data presented here contradicts the dual parentage of *C. tinctorius* by *C. oxyacanthus* and *C. palaestinus* suggested by chloroplast data (Sehgal et al. 2008). This inconsistency may be because the microsatellites are of a more likely of nuclear origin and thus showing a

more complex picture. Also, in the present study, no *C. tinctorius* plants without spines were included, which would correspond to the second variety of the species. Thus, considering molecular phylogenetic analyses presented here and elsewhere (Chapman and Burke 2007) and morphological data (Ashri and Efron 1964), we hypothesize the closest relative to *C. tinctorius* is *C. palaestinus.*

Conclusions

The combined efforts of nuclear and chloroplast sequence along with microsatellite data provided insights into relationships within *Carthamus.* The sections *Carthamus* and *Atractylis* as proposed by Vilatersana et al. (2000a, 2005) are further supported here as monophyletic clades. Sequence data is also consistent with hybrid origins of *C. creticus* and *C. turkestanicus* (Khidir and Knowles 1970; Garnatje et al. 2006; Vilatersana et al. 2007). The inconsistency between traditionally described species and molecular data provided here, suggest that further study and possible taxonomic revisions are needed to define species in *Carthamus*. *Carthamus palaestinus* is likely the closest relative of safflower based on a combination of sequence data and previously reported data (Ashri and Efron 1964; Chapman and Burke 2007).

Literature Cited

Alvarez AE, van de Wiel CCM, Smulders MJM, Vosman B (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theoretical Applied Genetics 103:1283-1292.

Ashri A, Efron Y (1964) Inheritance studies with fertile interspecific hybrids of three *Carthamus* L. species. Crop Science 4:510-514.

Ashri A, Knowles PF (1960) Cytogenetics of Safflower (*Carthamus* L.) species and their hybrids. Agronomy Journal 52: 11-17.

Ashri A, Rudich J (1965) Unequal reciprocal natural hybridization rates between two *Carthamus* L. species. Crop Science 5: 190-191.

Bassiri A (1977) Identification and polymorphism of cultivars and wild ecotypes of Safflower based on isoenzyme patterns. Euphytica 26: 709-719.

Chapman MA, Burke JM (2007) DNA sequence diversity and the origin of cultivated safflower (*Carthamus tinctorius* L.; Asteraceae). Bmc Plant Biology 7: 60.

Deshpande RB (1952) Wild safflower (*Carthamus oxyacantha* Bieb.) - a possible oilseed crop for the desert and arid regions. Indian Journal of Genetic Plant Breeding 12: 10-14.

Doyle JF, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13-15.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539-563.

Estilai A, Knowles PF (1976) Cytogenetic studies of *Carthamus divaricatus* with 11 pairs of chromosomes and its relationship to other *Carthamus* species (Compositae). American Journal of Botany 63: 771-782.

Fitch WM (1971) Towards defining the course of evolution: Minimum change for a specific tree topology. Systematic Zoology 20:406-416.

Garnatje T, Garcia S, Vilatersana R, Vallès J (2006) Genome size variation in the *Carthamus* (Asteraceae, Cardueae): Systematic implications and additive changes during allopolyploidization. Annals of Botany 97:461-467.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463-471.

Goldstein DB, Pollock DD (1997) Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. Journal of Heredity 88: 335-342.

Hamilton MB, Pincus EL, Fiore AD, Fleisher RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries. Biotechniques 27: 500-507.

Hanelt P (1963) Monographische Ubersicht der Gattung *Carthamus* L. (Compositae). Feddes Repertorium 67: 41-180.

Hillis DM, Huelsenbeck JP (1992) Signal, noise and reliability in molecular phylogenetic analyses. Journal of Heredity 83:189-195.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics 17: 754-755.

Imrie BC, Knowles PF (1970) Inheritance studies in interspecific hybrids

between *Carthamus flavescens* and *C. tinctorius*. Crop Science 10: 349-352.

Johnston AM, Tanaka DL, Miller PR, Brandt SA, Mielsen DC, Lafond GP, Riveland NR (2002) Oilseed crops for semiarid cropping systems in the Northern Great Plains. Agronomy Journal 94:231-240.

Khidir MO, Knowles PF (1970) Cytogenetic studies of *Carthamus* species (Compositae) with 32 pairs of chromosomes II. Intersectional Hybridization. Canadian Journal of Genetics and Cytology 12:90-99.

Knowles PF (1976) Safflower In: Evolution of Crop Plants. New York pg. 31

Knowles PF, Schank SC (1964) Artificial hybrids of *Carthamus nitidus* Boiss. and *C. tinctorius* L. (Compositae). Crop Science 4:596-599.

Keil DJ (2006) *Carthamus*. In: Flora of North America. Edited by Barkley TM, Brouillet L, Jeude H, Strother JL, Gandhi K, Kiger RW, Yatskievych K, Sarucchi JL. New York, New York, Vol.19 pp178-181

López-González G (1989) Acerca de la Clasificación natural del género *Carthamus* L., s. l. Anales del Jardín Botánico de Madrid 47: 11-34.

Maddison WP, Maddison DR (2009) Mesquite: a modular system for evolutionary analysis. Version 2.72 http://mesquiteproject.org

Maddison DR, Maddison WP (2005) MacClade version 4.07. Computer program and documentation. Sinauer Associates, Sunderland, Mass.

Mason-Gamer RJ, Kellogg EA (1996) Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Graminae). Systematic Biology 45:524-545.

McPherson MA, Good AG, Topinka AKC, Hall LM (2004) Theoretical

hybridization potential of transgenic safflower (*Carthamus tinctorius* L.) with weedy relatives in the New World. Canadian Journal of Plant Science 84: 923-934.

Mündel HH, Blacksaw RE, Byers JR, Huang HC, Johnson DL, Keon R, Kubik J, McKenzie R, Otto B, Roth B, Stanford K (2004) Safflower production on the Canadian prairies: revisited in 2004. Agriculture and Agri-Food Canada, Lethbridge Research Center, Alberta. Pg 1.

Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics 143:1021-1032.

Nylander, J. A. A. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Petren K, Grant BR, Grant PR (1999) A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proceedings of the Royal Society of London Series B-Biological Sciences 266: 321-329.

Pollock DD, Bergman A, Feldman MW, Goldstein DB (1998) Microsatellite behavior with range constraints: Parameter estimation and improved distances for use in phylogenetic reconstruction. Theoretical Population Biology 53: 256-271.

Richard M, Thorpe RS (2001) Can microsatellites be used to infer phylogenies? Evidence from population affinities of the Western Canary Island lizard (*Gallotia galloti*). Molecular Phylogenetics and Evolution 20: 351-360.

Ritz LR, Glowatzki-Mullis ML, MacHugh DE, Gaillard C (2000) Phylogenetic analysis of the tribe Bovini using microsatellites. Animal Genetics 31: 178-

185.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers Bioinformatics Methods and Protocols: Methods in Molecular Biology, pp. 365-386. Humana Press, Totowa, NJ.

Sasanuma T, Sehgal D, Sasakuma T, Raina SN (2008) Phylogenetic analysis of *Carthamus* species based on the nucleotide sequence of the nuclear SACPD gene and chloroplast *trn*L-*trn*F IGS region. Genome 51: 721-727.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18: 233-234.

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

Sehgal D, Raina SN, Devarumath RM, Sasanuma T, Sasakuma T (2009) Nuclear DNA assay in solving issues related to ancestry of the domesticated diploid safflower (*Carthamus tinctorius* L.) and the polyploid (*Carthamus*) taxa, and phylogenetic and genomic relationships in the genus *Carthamus* L. (Asteraceae). Molecular Phylogenetics and Evolution 53:631-644.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu WS, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92: 142-166.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. American Journal of Botany 94: 275-288.

Smith JR (1996) Safflower. AOCS Press, United States of America. Pg. 2-4

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of 3 noncoding regions of chloroplast DNA. Plant Molecular Biology 17: 1105-1109.

Vilatersana R, Brysting AK, Brochmann C (2007) Molecular evidence for hybrid origins of the invasive polyploids *Carthamus creticus* and *C. turkestanicus* (Cardueae, Asteraceae). Molecular Phylogenetics and Evolution 44: 610-621.

Vilatersana R, Garnatje T, Susanna A, Garcia-Jacas N (2005) Taxonomic problems in *Carthamus* (Asteraceae): RAPD markers and sectional classification. Botanical Journal of the Linnean Society 147: 375-383.

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

Vilatersana R, Susanna A, Garcia-Jacas N, Garnatje T (2000b) Karyology, generic delineation and dysploidy in the genera *Carduncellus, Carthamus* and *Phonus* (Asteraceae). Botanical Journal of the Linnean Society 134:425-438.

Weising KN, Nybom H, Wolff H, Kahl G (2005) DNA Fingerprinting in Plants: Principles, Methods and Applications, 2nd edition, pp. 41-45. Taylor & Francis Group, Boca Raton, FL.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pg. 315-322. PCR Protocols: A guide to Methods and Applications. Academic Press, San Diego, California.

Yao XH, YE QG, Fritsch PW, Crux BC, Huang HW (2008) Phylogeny of *Sinojackia* (Styracaceae) based on DNA sequene and microsatellite data: Implications for taxonomy and conservation. Annals of Botany 101: 651-659.

Zwickl D J (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

Figure legends

Figure 2.1. Strict consensus of 8500 most parsimonious trees from analysis of combined ITS, *trn*T-*trn*L and *trn*L-*trn*F data. Bayesian posterior probabilities are above branches and MP/ ML support values are below. Species names indicated with a * have been classified as noxious weeds, and bolded names represent individuals of commercially grown Safflower.

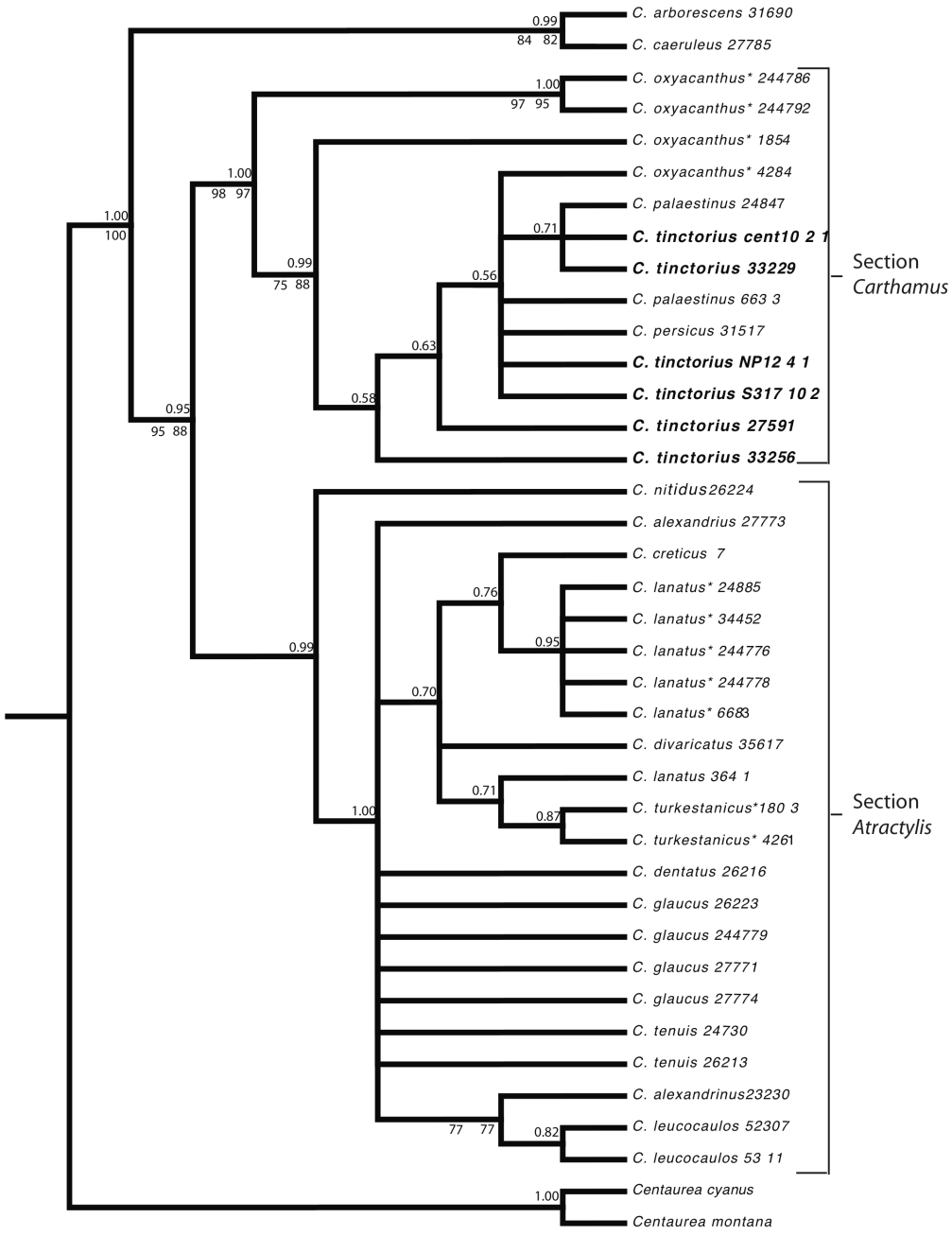

C. arborescens 31690
C. caeruleus 27785
C. oxyacanthus* 244786
C. oxyacanthus* 244792
C. oxyacanthus* 1854
C. oxyacanthus* 4284
C. palaestinus 24847
C. tinctorius cent10 2 1
C. tinctorius 33229
C. palaestinus 663 3
C. persicus 31517
C. tinctorius NP12 4 1
C. tinctorius S317 10 2
C. tinctorius 27591
C. tinctorius 33256
C. nitidus26224
C. alexandrius 27773
C. creticus 7
C. lanatus* 24885
C. lanatus* 34452
C. lanatus* 244776
C. lanatus* 244778
C. lanatus* 6683
C. divaricatus 35617
C. lanatus 364 1
C. turkestanicus*180 3
C. turkestanicus* 4261
C. dentatus 26216
C. glaucus 26223
C. glaucus 244779
C. glaucus 27771
C. glaucus 27774
C. tenuis 24730
C. tenuis 26213
C. alexandrinus23230
C. leucocaulos 52307
C. leucocaulos 53 11
Centaurea cyanus
Centaurea montana
Section Carthamus
Section Atractylis

Figure 2.2. Neighbour joining tree using Nei's genetic distance measure of scored microsatellites within sect. *Carthamus*. Bolded names represent individuals of commercially grown safflower.

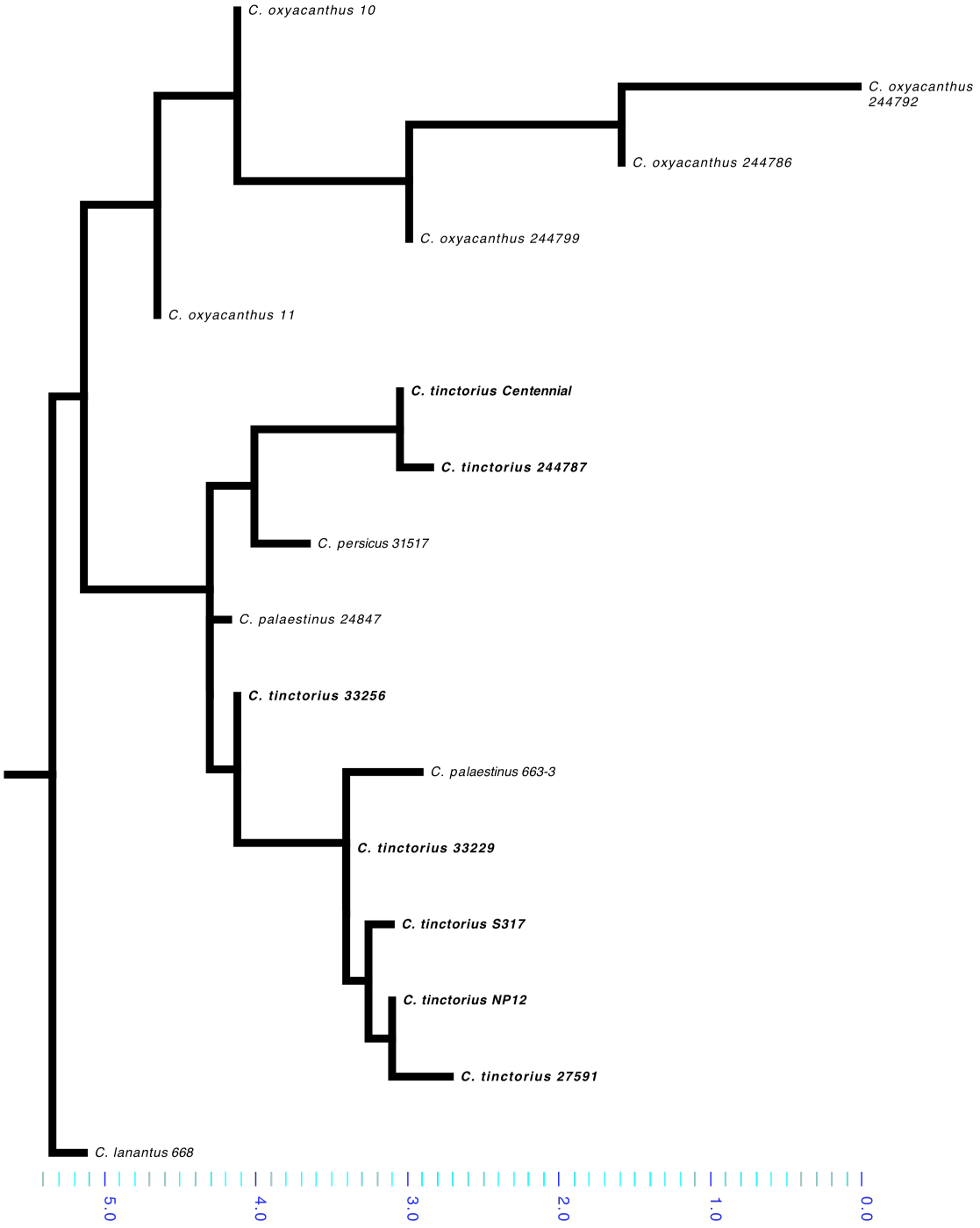

C. oxyacanthus 10
C. oxyacanthus 244792
C. oxyacanthus 244786
C. oxyacanthus 244799
C. oxyacanthus 11
C. tinctorius Centennial
C. tinctorius 244787
C. persicus 31517
C. palaestinus 24847
C. tinctorius 33256
C. palaestinus 663-3
C. tinctorius 33229
C. tinctorius S317
C. tinctorius NP12
C. tinctorius 27591
C. lanantus 668
5.0
4.0
3.0
2.0
1.0
0.0

Table 2.1: Classification systems for *Carthamus*.

Hanelt (1963)		Estilai and Knowles (1976)	
Sect. *Carthamus*	*C. curdicus* Hanelt *C. gypsicola* Iljin *C. oxyacanthus* M. Bieb. *C. palaestinus* Eig *C. persicus* Willd. *C. tinctorius* L.	Section I	*C. oxyacanthus* M. Bieb. *C. palaestinus* Eig *C. persicus* Willd. *C. tinctorius* L.
Sect. *Odonthagnathius*	*C. dentatus* Vahl ssp*. dentatus* Hanelt ssp*. rubber* (Link) Hanelt	Section II	*C. alexandrinus* (Boiss. & Heldr.) Bornm. *C. dentatus* Vahl *C. leucocaulos* Sibth. & Sm.
Sect. *Atractylis*	*C. lanatus* L. ssp*. lanatus* Hanelt ssp*. creticus* (L.) Holmb. ssp*. Montanus* (Pomel) Jahand. Et Maire ssp*. turkestanicus* (Popov) Hanelt	Section III Section IV	*C. lanatus* L. *C. creticus* L. (= *C. baeticus* (Boiss. & Reut.) Nym) *C. divaricatus* Beguinot & Vacc. *C. turkestanicus* Popov
Sect. *Thamnacanthus*	*C. arborescens* L. *C. riphaeus* Font Quer & Pau	Other	*C. nitidus* Boiss.
Sect. *Lepidopappus* Series *Lepidopappi*	*C. glaucus* M. Bieb. ssp*. glaucus* ssp*. glandulosus* Hanelt ssp*. anatolicus* (Boiss.) Hanelt ssp*. alexandrinus* (Boiss. et Heldr.) Hanelt *C. boissieri* Halácsy *C. tenuis* (Boiss. & Blanche) Bornm.		

	ssp*. tenuis*
	ssp*. gracillimus* (Rech. f.) Hanelt
	ssp*. foliosus* (Boiss.) Hanelt
	C. leucocaulos Sibth. et Sm.
Series *Leucauli*	*C. nitidus* Boiss.

Table 2.2: Samples used in this study.

Genus	Specific epithet	Voucher	ID number for figures	Source	Biogeography	*trn*L-*trn*F Genbank accession numbers	ITS Genbank accession numbers	*trn*T-*trn*L Genbank accession numbers
Carthamus	*alexandrinus* (Boiss. & Heldr.) Bornm.	Crampton & Shank,	27773	AHUC	Egypt	HM002861	GU969617	HM002822
Carthamus	*alexandrinus* (Boiss. & Heldr.) Bornm.	Crampton & Ashri	23230	AHUC	Australia	HM002860	GU969616	HM002821
Carthamus	*arborescens* L.	Crampton	31960	AHUC	unknown	HM002862	GU969618	HM002823
Carthamus	*caeruleus* L.	Crampton & Shank,	27785	AHUC	Portugal	HM002864	GU969620	HM002825
Carhtamus	*creticus* L.	seed stock	7	SemBioSys	Chile	HM002863	GU969619	HM002824
Carthamus	*dentatus* Vahl	Shank	26216	AHUC	Turkey	HM002866	GU969622	HM002827
Carthamus	*divaricatus* Beguinot. & Vacc.	Estilai	35617	AHUC	unknown	HM002867	GU969623	HM002828
Carthamus	*glaucus* M. Bieb.	Crampton & Shank,	27771	AHUC	Iran	HM002870	GU969626	HM002831
Carthamus	*glaucus* M. Bieb.	Schank	26223	AHUC	Turkey	HM002868	GU969624	HM002829

Carthamus	*glaucus* M. Bieb.	Fred Meyers	244779	GH	Jerusalem	HM002869	GU969625	HM002830
Carthamus	*glaucus* M. Bieb.	Crampton & Shank,	27774	AHUC	Syria	HM002871	GU969627	HM002832
Carthamus	*lanatus* L.	seed stock, PI235668	668-3	USDA	Netherlands	HM002877	GU969633	HM002838
Carthamus	*lanatus* L.	seed stock, PI326364	364-1	USDA	F. Soviet Union	HM002876	GU969632	HM002837
Carthamus	*lanatus* L.	Crampton & Shank,	27730	AHUC	Egypt			
Carthamus	*lanatus* L.	Ashri	24885	DAV	US	HM002872	GU969628	HM002833
Carthamus	*lanatus* L.	Fuller	34452	DAV	US	HM002873	GU969629	HM002834
Carthamus	*lanatus* L.	A. Contardo	244778	GH	Toscana	HM002875	GU969631	HM002836
Carthamus	*lanatus* L.	L. Adamovic	244776	GH	Pirot	HM002874	GU969630	HM002835
Carthamus	*leucocaulos* Sibth. et Sm.	seed stock, PI54-53	53-11	USDA	Unknown	HM002879	GU969635	HM002840
Carthamus	*leucocaulos* Sibth. et Sm.	Fuller	52307	DAV	USA	HM002878	GU969634	HM002839
Carthamus	*nitidus* Boiss	Schank	26224	AHUC	Jordan	HM002881	GU969637	HM002842
Carthamus	*oxyacanthus* M. Bieb.	seed stock, PI426185	185-4	USDA	Afghanistan	HM002884	GU969640	HM002845
Carthamus	*oxyacanthus* M. Bieb.	seed stock, PI426428	428-4	USDA	Pakistan	HM002885	GU969641	HM002846
Carthamus	*oxyacanthus* M. Bieb.	L. E. Rodin	244786	GH	Afghanistan	HM002882	GU969638	HM002843
Carthamus	*oxyacanthus* M.	H. Chandhuri	244792	GH	Lahore	HM002883	GU969639	

	Bieb.							HM002844
Carthamus	*oxyacanthus* M. Bieb.	Herb Giffith	244799	GH	Afghanistan			
Carthamus	*palaestinus* Eig	seed stock, PI235663	663-3	USDA	Israel	HM002887	GU969643	HM002848
Carthamus	*palaestinus* Eig	Ashri	24847	DAV	Israel	HM002886	GU969642	HM002847
Carthamus	*persicus* Willd.	Allizzi & S. Omar, 36183	31517	K	Iraq	HM002888	GU969644	HM002849
Carthamus	*tenuis* (Boiss. & Blanche) Bornm.	Schank	26213	AHUC	Israel	HM002890	GU969646	HM002851
Carthamus	*tenuis* (Boiss. & Blanche) Bornm.	Ashri	24730	DAV	Israel	HM002889	GU969645	HM002850
Carthamus	*tinctorius* L.	seed stock, Centennial	Cent10-12	Good Lab		HM002891	GU969647	HM002852
Carthamus	*tinctorius* L.	seed stock, NP12	NP12-41	Good Lab		HM002892	GU969648	HM002853
Carthamus	*tinctorius* L.	seed stock, S317	S317-10	Good Lab		HM002893	GU969649	HM002854
Carthamus	*tinctorius* L.	Knowles	33229	AHUC	Iran	HM002896	GU969652	HM002857
Carthamus	*tinctorius* L.	Knowles	33256	AHUC	unknown	HM002895	GU969651	HM002856
Carthamus	*tinctorius* L.	Pollard	27591	DAV	US	HM002894	GU969650	HM002855
Carthamus	*tinctorius* L.	R. M. Homan	244787	GH	Philippines			

Carthamus	*turkestanicus* Popov	seed stock, PI426180	180-3	USDA	Afghanistan	HM002897	GU969653	HM002858
Carthamus	*turkestanicus* Popov	seed stock, PI426426	426-1	USDA	Afghanistan	HM002898	GU969654	HM002859
Centaurea	*cyanus L.*	M. Gwizdala	116102	ALTA		HM002865	GU969621	HM002826
Centaurea	*montana L.*	G.W. & G. G. Douglas	53646	ALTA		HM002880	GU969636	HM002841

ID numbers are referenced for specific samples in the figures. Source abbreviations are as follows: AHUC and DAV, Davis Herbarium; ALTA, the University of Alberta Herbarium; GH, Harvard University Herbaria; K, Kew Gardens Herbarium; USDA, the United States Department of Agriculture.

Table 2.3: Microsatellite loci and primers

Primer name	Primer Sequences (5'-3')	Repeat motif	Expected size (bp)	Variable within species	Number of alleles per species*
VL001	TCCAAAATGCTTCAATGCAC GAGGAGGCCCTTTCATCTCT	CT	170	*C. oxyacanthus, C. tinctorius*	1, 2, 1, 1, 4
VL003	GTTAATGGGCTTGGGTTTGA TGGACAACTCTCAATGCAAGA	GT	173	*C. palaestinus, C. tinctorius*	1, 1, 2, 1, 2
VL004	TAAGCGTGAAAGAAGCAGCAT ACCATCATTCCCACCACAAT	GA	204	*C. oxyacanthus, C. tinctorius*	1, 2, 1, -, 3
VL005	AAAGCAGCATCCACCAGAAG ATCAACCGCCTTTGATTCAC	GA	163	*C. oxyacanthus, C. tinctorius*	1, 4, 2, 1, 3
VL014	AAACCTACCACTTGCAAACTCC GCCCCTAGTTTGCATTTGTTG	CA/CT	185	*C. oxyacanthus*#, *C. tinctorius*#	1, 4, 2, 1, 2
VL016	GGGCCATTGGCTTAGTGTTA CCTTCTTGAAGTGCTTGAGAGTT	GT	207	*C. lanatus*#, *C. oxyacanthus, C. persicus, C. tinctorius*	2, 3, 1, 2, 3
VL017	ATTCAAGTGGGTGTGGAGGA AAAGACGCTCCTTATAGCTAGGC	GT	236	*C. lanatus*#, *C. oxyacanthus, C. persicus, C. palaestinus, C. tinctorius*	2, 3, 2, 2, 4
VL021	TTTTGGGCGAGCTCACATAC CAAGATCCGCAGCTGTCATA	CA	150	*C. palaestinus, C. tinctorius*	1, 1, 2, 1, 2
VL030	TCATGATTCGAGTCCCGTTA CGACCAGTTAATATAGTTCGTTTTG	CA	102	*C. oxyacanthus*#, *C. persicus*	1, 3, 1, 2, 1
VL031	TCAAGTGGGTATGGAGGACA CGCTTTTAACTAATACATGCCAAA	GT	101	*C. oxyacanthus, C. persicus, C. tinctorius*	1, 3, 1, 2, 2
VL040	CATATTCCACCATGGTTTTGC TACACAACAAGAATAAGCATGTCAA	CA	265	*C. oxyacanthus*#, *C. persicus*	1, 2, 1, 2, 1
VL042	TGCAGGCAGAAAACACAATC CATGAGTACATAAGCCCTTCCA	GA	143	*C. lanatus, C. oxyacanthus*#, *C. tinctorius,*	2, 4, 1, 1, 2

VL044	GCACTGTAACACCCCACCTT TCCTCTCTTAGCCTCTCACCA	GT	207	*C. oxyacanthus, C. tinctorius*	1, 4, 1, 1, 3
VL046	TTCACGATTTGGAGTCTGAAA TAAGCCATATCCCACCTTCC	GA	284	*C. oxyacanthus#*	1, 3, 1, 1, 1
VL071	TGACCTTAATTTCCTCCCTTAATTT CAAGATCCGCAGCTGTCATA	(CA)12	210	*C. oxyacanthus#, C. tinctorius*	-, 3, 1, 1, 2
VL074	CATCGAACAGGTGGCAGAT CCTTCTCCTTCGGTTAGTTTTG	(GA)6	136	*C. oxyacanthus, C. persicus*	1, 4, 1, 2, 1
ct006	CAATTCGCTTCCACCAAGAT TACTCCTACCCGCCACAAAC	(GT)12	284	*C. oxyacanthus#, C. tinctorius*	1, 3, 1, 1, 2
ct014	TCTTCCATTCAGCCCAAGTT GGTGCATGCTTCATTCCTTT	(TC)12	277	*C. oxyacanthus, C. persicus, C. tinctorius*	1, 3, 1, 2, 3
ct050	CCTGCTGCTGGTCTTCCTAC CCAACACCTACGCATCCTTT	(GA)14	156	*C. oxyacanthus#, C. palaestinus, C. tinctorius*	1, 3, 2, 1, 2
ct127	GGAGTCGGTGTTTATCCCCT CCAGAGCACTGCAAGTGAAA	(TC)12	395	*C. tinctorius*	1, 1, 1, -, 3
ct137	AAGCGCTTCACTTCCCACTA TCCATCGTCGTTTGTAACCA	(AC)9	208	*C. oxyacanthus, C. tinctorius*	1, 4, 1, 1, 3
ct171	CCCCCATTGCTTCAATAAGA TCAACAATCAGTCAGCCAGC	(TC)9	291	*C. oxyacanthus#, C. tinctorius*	1, 4, 1, -, 2
ct185	ATTTGTCGCCATTATCGAGC CAATCTAAACCCTCTCGCCA	(TG)9	193	*C. oxyacanthus#, C. palaestinus, C. tinctorius*	1, 6, 2, 1, 3

Primers VL001 through VL074 were designed from the microsatellite enriched genomic library and the primers ct006 through ct185 were designed from EST sequences from the publically available library. Forward primers for each locus are the top primers, and the reverse primers are found below. Expected size is from the sequence data for *C. tinctorius*, S317. **C. lanatus, C. oxyacanthus, C. palaestinus, C. persicus, C. tinctorius*. #Species contain a unique allele.

Table 2.4: Phylogenetic information on nuclear and chloroplast regions utilized in this study.

	ITS	*trnT-trnL*	*trnL-trnF*	Combined
Length (bp)	892	622	1242*	2693
Number of informative characters	54	28	28	105
Models (MrModeltest, AIC[5])	GTR[1] + I[2] + G[3]	GTR + I + G	HKY[4] + I	GTR + I + Γ

* Indel of ~290bp
[1]Independent rates of substitution
[2]Among site rate variation modeled
[3]Some sites considered invariant
[4] Transistion and transversions occur at different rates, variation in base frequencies allowed
[5]Akaike Information Criterion

Chapter 3: Introgression of *C. oxyacanthus* into the *C. tinctorius* genome, utilizing a BC1S1 generation of an interspecific cross and existing genetic maps.

Victoria G. Bowles

Introduction:

The crossing of wild species with domesticated species can have both positive and negative effects. The occurrence of crossing in controlled breeding programs can result in the introgression of beneficial traits into the domesticated species such as disease resistance (Okogbenin et al. 2007) or cytoplasmic male sterility (CMS) restorer genes (Laughnan and Gabay-Laughnan 1983). Introgression in natural populations may allow species to adapt quickly by gaining new genes from related species, rather than having to accumulate numerous mutations in a gene to change the function (Rieseberg 2009). As well as providing new genetic variation, introgression may also serve as a repair mechanism by providing a copy of an allele to replace an existing damaged one (Rieseberg 2009, Ellstrand and Schierenbeck 2000).

However, the occurrence of introgression in the wild may also lead to an increase in the weediness of the hybrids or the extinction of the wild relatives (Ellstrand 1999). The introgression of the crop genome into that of its wild relatives is also a concern where genetically modified crops are being utilized. With the crossing of the genetically modified crop to a wild relative, there is the risk of the transgene being transferred into the wild population. Escape of herbicide resistance transgenes into wild populations has already been seen in several species including wheat, rice and oilseed rape (Sanchez Olguin et al. 2009, Seefeldt et al. 1998, Warwick et al. 2003).

The genus *Carthamus* contains both the crop species, *C. tinctorius* (safflower), and a number of weedy species. *Carthamus oxyacanthus* (wild safflower or jeweled daft thistle), a noxious weed in North America, is in the same section of the genus as *C. tinctorius* (Sect. *Carthamus*; Vilatersana et al. 2005, Bowles et al. 2010). *Carthamus oxyacanthus* has been hypothesized to be the closest relative or ancestor to *C. tinctorius* based on morphological and cytological information (Ashri and Knowles 1960; Hanelt 1963; Knowles 1958). The close relationship has been confirmed with molecular data (Vilatersana et al. 2000, Chapman and Burke 2007, Bowles et al. 2010) and there are no clear molecular boundaries between the species (Bowles et al. 2010). It has been proposed that *C. oxyacanthus* and *C. tinctorius* are races of the same species (Ashri and Efron 1964; Imrie and Knowles 1970) based on crossing and inheritence studies.

Interspecific hybridization within *Carthamus* is possible (McPherson et al. 2004) including between *C. tinctorius* and *C. oxyacanthus* which both have a chromosome number of n=12. The ability of *C. tinctorius* and *C. oxyacanthus* to produce fertile offspring when crossed has been documented (Imrie and Knowles 1970, Ashri and Efron 1964, Despande 1952), but with differing degrees of success (Mayerhofer et al. submitted, Ramanamuthy 1964, Ashri and Efron 1964). Mayerhofer et al. (submitted) found that crossing *C. tinctorius* and *C. oxyacanthus* had a success rate as high as 56%, with the resulting offspring being fertile, utilizing normal sexual crossing. However Ramanamurthy (1964) found that only 9% of the F1 hybrids from the *C. oxyacanthus* x *C. tinctorius* cross were fertile. Natural hybrids

between the two species have also been observed between the two species (Ashri and Knowles, 1960).

The close relationship between *C. tinctorius* and *C. oxyacanthus*, as well as their ability to produce viable hybrids, is of particular importance because these species often overlap in distributions. Within the United States, the major production area for *C. tinctorius* is California, where it is suited to the Mediterranean like climate. Along with being considered a noxious weed across the USA, *C. oxyacanthus* is also a common weed in many other places that *C. tinctorius* is grown including Pakistan and India (Deshpande 1952; Ashri and Knowles 1960), the worlds largest producer of *C. tinctorius.* Shared distribution patterns are particular concern now that *C. tinctorius* is being modified to produce insulin linked to the oil bodies by SemBioSys Genetics Inc. along with other pharmaceuticals. Therefore, the ability of *C. tinctorius* to cross with a noxious weed and exchange genetic material is a concern, if the genetically modified (GM) crops are to be released.

Utilizing a cross between *C. tinctorius* and *C. oxyacanthus,* Mayerhofer et al. (2010) have produced an interspecific linkage map, which is useful in crop breeding programs for marker assisted trait selection, and permits tracking the introgression of the *C. oxyacanthus* genome into the *C. tinctorius* genome. The *C. tinctorius* x *C. oxyacanthus* map shows synteny with the *C. tinctorius* x. *C. tinctorius* map, although the degree of similarity is hard to determine, due to low saturation of the maps and differing markers (Mayerhofer et al. 2010). In this study, the next generation of material derived from the interspecific mapping population developed by

Mayerhofer et al. (2010) was characterized. Markers from the existing map were used to track which areas of the *C. oxyacanthus* genome introgressed into the *C. tinctorius* genome. The amount of morphological variation was evaluated within the BC1S1 (backcross generation 1, selfed) population. In addition, various plant traits were measured to determine if meaningful variation is present among the genetic lineages for mapping and possibly crop development.

Materials and Methods

Population:

The population used in this study was derived from a cross between *C. tinctorius* (Centennial, USDA PI 538779) and *C. oxyacanthus* (USDA PI 426185) backcrossed to *C. tinctorius* (S317, USDA PI 599253) (Mayerhofer et al. 2010; Figure 3.1). The resulting plants, designated at the BC1 generation, were selfed to produce cypselae used for the 2008 field season. The *C. oxyacanthus* genetic map (Mayerhofer et al. 2010) was produced using 66 plants from the BC1 generation of the interspecific cross. Field and introgression work was done using the selfed cypselae (designated as the BC1S1 generation) of 31 members of the mapping population (Mayerhofer et al. 2010. All offspring from a single BC1 plant were considered a single genetic line (Fig. 3.1b). A single BC1S1 plant from each line was randomly chosen for marker analysis after harvest for a total of 31 plants analyzed (Table 3.2).

Field Season 2007:

The field site was east of Warner Alberta and was sheltered by trees, but was not fenced. The location of the site was chosen as safflower was known to grow in the area, and based on the availability of land with our cooperator. Plants were not watered, or fertilized at any point during the trial. Cypselae from five species (*C. lanatus, C. leucocalous, C. oxyacanthus, C. tinctorius, C. turkestanicus*), including parental species for the BC population, were planted by hand under 1" of soil, approximately 1.5 to 2" apart, in rows approximately 1 meter apart.

At harvest, plants were cut off close to the soil and placed in labeled paper bags. Plants were dried by placing the open bags in a growth chamber. Once dried, the plants were threshed and the cypselae were harvested.

Field Season 2008:

The field site was located in Southern Alberta (49°19' 12.51" N, 111°56' 32.05" W) and was chosen because our co-operator was experienced at growing safflower and had grown the crop on these fields over the previous 15 years. The field location was in an open field surrounded by cultivated safflower.

Cypselae from 31 selfed BC1 plants (BC1S1 population) were planted in a fenced plot, along with cypselae from each of the three parental varieties (*C. oxyacanthus, C. tinctorius* Centennial, *C. tinctorius* S317) used in the cross. Thirty-two cypselae were planted from each of the 31 selfed BC1 plants (referred to as a line) with 16 individuals being planted in two sections randomly distributed in the plot. Cypselae were had planted 1-2" deep (in soil moisture layer). Cypselae were

placed about 2" apart within a row, and rows were separated by 1m. The plot was surrounded by a 6' orange snow fence.

Plants were measured twice during the field season. At seven weeks after planting, the plants were checked for emergence. At that time, the number of true leaves and stem height were measured and recorded. Approximate stem height (measured from the ground to the top of the primary stem) of the shortest and tallest plants in a line (and all plants for in the line for 23), any flowering plants and branching patterns were measured seven weeks after the first measurements were taken (14 weeks after planting).

Plants were harvested by cutting off the plants below the lowest branches and placing the plant in a labeled paper bag. Harvested material was stored in the paper bags until threshing. Plants were dry at the time of harvest and, as a result, did not have to be dried after harvest. At threshing, the number of flower heads was recorded from one to five randomly selected individuals of the BC1S1 population. After threshing the cypselae of the 31 plants from the BC1S1 population chosen for marker analysis were measured for height, width and length (Fig. 3.2), as well as total weight of cypselae per plant and weight per 100 cypselae. Averages were determined for within line measurements (stem and plant heights, number of true leaves) and cypsela measurements were checked for correlations.

Marker Selection:

Markers for the analysis presented here were chosen from the *C. oxyacanthus* map produced by Cathy Archibald (Mayerhofer et al. 2010). The primers were selected based on even spacing and reliability of amplification from the five large, well-defined linkage groups of the Archibald map. Where possible, spacing between primers on the Archibald map was no more than 30 m.u., thus 40 primers were analyzed to cover the linkage groups through to the ends. Markers that had shown segregation distortion on the original map or were problematic for scoring were avoided during the selection process.

Primer Amplification

Marker regions were amplified using a three-primer system from Schuelke (2000) for fluorescently labeling the products. Forward primers were 5' tailed with an M13 sequence (TGTAAAACGACGGCCAGT) allowing for labeling of the PCR products. M13 sequence primers were labeled with four fluorescent labels for the different labeling reactions. PCR reactions contained 0.75mM $MgCl_2$, 0.2mM dNTPs, 0.067mM reverse and M13 labeled primers, 0.267mM forward primer, 2.5 units of Taq DNA polymerase and 50-100ng of template in 15ml. Thermocycling conditions were as follows: 94°C (5 min.); 30 cycles of 94°C (30sec), 56°C (45sec), 72°C (45sec); 9 cycles of 94°C (30 sec), 53°C (45 sec), 72°C (45 sec); ending with 72° for 10 minutes.

Marker Analysis

After PCR amplification of the markers, the products were diluted (1 in 10) and 2 ul was added to 8ul Formamide before being run on an ABI 3730 DNA Analyzer (Applied Biosystems Inc) and the output sized using Genemapper (Applied Biosystems, Foster City, CA). Products were sized using the Genescan LIZ-600 marker for reference.

After sizing, the allele sizes were determined to belong to either *C. tinctorius* or *C. oxyacanthus* based on parental screening data provided by C. Archibald. To aid in analysis alleles were converted to A for *C. tinctorius* alleles and B for *C. oxyacanthus* alleles, individuals containing both a *C. tinctorius* and *C. oxyacanthus* allele were scored as H (Figure 3.7 a-e) for each marker. After scoring, each marker was checked for the presence of segregation distortion (SD) using a chi-squared test with an expected ratio of 5:2:1 (AA:AB:BB) for the backcross population (Appendix 3.2). Markers were placed on in the linkage group in the ordered determined by the original map (Mayerhofer et al. 2010) (Fig. 3.3). Scoring data for the markers was placed in this order for each linkage group (Fig. 3.4). Recombination frequencies (RF) were calculated along each linkage group based on the 31 individuals. However, due to the diploid nature of the plants, there is the possibility of "masked" recombination events in the data (Figure 3.8). Being unable to determine where these recombination events may have occurred over a sequential set of heterozygous markers (Figure 3.8), recombination frequencies were based on the most parsimonious explanation (using a minimum number of recombination events).

Results:

Field Season 2007

In the summer of 2007, a preliminary field season of the parental species was performed for familiarization with the species' growth habit in natural settings and to ensure that both parental types would be able to fully mature during the growing season. At harvest there were notable differences between parental *C. tinctorius* and *C. oxyacanthus* genotypes in height, branching and flower heads (Fig. 3.5). Both parental species were able to grow to maturity before harvest and set cypselae. However, harvesting on September 29, 2007 was later than is typical for crops in the area. At harvest, predation of the plants was suspected due the presence of stems without heads and uprooted plants.

Field Season 2008

In the summer of 2008 BC1S1 plants and parental varieties were planted in the field for the purpose of assessing phenotypic variation within and between the offspring of the mapping population, as well as increasing the cypsela stocks and obtaining material for DNA extraction. At the time of measurements for emergence (seven weeks after planting), not all seedlings had emerged. Of the emerged seedlings (parental and BC1S1), the number of true leaves varied from zero to nine, with an average of 4.6 across both parental and BC1S1 lines, and stem height from 0 to 36 mm with an average height of 9.6 mm (Table 3.1).

The second visit revealed that the plants from both parental species (*C. oxyacanthus* and *C. tinctorius*) were larger than expected based on the previous year's field season (Fig. 3.6), with only subtle differences between the parental species. BC1S1 individuals within a line (offspring of same parental plant) were fairly uniform in height, with a difference of about 25cm as the greatest different in height within a line (Table 3.2). Accurate height measurements of individual plants were difficult to obtain due to do the closeness of the individual plants and the intertwined branches of multiple plants. Across all parental and BC1S1 lines, the tallest plant was 31cm and the shortest was 90cm, with an average height of 62cm. The majority of plants were highly branched and only three individuals had begun to flower. Each of the three flowering plants had only the primary capitulum. There was no sign of herbivore predation on any of the plants, as evidenced by missing plants, or a change in growth habit.

All plants were harvested after 25 weeks of growth. The following were recorded from one randomly chosen plant per line and parental species: the number of capitula and cypsela measurements. Thus a total of 33 plants were measured in the morphological analyses. Number of capitula ranged from 7 to 148, with parental lines having capitula 43 (*C. tinctorius*) and 100 (*C. oxyacanthus*). Across all BC1S1 lines, the average cypselae number was 36 with a range from 2 to 3088 (Fig. 3.5) while the average cypselae weight per 100 cypselae was 2.89g with a range from 1.7g to 3.6g. (Fig. 3.7, Table 3.2). Cypsela coloration ranged from white to tan cypselae with brown stripes present on some individuals. The size of the cypselae

ranged from 6.12mm to 8.52 mm in length with an average of 8.03mm, 2.82 to 4.13 mm in width (average = 3.62mm) and 2.8 to 3.28 mm in height (average = 3.03mm).

Marker Analysis

The five linkage groups chosen provided sufficient coverage of the linkage groups 1, 3, 4b, 5 and 6 from the *C. tinctorius* map (Mayerhofer et al. 2010). A total of 40 markers were selected across all linkages groups with seven and nine markers examined per linkage group (Fig. 3.3 a-e). Nineteen of the 40 markers had amplification problems in select individuals such that these individuals have missing data for these markers. Fifteen of the chosen makers were able to distinguish between the two parental *C. tinctorius* genotypes (Centennial and S317 – Fig 3.4 a-e).

After testing each of the markers for SD, seven of the 40 (17.5%) markers (ct 657, ct297, ct 32, ct266, ct 642, ct639 and ct 384) showed significant deviation from the expected pattern. Distorted markers were present on each linkage group with the exception of linkage group 4b. Linkage groups 5 and 6 each had more than one marker showing SD, but those markers were spread across the linkage group and not found clumped together (Fig. 3.4a-e).

A decrease in RF from the original map to the BC1S1 generation is seen between markers ct 381 and ct201 (Fig. 3.4e), ct495 and ct605 (Fig. 3.4c) and ct598 and ct405 (Fig. 3.4a). The RF between other markers was otherwise similar to the previous generation or saw an increased in recombination. Linkage group 5 (Fig.

3.4 d) showed a large increase in recombination between markers along the entire linkage group, resulting in markers ct458 and ct353 appearing unlinked (RF = 56.7).

Recombination between the *C. tinctorius* and *C. oxyacanthus* genomes was observed on all linkage groups (Fig. 3.4 a-e). There were no individuals that were homozygous for *C. oxyacanthus* alleles along an entire linkage group, although individuals retained an entire *C. oxyacanthus* copy of the linkage group in 3, 4b and 6 (Fig 3.4 b, c, e). In addition, these linkage groups contained sections of the *C. oxyacanthus* and *C. tinctorius* copies. The majority of *C. oxyacanthus* sections were found grouped together on a linkage group as opposed to interspersed with *C. tinctorius* DNA along the linkage group.

Linkage group 1 (Fig. 3. 4a) had 18 of 31 individuals containing DNA from *C. oxyacanthus* to varying degrees. No individuals contained a *C. oxyacanthus* allele for the marker ct657, making it the only marker of the linkage group to be without alleles from *C. oxyacanthus*, despite *C. oxyacanthus* alleles being present in BC1 generation.

Linkage group 3 (Fig. 3.4b) had 11 individuals containing *C. oxyacanthus* DNA, with four of those individuals having a *C. oxyacanthus* chromosome (*C. oxyacanthus* allele at each marker). Only three individuals were homozygous for any *C. oxyacanthus* allele along the linkage group.

On linkage group 4b (Fig 3.4c) several individuals had single *C. oxyacanthus* alleles surrounded by *C. tinctorius* alleles. One individual contained *C. oxyacanthus*

alleles for each marker on the linkage group, while 10 individuals were homozygous for at least one *C. oxyacanthus* allele.

Fifteen individuals were homozygous for *C. oxyacanthus* for a minimum of one marker on linkage group 5 (Fig 3.4d). Single *C. oxyacanthus* alleles flanked by *C. tinctorius* alleles were found in 20 individuals (Fig. 3.4d). Markers ct458 and ct353 showed so much recombination that the two markers now appear unlinked (RF 56, Fig 3.3d). Only six individuals had no *C. oxyacanthus* DNA by the BC1S1 generation and there were no individuals which had *C. oxyacanthus* alleles at all loci. There were six individuals present that contained *C. oxyacanthus* DNA based on a single marker locus, with three of the individuals being homozygous for the *C. oxyacanthus* allele at ct353.

Linkage group 6 (Fig 3.3e, 3.4e) had only a single individual with *C. oxyacanthus* alleles along the entire linkage group. Out of the 13 individuals that contained *C. oxyacanthus* DNA, six were homozygous for at least one marker.

Discussion

Access to a genetic map and markers allowed the successful tracking of introgression of the *C. oxyacanthus* genome into the *C. tinctorius* genome through a generation. The striking physical differences observed in the same species between the two field seasons, revealed the large effect environmental factors can have on the morphology of both parental species. Despite limited morphological differences between plants in 2008, there was variation observed in the introgression of the *C.*

oxyacanthus genome. In fact, every linkage group contained multiple individuals containing introgressed *C. oxyacanthus* DNA.

Morphological Variation:

Large morphological differences were observed in the parental lines between the field seasons of 2007 and 2008, especially in *C. oxyacanthus* (Figs. 3.5 and 3.6). In 2008, individuals of this species were considerably larger than in the previous year, with branches spreading farther from the main stem of the plant. Differences in location, shading, herbivore activity and weather likely contributed to the morphological variation seen in 2007 and 2008. In 2007, no herbivore deterrents were employed and evidence of herbivores was seen on the plants at harvest (plants were missing, others were pulled up, there were fewer heads present on the remaining plants). In contrast, a fence was used to deter herbivores in 2008. Changing field plots between 2007 and 2008 resulted in different surrounding vegetation, mainly in fewer trees overshadowing the plot in 2008. Variation in the weather from one year to the next may also have contributed to plant differences (Mundel et al. 2004; Smith 1996). Environmental factors can cause large changes in the branching and heights of *Carthamus* (Deshpande 1952) and is consistent with environmental influences leading to observed differences in the plants from one year to the next in this study.

Potential for Crop Development

Increased yield is always a goal with crop development programs, and in the case of safflower increase in either oil in a cypsela or number of cypselae would be desirable. However simply increasing the number of cypselae may result in smaller cypselae due to limited room in the capitula. Correlations of cypselae size and weight were examined to insure that an increase in one measurement did not result in a decrease of other cypselae traits. The number and total weight of cypselae were positively correlated with the number of capitula per plant (r^2=0.93 and 0.87 respectively, Appendix 1). Neither the number of capitula or number of cypselae was strongly correlated with the weight per 100 cypselae (r^2=0.1 for both, Appendix 1). The weight per 100 cypselae showed a small positive correlation with both the cypselae length and width (r^2 = 0.57 and 0.62 respectively, Appendix 1), but showed very little correlation with the cypselae per capitula (r^2=-0.15, Appendix 1). This shows that the fatter cypselae were indeed heavier, but heavier cypselae in a head didn't have a large negative effect on the size or number of the cypselae. While there are correlations with some of the cypsela measurements, previous work have shown that the number of heads and the number of cypselae per head is influenced by environmental factors (Smith 1996, pg 48), indicating that the variation seen between lines is due, at least in part, to environmental influences rather than genetic differences. More work is needed to determine the extent of the genetic vs. environmental factors in these important traits.

The presence of plants with large yield potential shows promise for breeding programs. One plant of particular interest in this group was 1.7.8.3, which had large

number of capitula (148, Fig 3.7) and cypsela number (3088, Fig 3.7) when compared to the rest of the individuals (average of 36 and 798 for number of capitula and cypselae; Fig.3.7, Table 3.2). Although the cypselae from this plant were lower on the weight per 100 cypselae (2.3g Fig 3.7), but may have potential for crop development and breeding programs. Also of interest is individual 1.7.36.9, which was high in cypselae number (2190, Fig 3.7) and the weight per 100 cypselae (3.8g, Fig 3.7). More work is needed to see if any of the traits can be mapped to any of the linkage groups to help with marker assisted breeding programs, or finding genes associated with plant yield and cypselae development.

Genetic Variation and Introgression:

All markers with distortion showed a preference for the *C. tinctorius* allele. Distortion of the segregation pattern of alleles was seen for seven markers (Fig. 3.4a-e). These markers were located on linkage groups 1, 3, 5 and 6, with linkage group 5 having three markers with segregation distortion and linkage group 6 containing two distorted markers. Markers ct139 and ct266 had no heterozygous individuals while markers ct 297 and ct32 had no homozygous *C. oxyacanthus* individuals. Marker 697 showed no *C. oxyacanthus* alleles. The lack of grouping of distorted alleles suggests that the distortion is not due to areas with little to no recombination between the two genotypes.

Because all the distortion occurs in favour of the *C. tinctorius* genome, this finding suggests that the *C. oxyacanthus* genome around those markers may cause

some reduced fitness of the gametes, resulting in fewer individuals being produced with *C. oxyacanthus* DNA at that loci. However the most plausible explanation of the observed SD is the small sample size being tested, allowing random error to influence the results. Increasing the sample size may result in the disappearance of distortion for these markers, as they did not have distortion in the previous generation (Mayerhofer et al. 2010).

Introgression of DNA from the wild relative *C. oxyacanthus* into the *C. tinctorius* genome occurs during crosses (Mayerhofer et al. 2010), which allowed an interspecific genetic map to be produced. The RF between markers presented here increased from the original map for 26 regions (Mayerhofer et al. 2010, Fig. 3.4a-e), while four stayed roughly the same. Exceptions were the regions between markers ct381-ct201, ct 495-ct605 and ct598-ct405 (Fig. 3.4e, c, a) which showed a reduction in the RF. These differences in RF values are explained by examination of the parental scoring data for the markers (Fig. 3.4 a-e). For many of the individuals used, scoring data for many markers is incomplete for the parental BC1 plant (Fig. 3.4a-e), making accurate measurement of the RF difficult. With the more complete scoring data for all individuals presented in this work, more accurate RF values could be calculated, allowing a better estimation of the recombination between the genomes of *C. oxyacanthus* and *C. tinctorius*. Recombination frequencies for some markers estimated here may be slightly low due to the presence of masked recombination events (Fig. 3.8).

Previous to this work, in the inter-specific map many primers were seen clustered together with an RF of 0 (Fig 3.3 a-e, Mayerhofer et al. 2010). In the following BC1S1 generation, separation of two markers on linkage group 1 and 6 was observed (ct 531, ct783, Figure 3.7a and ct201, ct3331 Fig. 3.4e). This increase in RF suggests there is a low amount of recombination in some areas. With additional samples or generations, separation of more clustered markers may be possible, giving a more accurate chromosome maps. The ability to have good coverage of an area is key to marker assisted breeding programs, and the placement of relatively close markers on the chromosomes would be an asset.

Chromosome sections mostly moved between the two species in large blocks (Fig 3.4), with large differences in the amount of *C. oxyacanthus* DNA being retained from one linkage group to another. The number of individuals containing any *C. oxyacanthus* DNA varied from 11 (linkage group 3, Fig. 3.4b) to 25 (linkage group 5, Fig. 3.4d). The large differences from one chromosome to another may be due to the placement of the markers along the chromosomes affecting the amount of recombination that will occur, or it may be a result of selection for or against the *C. oxyacanthus* genome at certain areas.

Linkage group 5 is of particular interest in this population. The RF between each of the markers increased along this linkage group (Fig. 3.4d), to the point that markers ct 458 and ct 353 are no longer linked, showing a drastic increase in recombination between the two genomes over the previous generation. In addition, the individuals analyzed have *C. oxyacanthus* alleles interspersed along the

chromosome (Fig. 3.4 d), breaking the trend of the other linkage groups to have the *C. oxyacanthus* occur primarily in groups of close markers. The increased recombination and retention of small grouping or single markers of *C. oxyacanthus* indicates that either the markers were further apart on this chromosome than originally thought, or that the chromosomes of *C. tinctorius* and *C. oxyacanthus* for this linkage group are highly homologous allowing large amounts of recombination to occur. The high levels of recombination present on this linkage group is promising for crop breeding programs, as the incorporation of small amounts of *C. oxyacanthus* DNA is feasible. It is not known however, if any advantageous traits are present on this linkage group and more work will have to be done to see if there are potentially useful genes present.

Linkage group 3 and 6 show the retention of large groups of *C. oxyacanthus* marker, with linkage group 6 containing 4 individuals that have *C. oxyacanthus* alleles at every marker. Each linkage group had minimal number of individuals with single alleles of *C. oxyacanthus* flanked by *C. tinctorius* alleles (1, Linkage group 3 Fig. 3.4b; 2, linkage group 6 Fig. 3.4e). These linkage groups illustrate the relatively small number of recombination events occurring between the two genomes, indicating that at least for the these two linkage groups, there are differences between the chromosomes of the two species sufficient to inhibit recombination.

Conclusions:

The analysis of the BC1S1 lines performed in this study illustrate the ability of *C. oxyacanthus* DNA to be integrated into the *C. tinctorius* genome, the efficiency of which varies across the genome. Morphological variation was seen between all individuals, providing evidence that wild relatives of *C. tinctorius* may be a potential source of variation for breeding programs. The ability of the two genomes to combine, also indicates that gene transfer will occur between these two species, which is concerning for the release of transgenic varieties.

Literature Cited

Ashri A, Efron Y (1964) Inheritance studies with fertile interspecific hybrids of three *Carthamus* L. species. Crop Science 4:510-514.

Ashri A, Knowles PF (1960) Cytogenetics of Safflower (*Carthamus* L.) species and their hybrids. Agronomy Journal 52: 11-17.

Bowles VG, Mayerhofer R, Davis C, Good AG, Hall JC (2010) A phylogenetic investigation of *Carthamus* combining sequence and microsatellite data. Plant Systematics and Evolution (in press)

Chapman MA, Burke JM (2007) DNA sequence diversity and the origin of cultivated safflower (*Carthamus tinctorius* L.; Asteraceae). Bmc Plant Biology 7: 60.

Deshpande RB (1952) Wild safflower (*Carthamus oxyacantha* Bieb.) - a possible oilseed crop for the desert and arid regions. Indian Journal of Genetic Plant Breeding 12: 10-14.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539-563.

Ellstrand NC and Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? Proceedings of the National Acadamy of Science 97:7043-7050.

Hanelt P (1963) Monographische Ubersicht der Gattung *Carthamus* L. (Compositae). Feddes Repertorium 67: 41-180.

Imrie BC, Knowles PF (1970) Inheritance studies in interspecific hybrids between *Carthamus flavescens* and *C. tinctorius*. Crop Science 10: 349-352.

Laughnan JR and Gabay-Laughnan S (1983) Cytoplasmic male sterility in maize. Annual Review in Genetics 17:27-48.

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

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

McPherson MA, Good AG, Topinka AKC, Hall LM (2004) Theoretical hybridization potential of transgenic safflower (*Carthamus tinctorius* L.) with weedy relatives in the New World. Canadian Journal of Plant Science 84: 923-934.

Mündel HH, Blacksaw RE, Byers JR, Huang HC, Johnson DL, Keon R, Kubik J, McKenzie R, Otto B, Roth B, Stanford K (2004) Safflower production on the Canadian prairies: revisited in 2004. Agriculture and Agri-Food Canada, Lethbridge Research Center, Alberta. Pg 1.

Okogbenin E, Porto MCM, Egesi C, Mba C, Espinosa E, Santos LG, Ospina C, Marin J, Barrera E, Gutierrex J, Ekanayake I, Iglesias C, Fregene MA (2007) Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. Crop

Science 47:1895-1904.

Ramanurthy GV (1964) Relationships of Cultivated Safflower (*Carthamus tinctorius* L.) to the wild species *C. oxyacanthus*. Thesis, Doctor of Philosophy, University of California, Davis.

Rieseberg (2009) Evolution: replacing genes and traits through hybridization. Current Biology 19: 119 – 122.

Sanchez Olguin ER, Espinoza-Esquivel A, Lobo JA, Arrieta-Espinoza G (2009) Assessment of gene flow from a herbicide-resistant indica rice (*Oryza sativa* L.) to the Costa Rican weedy rice (*Oryza sativa*) in Tropical America: Factors affecting hybridization rates and characterization of F1 hybrids. Transgenic Research 18:633-647.

Seefeldt SS, Zementra R, Young FL, Jones SS (1998) Production of herbicide-resistant jointed goatgrass (*Aegilops cylindrical*) x wheat (*Triticum aestivum*) hybrids in the field by natural hybridization. Weed Science 46:632-634.

Smith JR (1996) Safflower. AOCS Press, United States of America. Pg. 2-4

Vilatersana R, Garnatje T, Susanna A, Garcia-Jacas N (2005) Taxonomic problems in *Carthamus* (Asteraceae): RAPD markers and sectional classification. Botanical Journal of the Linnean Society 147: 375-383.

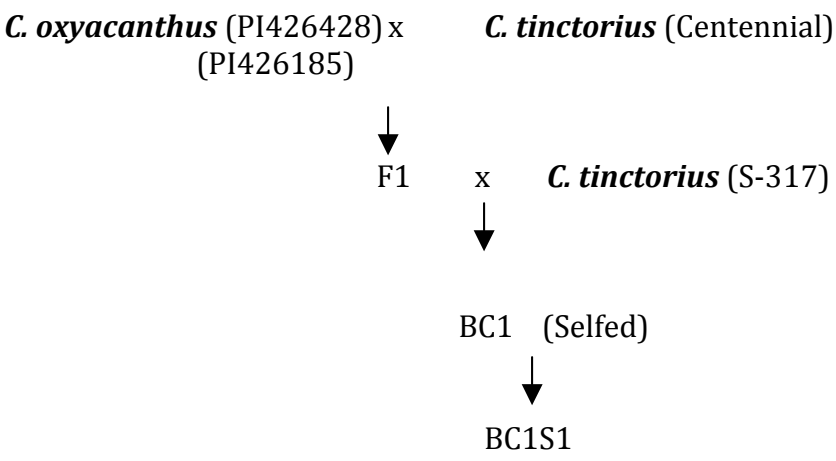
Vilatersana R, Susanna A, Garcia-Jacas N, Garnatje T (2000) Generic delimitation and phylogeny of the *Carduncellus-Carthamus* complex (Asteraceae) based on ITS sequences. Plant Systematics and Evolution 221: 89-105.

Warwick SI, Legere A, Simard MJ, James T (2003) Do escaped transgenes persist

in nature? The case of an herbicide resistance transgene in a weedy *Brassica rapa* population. Molecular Ecology 17:1387-1395.

Fig. 3.1 Crossing Scheme and Naming system

a. Crossing Scheme for the production of the interspecific (*C. oxyacanthus, C. tinctorius*) selfed backcross generation (BC1S1).



b. Plant naming system for the individual plants used in the production of the selfed backcross generation (BC1S1).

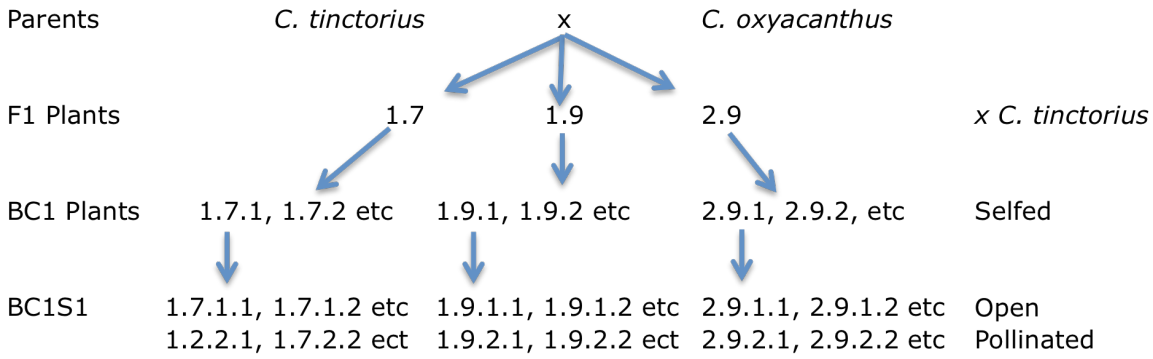


Fig. 3.2 Diagram for the measurement of height, width and length of Cypsela.

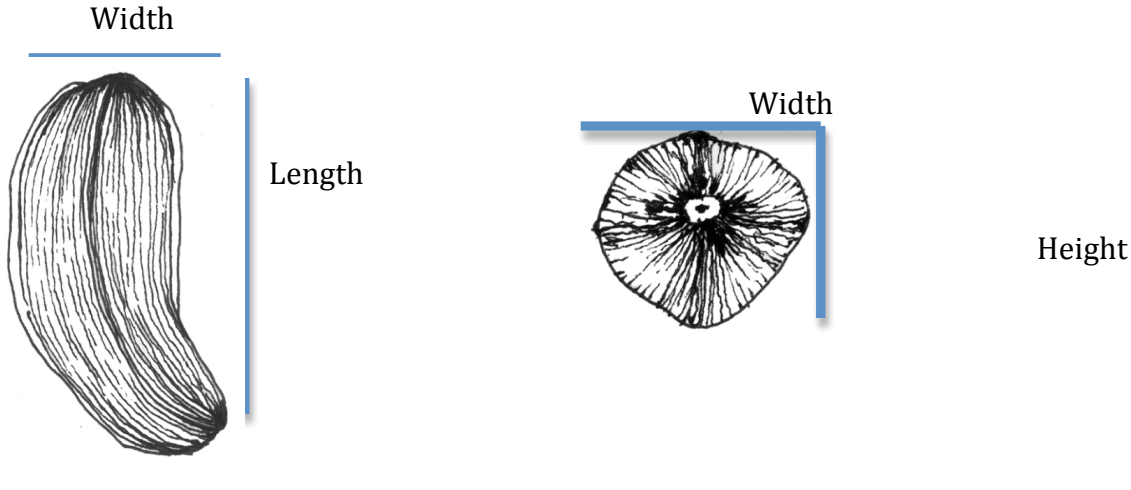
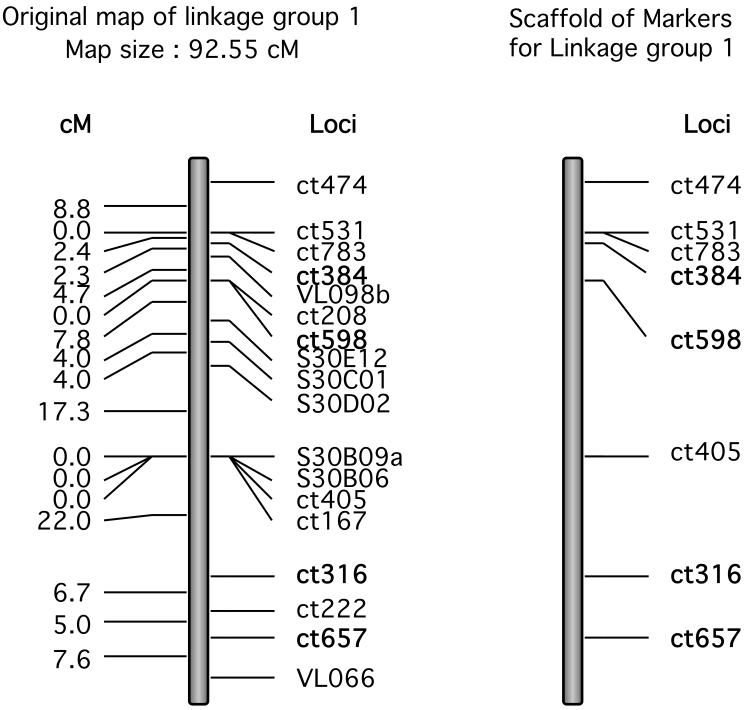
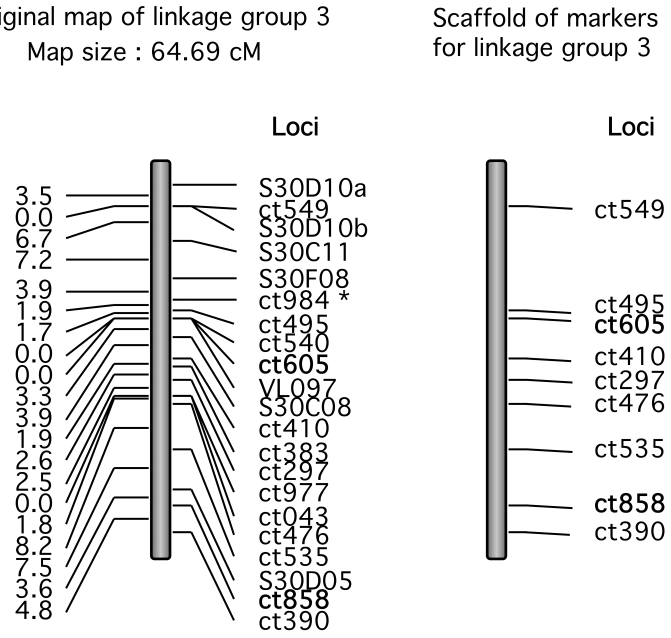


Figure 3.3 a-e. Linkage group maps and scaffolds for *C. oxyacanthus*.

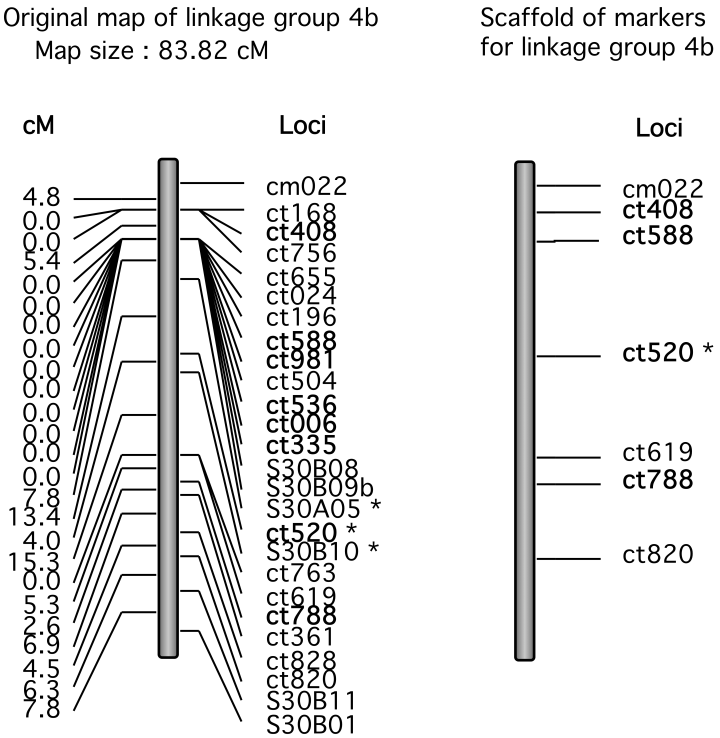
Maps of the linkage groups were adapted from Mayerhofer et al (2010). Scaffolds for marker scoring were adapted from the original maps, showing only the marker scored for the BC1S1 generation.

a.

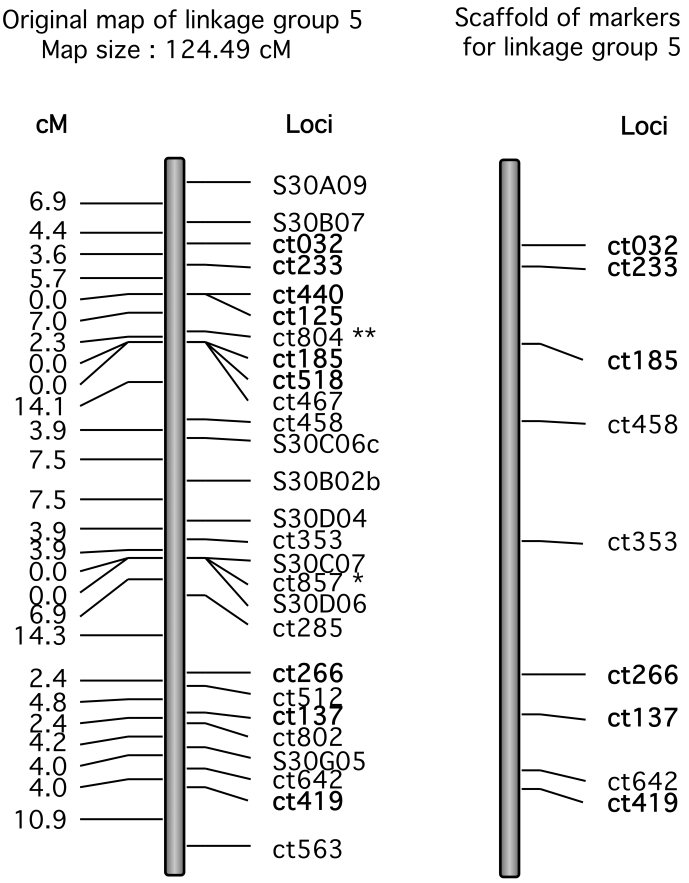



Original map of linkage group 3
Map size : 64.69 cM
Scaffold of markers
for linkage group 3
Loci
Loci
3.5
0.0
6.7
7.2
3.9
1.9
1.7
0.0
0.0
3.3
3.9
1.9
2.6
2.5
0.0
1.8
8.2
7.5
3.6
4.8
S30D10a
ct549
S30D10b
S30C11
S30F08
ct984 *
ct495
ct540
ct605
VL097
S30C08
ct410
ct383
ct297
ct977
ct043
ct476
ct535
S30D05
ct858
ct390
ct549
ct495
ct605
ct410
ct297
ct476
ct535
ct858
ct390

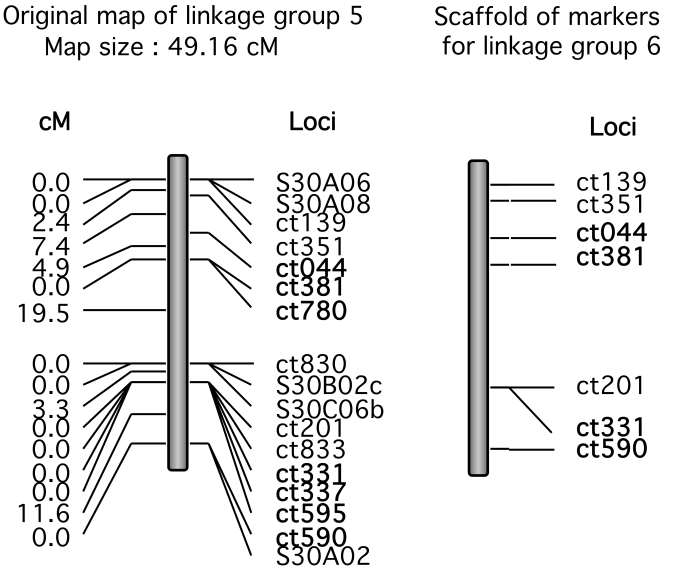
b.


Original map of linkage group 4b
Map size : 83.82 cM
Scaffold of markers
for linkage group 4b
cM
Loci
Loci
4.8
0.0
0.0
5.4
0.0
0.0
0.0
0.0
0.0
0.0
0.0
0.0
0.0
7.8
13.4
4.0
15.3
0.0
5.3
2.6
6.9
4.5
6.3
7.8
cm022
ct168
ct408
ct756
ct655
ct024
ct196
ct588
ct981
ct504
ct536
ct006
ct335
S30B08
S30B09b
S30A05 *
ct520 *
S30B10 *
ct763
ct619
ct788
ct361
ct828
ct820
S30B11
S30B01
cm022
ct408
ct588
ct520 *
ct619
ct788
ct820

c.


Original map of linkage group 5
Map size : 124.49 cM
Scaffold of markers
for linkage group 5
cM
Loci
Loci
6.9
4.4
3.6
5.7
0.0
7.0
2.3
0.0
0.0
14.1
3.9
7.5
7.5
3.9
3.3
3.9
0.0
0.0
6.9
14.3
2.4
4.8
2.4
4.2
4.0
4.0
10.9
S30A09
S30B07
ct032
ct233
ct440
ct125
ct804 **
ct185
ct518
ct467
ct458
S30C06c
S30B02b
S30D04
ct353
S30C07
ct857 *
S30D06
ct285
ct266
ct512
ct137
ct802
S30G05
ct642
ct419
ct563
ct032
ct233
ct185
ct458
ct353
ct266
ct137
ct642
ct419

d.


Original map of linkage group 5
Map size : 49.16 cM
Scaffold of markers
for linkage group 6
cM
Loci
Loci
0.0
0.0
2.4
7.4
4.9
0.0
19.5
0.0
0.0
3.3
0.0
0.0
0.0
0.0
11.6
0.0
S30A06
S30A08
ct139
ct351
ct044
ct381
ct780
ct830
S30B02c
S30C06b
ct201
ct833
ct331
ct337
ct595
ct590
S30A02
ct139
ct351
ct044
ct381
ct201
ct331
ct590

e.

Figure 3.4 a-e Scoring data for linkage groups of the BC1 and BC1S1 generations of the *C. oxyacanthus* x. *C. tinctorius* interspecific cross.

Scoring data for the parent BC1 population is seen on the top, with recombination frequencies for each set of markers (Done by Cathy Archibald). The bottom matrix is the BC1S1 data, with corresponding recombination frequencies calculated for the new scoring data. Bolded markers were able to distinguish between the two genotypes of *C. tinctorius* (Centennial and S317). Chi squared values for segregation distortion are shown, with bolded values indicating significant deviation. RF (recombination frequency) is given between each set of adjacent markers.

Figure 3.4a Linkage group 1

Plant F1, BC1		2.9	1.7	1.9	2.9	2.9	1.7	1.7	1.7	2.9	2.9	1.7	2.9	2.9	2.9	1.7	1.9	2.9	1.9	2.9	1.7	1.9	2.9	2.9	1.7	2.9
RF	**Marker**	**12**	**11**	**10**	**4**	**55**	**35**	**16**	**32**	**26**	**49**	**16**	**37**	**36**	**33**	**11**	**2**	**54**	**8**	**35**	**16**	**47**	**15**	**39**	**44**	**5**	**12**	**11**	**10**	**1**	**21**	**57**
8.8	474	A	A	-	A	A	H	A	A	A	-	-	H	A	A	A	A	-	A	-	H	-	A	H	-	-	H	-	-	-	H	H
	531	A	-	A	A	A	H	A	A	A	-	H	H	A	A	A	H	-	A	-	-	-	A	H	A	-	-	-	H	-	H	H
0	783	A	H	A	A	A	H	A	A	A	A	H	H	A	A	A	H	A	A	H	A	A	A	H	A	H	A	A	H	A	H	H
2.4	384	-	H	-	-	-	H	A	A	A	-	-	H	A	-	-	-	-	A	-	-	-	-	H	-	-	A	-	-	-	-	-
7	598	-	H	-		-	H	-	-	A	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	A	-	-	-	H	-
33.1	405	H	H	H	A	H	H	A	A	A	-	H	H	A	A	A	H	A	H	A	A	A	A	H	A	H	A	A	H	H	H	H
27	316	-	H	-	-	-	H	A	A	A	-	-	-	A	-	-	-	-	H	-	-	-	-	-	-	-	A	-	-	-	A	-
11.7	657	-	H	-	-	-	A	A	-	A	-	-	H	-	-	-	-	-	H	-	-	-	-	H	-	-	A	-	-	-	A	-
Plant BC1S1		**6**	**6**	**8**	**4**	**9**	**2**	**8**	**9**	**2**	**11**	**4**	**9**	**9**	**2**	**4**	**2**	**14**	**3**	**9**	**8**	**3**	**9**	**13**	**8**	**3**	**2**	**14**	**3**	**2**	**8**	**5**
	474	A	A	A	A	A	H	A	A	A	A	A	H	A	A	A	A	A	A	H	A	A	A	A	H	H	B	A	H	A	H	A
10	531	A	A	A	A	A	H	A	A	A	A	H	H	A	A	A	-	A	A	H	A	A	A	A	A	H	A	A	H	A	H	A
3.3	783	A	A	A	A	A	H	A	A	A	A	H	H	A	A	A	B	A	A	B	A	A	A	A	A	H	A	A	H	A	H	A
3.2	384	A	A	A	A	A	H	A	A	A	A	H	H	A	A	A	B	A	A	H	A	A	A	A	A	H	A	A	H	A	H	A
12.9	**598**	B	A	B	A	A	H	A	A	A	A	H	H	A	A	A	B	A	A	B	A	A	A	A	A	H	A	A	H	A	H	A
13.3	405		A	B	A	H	A	A	A	A	A	H	H	A	A	A	B	A	H	A	A	A	A	H	A	H	A	A	H	A	H	A
24.1	**316**	A	A	B	A	H	A	A	A	A	A	A	B	A	A	A	B	A	B	A	A	A	-	B	A	H	A	A	B	H	A	A
30	657*	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Figure 3.4b. Linkage group 3

Plant F1		2.9	1.7	1.9	2.9	2.9	1.7	1.7	1.7	2.9	1.7	2.9	2.9	2.9	1.7	2.9	1.9	2.9	1.9	1.7	1.9	2.9	2.9	1.7	2.9
Plant BC1		12	11	10	4	55	35	16	32	26	49	16	37	36	33	11	2	54	8	35	16	47	15	39	44	5	12	11	10	1	21	57
RF	Marker																															
19.7	549	H	H	-	A	A	H	H	A	H	-	A	A	A	A	A	A	A	H	A	A	-	H	H	H	H	A	-	A	H	A	A
	495	H	H	H	A	A	H	H	A	H	-	A	A	A	A	A	A	-	H	A	A	-	H	H	H	A	A	H	A	H	A	A
1.7	605	-	H	-	-	-	H	H	A	H	-	-	A	A	-	-	-	-	H	-	-	-	-	H	-	-	A	-	-	-	A	-
7.2	410	A	H	H	A	-	H	A	A	H	-	A	A	A	A	A	-	-	H	-	A	-	H	H	H	H	A	H	A	H	A	A
4.5	297	-	-	-	-	-	-	A	A	H	-	-	-	-	-	-	-	-	H	-	-	-	-	H	-	-	A	-	-	-	A	-
4.3	476	A	H	H	A	H	H	A	A	H	H	A	A	A	A	A	A	A	H	A	A	A	H	H	H	H	A	H	A	H	A	A
8.2	535	-	H	-	-	-	H	A	A	H	-	-	-	-	-	-	-	-	H	-	-	-	-	H	-	-	A	-	-	-	A	-
11.1	858	-	H	-	-	-	H	A	A	A	-	-	A	H	-	-	-	-	H	-	-	-	-	A	-	-	A	-	-	-	A	-
4.8	390	A	H	A	A	H	H	A	A	A	**-**	A	A	H	A	A	A	-	H	A	A	A	H	A	-	H	A	H	A	H	A	A
Plant BC1S1		**6**	**6**	**8**	**4**	**9**	**2**	**8**	**9**	**2**	**11**	**4**	**9**	**9**	**2**	**4**	**2**	**14**	**3**	**9**	**8**	**3**	**9**	**13**	**8**	**3**	**2**	**14**	**3**	**2**	**8**	**5**
3.3	549	B	H	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	H	H	A	A	A	H	A	A
	495	-	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	H	H	A	A	A	H	A	A
0	605	H	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	A	A	A	H	A	A
12	410	A	H	H	A	A	H	A	A	A	-	A	A	-	A	A	A	A	H	-	A	-	A	A	-	H	-	A	A	A	A	-
4	297*	A	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	A	A	A	A	A	A
4.2	476	A	H	H	A	A	H	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	-	-	-	-	-	-	-	-
13	**535**	A	H	A	A	A	H	A	A	-	H	A	A	A	A	A	A	B	H	A	A	A	-	A	H	H	A	A	A	A	A	A
6.7	**858**	A	H	A	A	A	H	A	A	A	H	A	A	H	A	A	A	B	B	A	A	A	A	A	H	H	A	A	A	A	A	A
6.3	**390**	A	H	A	A	A	H	A	A	A	H	A	A	H	A	A	A	H	H	A	A	A	A	A	H	H	A	A	A	A	A	A

Figure 3.4c Linkage group 4b

Plant F1, BC1		2.9	1.7	1.9	2.9	2.9	1.7	1.7	1.7	1.7	2.9	1.7	2.9	2.9	2.9	2.9	1.7	1.9	2.9	1.9	1.7	1.9	2.9	2.9	1.7	2.9
RF	Marker	12	11	10	4	55	35	16	32	26	49	36	33	11	2	54	8	16	47	15	39	44	5	12	11	10	1	21	57
4.8	22	A	H	H	A	A	-	H	A	H	H	A	A	H	H	H	A	H	A	H	H	A	H	A	H	A	A	A	A
	408	-	H	-	-	-	H	H	A	H	-	-	-	-	-	-	A	-	-	-	H	-	-	A	-	-	-	A	-
5.4	588	-	A	-	-	-	H	H	A	H	-	A	A	-	-	-	A	-	-	-	-	-	-	A	-	-	-	A	-
21.2	520	-	A	-	-	-	A	-	A	H	-	H	A	-	-	-	A	-	-	-	A	-	-	A	-	-	-	A	-
19.3	619	A	A	H	A	H	A	A	A	H	H	H	H	A	H	H	A	H	H	H	A	A	A	A	A	A	A	A	H
5.3	788	-	A	-	-	-	-	A	A	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	-
11.4	820	A	A	H	A	A	H	-	A	-	-	A	-	-	H	H	A	H	H	A	A	H	A	A	A	A	A	H	H
Plant BC1S1		**6**	**6**	**8**	**4**	**9**	**2**	**8**	**9**	**2**	**11**	**4**	**2**	**4**	**2**	**14**	**3**	**8**	**3**	**9**	**13**	**8**	**3**	**2**	**14**	**3**	**2**	**8**	**5**
6.7	22	A	H	H	A	H	A	A	A	A	B	A	A	H	A	B	A	H	A	H	H	A	A	A	A	A	A	A	A
10.3	408	A	H	H	A	H	A	A	A	A	B	A	A	H	A	B	A	H	A	H	B	A	A	A	H	A	A	A	-
34.5	**588**	A	A	H	A	H	A	H	A	A	B	A	A	H	H	B	A	H	A	H	-	A	A	A	H	A	A	A	A
30	**520**	A	A	A	A	H	A	H	A	A	B	A	A	H	A	A	A	B	H	H	A	A	A	A	A	-	A	A	A
30	619	A	A	H	A	H	A	A	A	A	H	H	A	A	B	A	A	A	H	H	B	A	A	A	A	A	A	A	H
20	**788**	A	A	H	A	H	A	A	A	A	H	B	A	A	H	A	A	B	-	A	A	A	A	A	A	A	A	B	H
	820	A	A	H	A	H	H	A	A	A	H	H	A	A	H	A	A	A	B	A	A	B	A	A	A	A	A	H	H

Figure 3.4d Linkage group 5

Plant F1, BC1	2.9	1.7	1.9	2.9	2.9	1.7	1.7	1.7	2.9	2.9	1.7	1.7	2.9	2.9	2.9	2.9	1.7	1.9	2.9	1.7	2.9	1.9	1.7	1.9	2.9	2.9	1.7	2.9
RF	Marker	12	11	10	4	55	35	16	32	26	49	16	37	36	33	11	2	54	8	35	16	47	15	39	44	5	12	11	10	1	21	57
3.6	32	-	A	-	-	-	-	A	A	A	-	-	A	H	-	-	-	-	-	-	-	-	-	A	-	-	H	-	-	-	A	-
	233	-	A	-	-	-	A	A	-	A	-	-	A	H	-	-	-	-	A	-	-	-	-	A	-	-	H	-	-	-	A	-
15	185	-	A	-	-	-	A	A	A	A	-	-	A	H	-	-	-	-	A	-	-	-	-	A	-	-	H	-	-	-	A	-
14.1	458	A	A	H	A	A	A	A	A	A	H	A	-	-	A	H	H	H	A	-	-	H	H	A	H	-	H	A	H	H	A	H
22.8	353	H	A	H	A	A	A	A	A	A	-	H	A	H	A	H	H	-	A	A	A	A	H	A	H	H	H	A	H	H	A	H
25.1	266	-	A	-	-	-	-	A	H	A	-	-	A	H	-	-	-	-	H	-	-	-	-	H	-	-	A	-	-	-	A	-
7.2	137	-	H	-	-	-	A	A	H	A	-	-	A	H	-	-	-	-	H	-	-	-	-	H	-	-	A	-	-	-	A	-
10.6	642	-	H	-	-	-	H	A	H	A	-	-	-	A	-	-	-	-	H	-	-	-	-	H	-	-	A	-	-	-	A	-
4	419	-	-	-	-	-	H	-	H	A	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-
Plant	**BC1S1**	**6**	**6**	**8**	**4**	**9**	**2**	**8**	**9**	**2**	**11**	**4**	**9**	**9**	**2**	**4**	**2**	**14**	**3**	**9**	**8**	**3**	**9**	**13**	**8**	**3**	**2**	**14**	**3**	**2**	**8**	**5**
40	32	A	A	A	A	A	A	A	A	A	A	A	A	H	A	A	A	A	A	A	A	A	A	A	A	B	H	A	A	A	A	A
	233	H	A	A	A	A	A	A	A	A	H	H	A	H	A	A	H	B	A	A	-	B	B	A	H	A	H	A	B	H	A	B
30	185	A	A	H	A	A	A	A	A	A	H	A	A	H	A	H	B	A	A	A	A	B	B	A	H	A	H	A	A	A	A	A
19.4	458	A	A	A	A	A	A	A	A	A	H	A	A	H	A	A	B	B	A	A	A	B	B	A	H	A	H	A	B	H	A	B
56.7	353	H	-	A	A	A	A	A	A	A	B	A	A	A	A	H	A	A	H	B	A	A	A	B	B	A	B	B	H	A	A	A
36.7	266	A	A	A	A	A	A	A	A	A	B	A	A	A	A	A	B	A	B	A	A	A	A	A	A	B	A	A	A	A	A	B
25.8	137	A	B	H	A	A	A	A	H	A	B	H	A	H	A	A	B	H	B	A	A	A	A	A	A	B	A	A	A	H	A	B
25	**642**	A	B	A	A	A	A	H	A	A	-	A	A	A	A	A	A	A	-	A	A	A	A	-	A	B	A	A	A	A	A	B
22.2	419	A	B	H	A	A	H	A	H	A	B	H	H	A	A	A	A	A	B	A	A	A	A	-	A	B	A	A	A	A	A	B

Figure 3.4e Linkage group 6

Plant F1, BC1		2.9	1.7	1.9	2.9	2.9	1.7	1.7	1.7	2.9	2.9	1.7	1.7	2.9	2.9	2.9	1.7	2.9	1.9	2.9	1.9	1.7	1.9	2.9	2.9	1.7	2.9
RF	Marker	12	11	10	4	55	35	16	32	26	49	16	37	36	33	11	2	54	8	35	16	47	15	39	44	5	12	11	10	1	21	57
2.4	139	-	A	-	-	-	A	H	A	A	-	-	A	H	-	-	-	-	H	-	-	-	-	A	-	-	A	-	-	-	A	-
	351	A	A	A	H	H	A	H	A	A	A	A	H	H	A	A	A	A	H	H	A	A	H	A	A	H	A	A	A	H	A	A
7.4	44	-	A	-	-	-	A	H	A	H	-	-	A	H	-	-	-	-	H	-	-	-	-	A	-	-	A	-	-	-	H	-
4.9	381	-	A	-	-	-	A	H	A	A	-	-	A	A	-	-	-	-	H	-	-	-	-	A	-	-	A	-	-	-	A	-
21.8	201	-	A	-	-	-	H	H	A	A	-	-	A	A	-	-	-	-	H	-	-	-	-	A	-	-	A	-	-	-	A	-
0	331	-	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	A	-
11.6	590	-	H	-	-	-	H	-	A	A	-	-	H	A	-	-	-	-	H	-	-	-	-	H	-	-	A	-	-	-	A	-
Plant BC1S1		**6**	**6**	**8**	**4**	**9**	**2**	**8**	**9**	**2**	**11**	**4**	**9**	**9**	**2**	**4**	**2**	**14**	**3**	**9**	**8**	**3**	**9**	**13**	**8**	**3**	**2**	**14**	**3**	**2**	**8**	**5**
20	139*	A	A	A	A	B	A	H	A	A	B	A	A	A	A	A	A	A	A	B	A	A	B	A	B	A	A	A	A	A	A	A
	351	A	A	A	H	B	A	H	A	A	B	A	A	A	A	B	H	A	A	B	A	A	B	A	A	H	A	A	A	H	-	-
16	**44**	A	-	A	H	A	A	H	-	A	A	A	A	A	A	B	A	A	H	B	A	A	-	-	A	H	A	A	A	H	A	A
5	**381***	A	A	A	H	A	A	H	A	A	A	A	A	A	A	B	A	A	A	B	A	A	A	A	-	-	-	-	-	-	-	-
4.3	201	A	A	A	H	A	B	H	A	A	A	A	A	A	A	B	A	A	A	B	A	A	A	A	-	H	A	A	A	H	A	A
4	331	A	A	A	H	A	B	H	A	A	A	A	A	A	A	B	A	H	A	B	A	A	A	-	-	-	-	-	-	-	-	-
9	590	A	A	A	H	A	H	H	A	A	A	A	A	A	A	H	A	H	A	H	A	A	A	-	A	H	A	A	A	A	A	A

Fig.3.5 Morphology of parental genotypes from field grow out, summer 2007

C. tinctorius Centennial

C. tinctorius S317

C. oxyacanthus

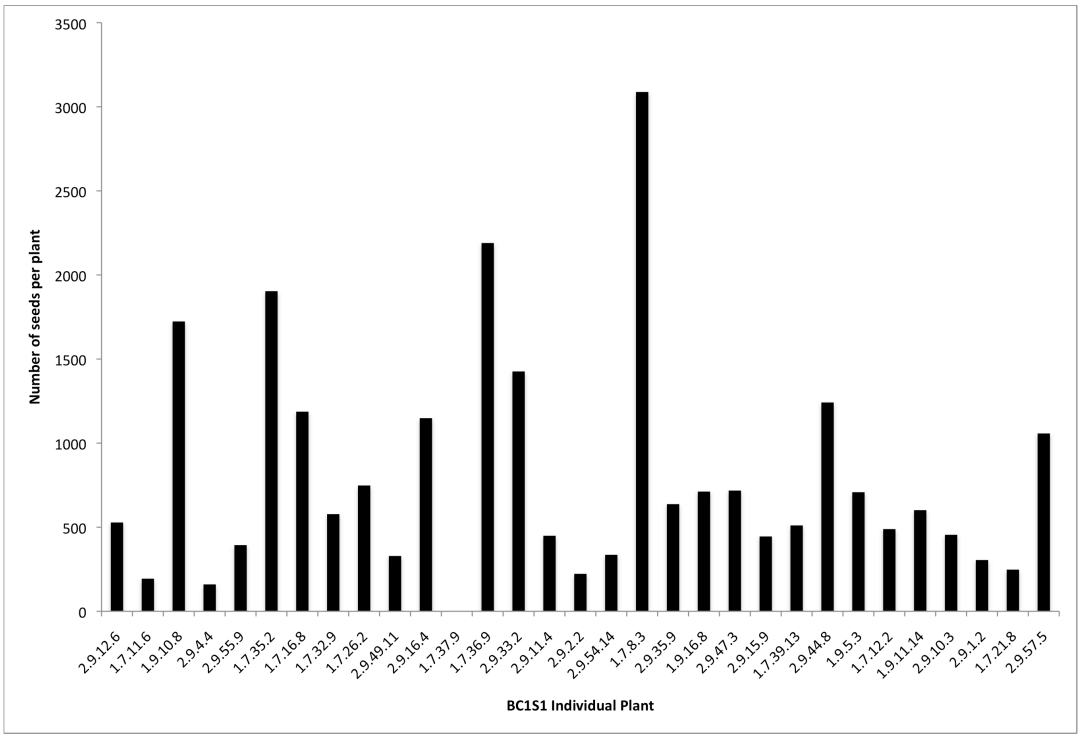
Fig. 3.6 Field Season 2008

a. 

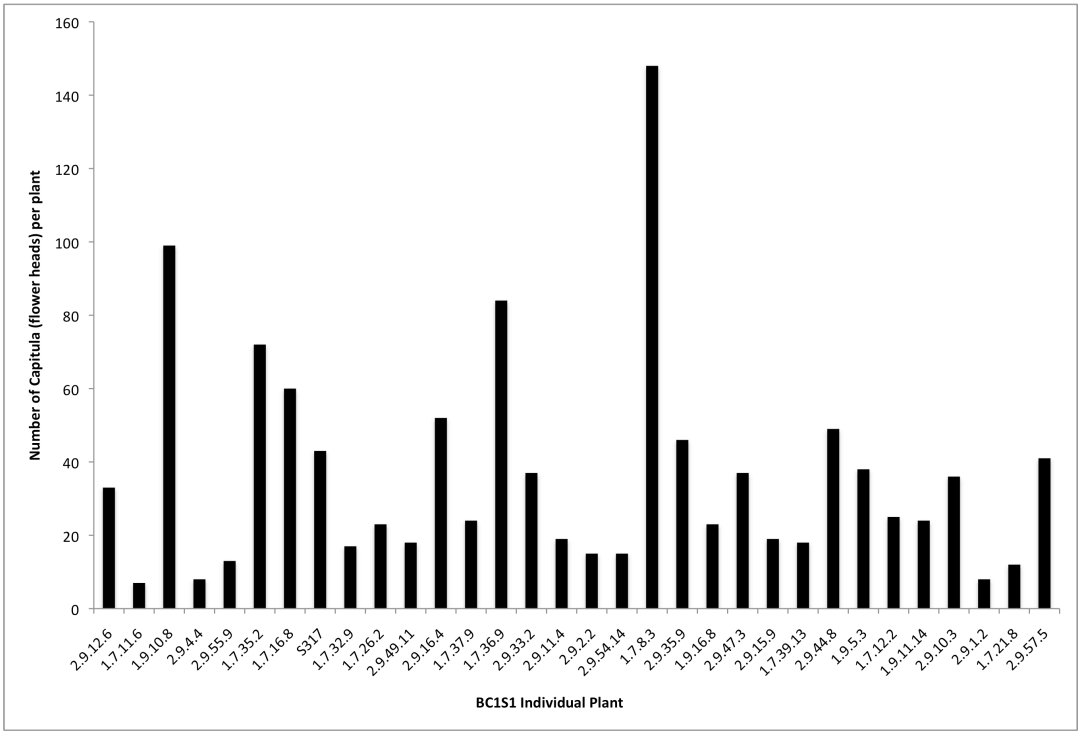
b. 

BC1S1 and parental lines prior to and during flowering for the field season in the summer of 2008. Photograph a shows rows 7 and 8, with the remaining rows in the background from the north end of the plot facing south. Taken prior to height measurements or flowering. Photograph b shows 1 and 2, from the south end of the plot facing north. After height measurements were taken, during flowering.

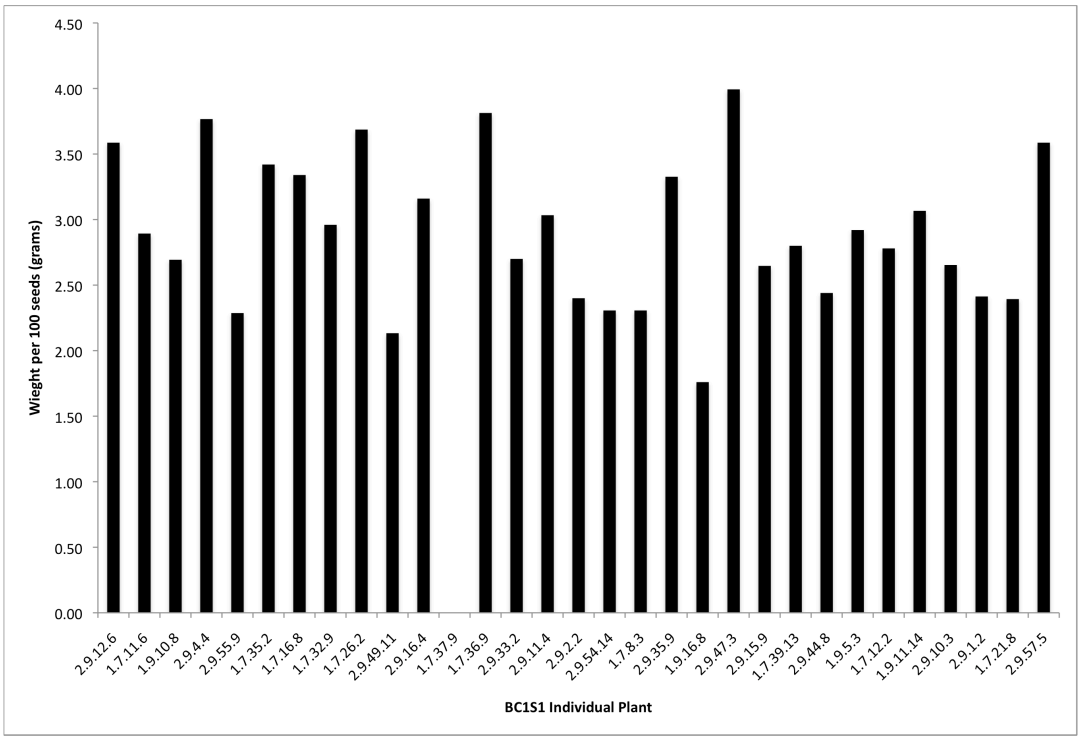
Fig. 3.7 Phenotypic measurements of BC1S1 plants



a. The number of cypselaes present for each individual BC1S1 plant, estimated by the total weight and weight per 100 cypselae.



b. The number of capitula present for each individual plant, as counted at the time of threshing.



c. The weight per 100 cypselae for each individual plant, an average of 3 weights of randomly counted 100 cypselae.

Figure 3.8 – Masked recombination events in the BC1S1 generation of the *C. oxyacanthus* x. *C. tinctorius* interspecific cross.

"Masked" Recombination

Genotype: T H H T 0

Possible Chromosome Structures:

T	O	O	T	O	
T	T	T	T	O	
		OR			
					Hidden
T	T	O	T	O	Recombination
T	O	T	T	O	Event

Genotype of the individual scored as T= homozygous *C. tinctorius*, O= homozygous *C. oxyacanthus*, H = heterozygous (one allele from *C. oxyacanthus*, one allele from *C. tinctorius*). Alleles on the chromosome structures are represented as T = *C. tinctorius* and O = *C. oxyacanthus.*

Table 3.1: Raw measurements of lines from field season 2008.

	Number of True Leaves		Emergence Stem Height		Flowering Plant Stem Height	
Line	Range	Average	Range (mm)	Average (mm)	Range (cm)	Average (cm)
2.9.12	3 to 8	5, n=15	3 to 21	11.33, n=15	55 to 77	72.29
1.7.11	1 to 9	4.95, n=20	1 to 20	9.45, n=20	55.5 to 80	67.59
1.9.10	2 to 6	3.33, n=15	0 to 14	8.27, n=15	65 to 90	NA*
2.9.4	2 to 7	4.22, n=18	0 to 20	9.67, n=18	50 to 75	61.75
2.9.55	1 to 10	5, n=17	0 to 21	6.71, n=17	55 to 75	NA*
1.7.35	3 to 10	5.5, n=8	0 to 20	8.5, n=8	52 to 70	59.75
1.7.16	2 to 7	4.3, n=20	1 to 22	10.9, n=20	55 to 74	NA*
1.7.32	1 to 9	4.71, n=21	0 to 18	7.43, n=21	53 to 75	62.58
1.7.26	0 to 8	4.14, n=22	0 to 20	8.36, n=22	70 to 80	65.50
2.9.49	2 to 6	3.5, n=20	0 to 33	7.4, n=20	54 to 78	66.55
2.9.16	0 to 7	4.3, n=20	0 to 21	9.45, n=20	58 to 70	NA*
1.7.37	2 to 9	4.71, n=24	8 to 23	14.04, n=24	31 to 75	54.13
1.7.36	1 to 11	6.18, n=17	0 to 16	5.29, n=17	63 to 80	NA*
2.9.33	1 to 8	4, n=12	0 to 19	12.83, n=12	60 to 65	63.63
2.9.11	3 to 8	5, n=17	3 to 21	12.88, n=17	64 to 77	69.38
2.9.2	2 to 5	3.1, n=10	2 to 16	9.4, n=10	59 to 65	60.60
2.9.54	3 to 8	4.64, n=28	0 to 14	6.68, n=28	47.5 to 70	58.83

1.7.8	1 to 10	4, n=9	0 to 18	6, n=9	65 to 70	NA*
2.9.35	3 to 8	5.21, n=19	0 to 20	10.11, n=19	60 to 67	NA*
1.9.16	0 to 9	4.3, n=10	0 to 22	5, n=10	65 to 70	NA*
2.9.47	2 to 12	6.39, n=18	0 to 36	14.06, n=18	65 to 70	62.15
2.9.15	0 to 7	3.91, n=22	0 to 13	4.59, n=22	65 to 75	NA*
1.7.39	3 to 6	4.32, n=25	0 to 25	10.64, n=25	61 to 78.5	69.42
2.9.44	2 to 7	4.58, n=26	0 to 18	7.69, n=26	35 to 66	50.67
1.9.5	2 to 6	3.5, n=12	0 to 12	4.58, n=12	54 to 75	NA*
1.7.12	2 to 9	6.15, n=20	9 to 26	17.45, n=20	60 to 70	65.78
1.9.11	2 to 6	3.96, n=26	0 to 21	9.69, n=26	52 to 70	59.73
2.9.10	3 to 6	4.82, n=22	8 to 20	13.14, n=22	34 to 75	61.78
2.9.1	0 to 6	2.92, n=13	0 to 19	7.83, n=13	49 to 70	59.17
1.7.21	0 to 7	4.57, n=23	0 to 31	13.57, n=23	60 to 70	NA*
2.9.57	0 to 12	5, n=13	0 to 25	11.08, n=13	34 to 80	53.33
S317-9	2 to 6	4.12, n=17	1 to 24	8.18, n=17	51 to 71	65.50
S317-7	0 to 6	3.3, n=10	1 to 27	11.3, n=10	65 to 72	68.25
Cent	3 to 9	4.44, n=9	0 to 21	13.22, n=9	60 to 67	64.40
Oxy	4 to 6	5, n=12	0 to 10	2.92, n=12	65 to 70	NA*

Ranges are given for all plants in a line. Averages are given were possible for all plants in a line, n being the number of individuals in the line. *Averages were not available due to difficulties in measuring and time constraints, only the tallest and shortest plants of the line were measured to give a range.

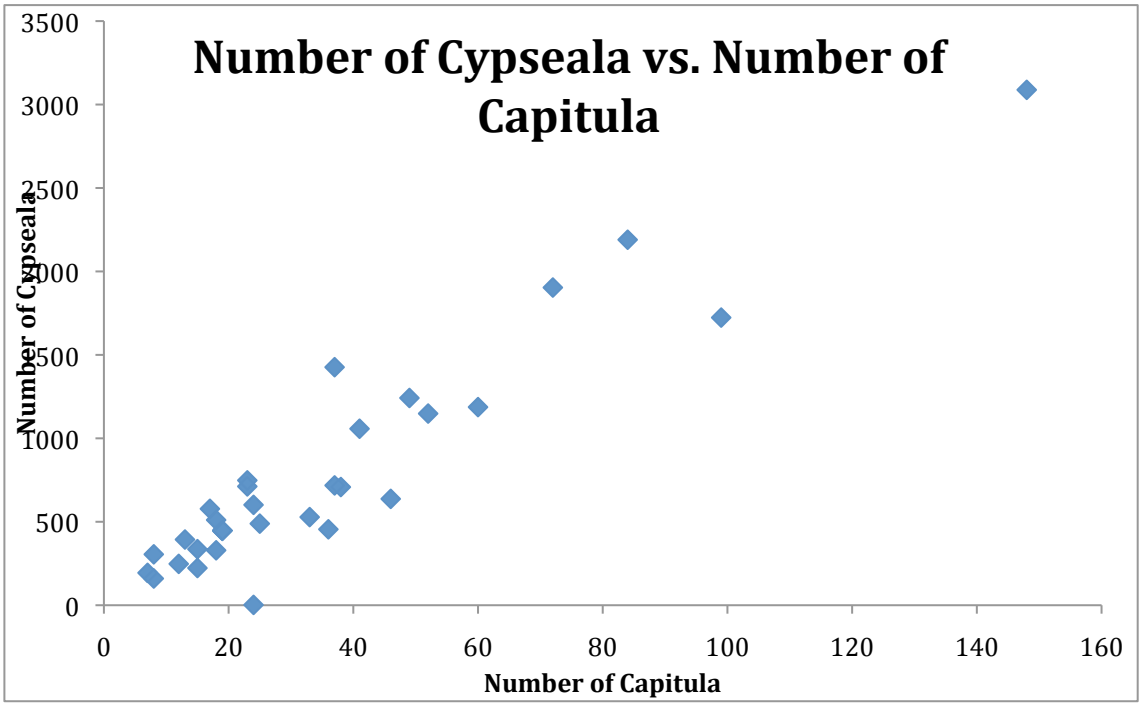
Table 3.2: Raw data of cypselae measurement for the 31 randomly chosen BC1S1 plants.

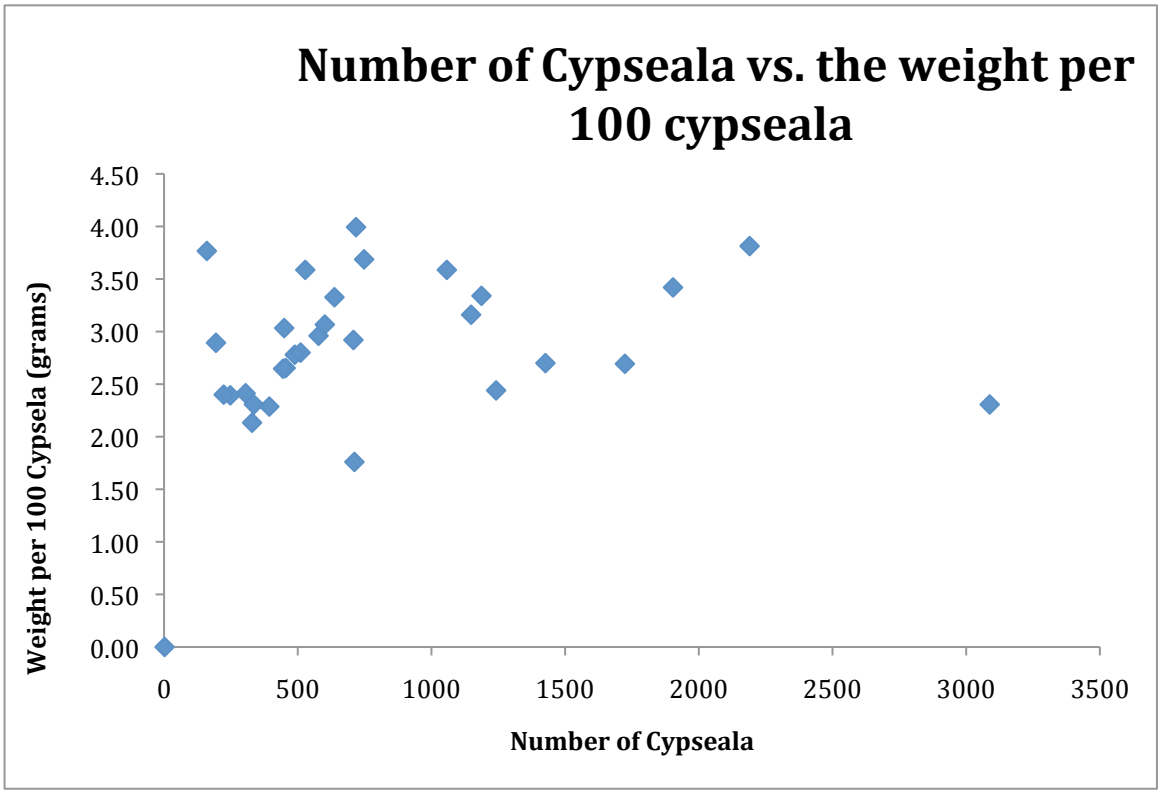
Plant	Number of Capitula	Number of Cypsela	Total weight of Cypsela	weight per 100 Cypsela	Cypsela height	Cypsela width	Cypsela length
2.9.12.6	33	528	18.94	3.59	0.8242	0.3858	0.3268
1.7.11.6	7	194	5.26	2.89	0.7836	0.3498	0.2892
1.9.10.8	99	1724	46.42	2.69	0.6852	0.3586	0.2974
2.9.4.4	8	160	5.96	3.77	0.8517	0.3678	0.3420
2.9.55.9	13	394	9	2.29	0.7760	0.3512	0.2902
1.7.35.2	72	1904	65.1	3.42	0.8554	0.3998	0.3146
1.7.16.8	60	1187	39.65	3.34	0.7690	0.4044	0.3174
1.7.32.9	17	578	17.1	2.96	0.7183	0.3540	0.2800
1.7.26.2	23	748	27.58	3.69	0.8062	0.3896	0.3180
2.9.49.11	18	329	7.02	2.13	0.6814	0.3446	0.2784
2.9.16.4	52	1148	36.29	3.16	0.7762	0.3244	0.2758
1.7.37.9	24	2	0.03	-	0.7380	0.2820	0.2560
1.7.36.9	84	2190	83.51	3.81	0.7700	0.3776	0.3150
2.9.33.2	37	1426	38.51	2.70	0.7394	0.3294	0.2780
2.9.11.4	19	449	13.63	3.03	0.7760	0.3788	0.3239
2.9.2.2	15	223	5.35	2.40	0.7960	0.3506	0.2836
2.9.54.14	15	336	7.44	2.31	0.6120	0.3662	0.3052
1.7.8.3	148	3088	71.23	2.31	0.8330	0.3220	0.2528
2.9.35.9	46	637	21.2	3.33	0.8226	0.4134	0.3990
1.9.16.8	23	712	12.53	1.76	1.7382	0.3256	0.2802
2.9.47.3	37	718	28.67	3.99	0.8430	0.3684	0.3122
2.9.15.9	19	445	11.78	2.65	0.8044	0.3822	0.3284
1.7.39.13	18	511	14.3	2.80	0.7790	0.3794	0.3188
2.9.44.8	49	1242	30.3	2.44	0.6800	0.3388	0.2874

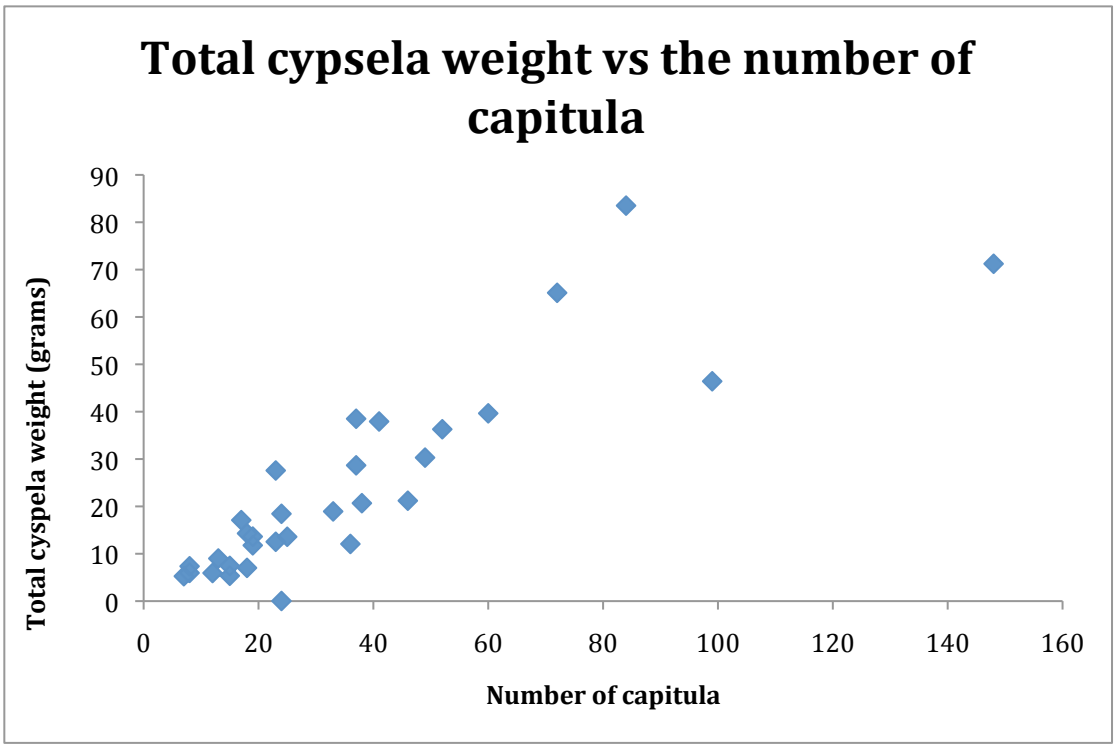
1.9.5.3	38	708	20.68	2.92	0.7912	0.3610	0.2790
1.7.12.2	25	489	13.59	2.78	0.6570	0.3694	0.3076
1.9.11.14	24	601	18.44	3.07	0.7848	0.3734	0.3276
2.9.10.3	36	455	12.07	2.65	0.8584	0.3522	0.2848
2.9.1.2	8	305	7.36	2.41	0.8326	0.3794	0.3066
1.7.21.8	12	248	5.93	2.39	0.7760	0.3366	0.2856
2.9.57.5	41	1058	37.93	3.59	0.7404	0.3972	0.3276
Standard Dev.	30.54	678.84	20.56	0.57	0.18	0.03	0.03

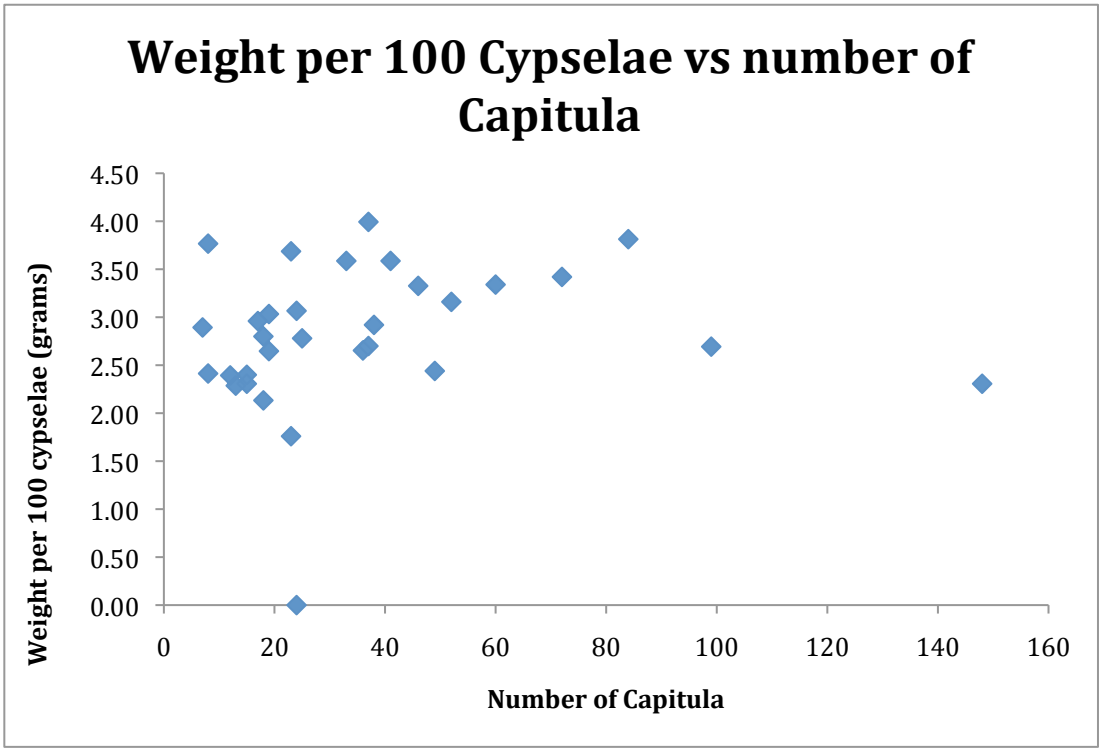
Individual plant height is given when possible.

Appendix : Scatter Plots of morphological correlations.






Total cypsela weight vs the number of capitula
Total cyspela weight (grams)
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
Number of capitula


Weight per 100 Cypselae vs number of Capitula
Weight per 100 cypselae (grams)
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
0
20
40
60
80
100
120
140
160
Number of Capitula

Chapter 4: General Conclusions

This thesis addresses two basic questions. First, what are the relationships within the genus *Carthamus*? Second, are hybrids of *C. oxyacanthus* and *C. tinctorius*, true genetic hybrids and does the genome of *C oxyacanthus*, a weedy close relative of *C. tinctorius*, introgress into the genome of *C. tinctorius*?

The low level of genetic variation was a major problem encountered in resolving relationships between *C. tinctorius* and close relatives. Several marker systems have been tried (RAPDs, Vilatersana et al. 2005; SNPs, Chapman et al. 2007) with varying degrees of success. Despite being considered a highly polymorphic marker system, the microsatellite markers used in Chapter 2 were still unable to resolve the relationships within the sect. *Carthamus*. There are several possible reasons for the low genetic variation among safflower and relatives. First, it has been shown that many traits between species in sect. *Carthamus* are under the control of single genes (Ashri and Efron 1964). This indicates that the source for the morphological variation between species may be the result of genetic variation in specific regions of the genome. It is likely that these regions have not been included in the limited sequence data that has been explored to date and it would be difficult to determine whether this hypothesis is correct, with the minimal marker data and genetic characterization that existed until developed within our research group. Second, it has been suggested that the species of sect. *Carthamus* are all races of a single polymorphic species (Ashri and Efron 1964; Imrie and Knowles 1970). However, this hypothesis does not fully account for the considerable morphological

diversity between species. Phenotypic plasticity of *C. tinctorius* and *C. oxyacanthus* in Chapter 3, as well as crossing and inheritance studies (Ashri and Efron 1964; Imrie and Knowles 1970) support this theory. In this case *C. oxyacanthus*, *C. palaestinus* and *C. tinctorius* would all be races of *C. tinctorius*. Third, continued crossing may blur species boundaries.

Crossing has been observed between members of sect. *Carthamus* in both greenhouse (Mayerhofer et al., unpublished) and natural (Desphande 1952; Ashri and Knowles 1960) settings (reviewed in McPherson et al. 2004). In fact, the ability for species to cross has been used as evidence of the close relationship between members of sect. *Carthamus* (Ashri and Knowles 1960). Interspecific hybrids create the opportunity for regions of the genome to move between different species. Introgression occurs naturally between the two genomes (ie there was no help needed to induce recombination) as evidenced by Mayerhofer et al. (2010) and as seen in Chapter 3. The introgression of traits from one species to another can prove beneficial for crop development, and possibly in nature, by providing a source of variation (Hajjar and Hodgkin 2007). It has also been suggested that introgression may serve as a potential source of repair for badly damaged genes, by providing a new copy (Rieseberg 2009).

Crop breeding programs take advantage of the introgression of new variation or traits to improve the cultivars of a crop for yield or fitness. Although there are many instances (Hajjar and Hodgkin 2007), one case where introgression has

proven useful is the transfer of cytoplasmic male sterility (CMS) from *Raphanus sativus* (radish) into *Brassica napus* (Heyn 1976; Delourme et al. 1992; Delourme and Eber 1992). This example also highlights a phenomenon known as linkage drag, where closely linked loci which have a negative effect on fitness are introgressed with the desired or beneficial traits. When CMS, which is a mitochondrial trait was introduced into *B. napus*, an additional nuclear gene, *Rfo*, was required(Delorme et al. 1998). However, the genotypes with the *Rfo* locus were also lines which contained additional genes which had negative effects on the line, and the linkage between these loci ultimately required gamma ray irradiation to encourage recombination. Thus, high levels of recombination between the species is desirable, along with fine scaled genetic maps.

Although there is sufficient similarity between the genomes of *C. oxyacanthus* and *C. tinctorius* allowing for pairing and recombination, the extent to which this occurs varies across the genome (Fig. 3.4 a-e). The movement of DNA in the selfed generation from the backcross of the *C. tinctorius* x *C. oxyacanthus* interspecific crosses to *C. tinctorius* (BC1S1) shows a trend of *C. oxyacanthus* DNA being incorporated in large groups of markers, indicating recombination is reduced from the intraspecific cross of *C. tinctorius* (Mayerhofer et al. 2010). However, linkage group 5 (Fig 3.4 d) revealed the potential for introgression of smaller areas of linkage groups.

The ability of DNA to move between *C. oxyacanthus*, a noxious weed, and *C. tinctorius* raises biosafety concerns. *Carthamus oxyacanthus* is the divergent species

of sect. *Carthamus* with regards to both morphology (Deshpande 1952; but see Keil 2006) and genetic variation (Figs. 2.1, 2.2). The recombination of *C. tinctorius* with that of another species in the section suggests this may also be possible with other members of the section (e.g., *C. palaestinus*). As most relatives are weedy to some extent (USDA; Ashri and Efron 1960), the potential to transfer genes becomes more of an issue. The most serious outcome of gene transfer is between genetically modified crops and weedy relatives (McPherson et al 2004; Ellestrand et al. 1999; Snow et al. 2005). Depending on the gene and its effects on the host plant, it may result in a decrease in fitness of the plants, decreasing the risk of transgene escape.

Cultivars of *C. tinctorius* have been developed for plant molecular farming for a number of pharmaceuticals, the most well known being insulin (SemBioSys Genetics). The potential for escape of these pharmaceutical-producing genes into the wild relatives is undeniable. Transfer of transgenes from crop to wild populations has been observed in oilseed rape (Warwick et al. 2003), rice (Sanchez Olguin et al. 2009, Shivrain et al. 2007) and wheat (Seefeldt et al. 1998). In safflower, Mayerhofer et al. (submitted) have already demonstrated the transfer of a transgene between a number of species within *Carthamus* (*C. tinctorius x C. leucocaulos, C. oxyacanthus, C. palaestinus, C. tinctorius,* and *C. lanatus*). Moreover, not only is *C. tinctorius* able to cross with many other species (Ashri and Knowles 1960, McPherson et al. 2004), but it's the genomes and chromosomes are able to recombine (Mayerhofer et al. 2010, Chapter 3). Moreover, weedy species are present in many areas of *C. tinctorius* production in Pakistan, India and North

America (USDA, Ashri and Efron 1964, Keil 2006), making at least occasional crossing plausible. Outcrossing of the cultivar used for transgenic production has been shown to occur at relatively low levels and decrease with distance from a pollen source (McPherson et al. 2009), which will help in management of transgenic crops.

While the ability of *C. tinctorius* and *C. oxyacanthus* to cross and recombine genomes may be problem for the management of transgenic crops, it is beneficial for improving crop cultivars. *Carthamus tinctorius* is currently a minor crop (Mündel et al. 2004), but has potential for production in semi-arid areas due to its long taproot (Mundel et al. 2004; Johnston et al. 2002). The wild relatives of *C. tinctorius*, including *C. oxyacanthus*, may contain valuable traits for breeding programs, and the ability of the genomes to recombine shows potential for successful use in a breeding program. The ability of single markers to move into the genome of *C. tinctorius* (Fig. 3.7a-e) also shows potential for use in breeding programs. Imrie and Knowles (1970) have also noted that *C. palaestinus* has potential for use in crop breeding programs, due to the ease with which the plants hybridize and morphological differences seen between the two species. Many of the genes responsible for the differences between *C. tinctorius* and its relatives also appear to be under the control of single genes (Ashri and Efron 1964, Imrie and Knowles 1970).

Future Work:

To further clarify species boundaries in *Carthamus*, two main options are available: increased sampling and additional markers. Likely both are necessary in

for resolving relationships across the genus. Including representatives of species that were not used in this study is important. For example, *C. curdicus* and *C. gypsicola*, two species with limited geographical distributions in Sect. *Carthamus*, are likely to contribute to our understanding of species relationships within Sect. *Carthamus*. Sampling all members of sect. *Carthamus* is necessary to clarify all relationships within the section. Given the all species are not monophyletic, it is also important to include multiple individuals throughout the range of each species. Recent work with sampling across the geographic distribution of members of *Carthamus* has been informative (Bowles et al. 2010, Chapman et al. 2010). It is likely that even with increased sampling of both specimens and sequence, species will be unresolved, resulting in possible taxonomic changes.

Additional markers are also needed to increase the resolution of relationships, especially within sect. *Atractylis*, and help clarify species boundaries in sect. *Carthamus*. Microsatellites have shown to be useful for phylogenetic studies of other species and provided some resolution within sect. *Carthamus* (Bowles et al. 2010). As such, adding additional microsatellite markers used in the study may increase the resolution within the genus. Alternatively, more sequence data may be useful if a region can be found with appropriate variation. Furthermore, examining morphology data will address whether or not species in sect. *Carthamus* should be treated as species or races of a single species. The combination of both morphological and molecular data can increase phylogenetic resolution (e.g.. Columnea – Smith et al. 1994; Costaceae – Specht 2006). If the lack of species

boundaries is confirmed with other molecular data, it may be pertinent to revisit the hypothesis that *C. oxyacanthus, C. palaestinus* and *C. tinctorius* are races of a polymorphic species (Ashri and Efron 1964; Imrie and Knowles 1970).

The current map and crossing data for *Carthamus* is insufficient for a breeding program, however it is an advancement on what has been available to date. In the future, in order for the map to be useful in marker assisted breeding, the coverage and density of markers along the linkage groups would need to be increased. This is especially true for the linkage groups 2, 7, 8, 9, 10, 11 and 12 (Mayerhofer et al. 2010), which were not investigated in Chapter 3 due to low numbers of markers available. The existing map will be able to serve as an excellent starting point to increase the resolution of the map, by adding additional markers to the existing map. Microsatellite makers have proved useful to date for mapping, but due to the low levels of genetic variation in the genus (Bowles et al. 2010), additional markers, such as single nucleotide polymorphisms (SNiPs) may be necessary to increase the map resolution. The number of markers needed is variable between species (1000 for tomato Tanksley et al. 1992, 2275 for rice Harushima et al. 1998) and can not be predetermined.

In Chapter 3 the introgression of *C. oxyacanthus* DNA into the *C. tinctorius* genome is shown to occur, however the stability of the introgressed regions is unknown. Additional mapping studies of further backcross generations, or recombinant inbred lines would indicate the stability of introgressed regions identified in this study. Utilizing additional markers along the linkage groups will be

able to provide more detail on what regions are moving between genomes. The information of the introgression from the BC1S1 generation is a reference point for further generations to be compared against to find areas of introgressed DNA that may have been lost or have changed in size. A better understanding of what is happening to the *C. oxyacanthus* introgressed DNA will be helpful for breeding programs which often need information on the number of generations needed for the introgressed regions to stabilize.

The population utilized in Chapter 3 is backcrossed to the commercial *C. tinctorius* which does not allow study of how the genome of the commercial variety moves into that of the noxious weed, making it of limited use in determining the risk of transgene escape. Additional introgression work should be done utilizing a hybrid population that is backcrossed to the weedy relative to assess the movement of commercial DNA to the weedy relatives. Utilizing a commercial line that contains a transgene will allow for a real assessment of the movement of the transgene out of the commercial crop, and the effects that the transgene will have on the wild relatives. It is possible that the addition of a large gene producing proteins unnecessary for the survival of the plant will decrease the fitness of the plant, decreasing the risks of transgene escape.

Finally, additional research needs to be done on key traits that may improve safflower as a crop species. These might include time to flowering, yield, disease resistance, etc. Preliminary trait measurements in chapter 3 indicated that there may be lines with potentially useful morphological variation. Two plants from

different lines showed increased seed numbers over plants from other lines. Depending on the oil amounts and quality in these lines, they may be of use for increasing yield for crop lines to be grown in southern Alberta. Increased grow outs of any lines that contain variation that are desirable with more detailed trait measurements would provide a better of idea of lines to focus on for breeding programs.

Final Conclusions

In summary the research presented here has accomplished two main objectives. First, relationships in *Carthamus* have been clarified and the lack of molecular boundaries between species has been identified. Second, mapping studies have shown recombination between genomes of the crop species and a weedy relative.

Molecular data has confirmed the presence of two sections as suggested by Vilatersana et al. (2000) using more markers and increased sampling of the genus. A new marker system, microsatellites, highlighted the lack of distinct boundaries between the species but was unable to clarify the relationships within the sect. *Carthamus*. Utilizing the relationships provided by the sequence and microsatellite data, along with morphology and previous work, *C. palaestinus* was identified as the closest relative to safflower, with other members of sect. *Carthamus* being closely related.

Work with an interspecific cross of *C. tinctorius* and *C. oxyacanthus* has shown that the genomes of the two species are close enough to allow for

recombination to occur. The amount of recombination between the two genomes varies from one linkage group to another, indicating that there are differences in some areas of the genome that may inhibit recombination or decrease fitness. The ability of the genomes to recombine shows promise for crop breeding programs, as relatives provide a source of genetic variation.

Literature Cited:

Ashri A, Efron Y (1964) Inheritance studies with fertile interspecific hybrids of three *Carthamus* L. species. Crop Science 4:510-514.

Ashri A, Knowles PF (1960) Cytogenetics of Safflower (*Carthamus* L.) species and their hybrids. Agronomy Journal 52: 11-17.

Bowles VG, Mayerhofer R, Davis C, Good AG, Hall JC (2010) A phylogenetic investigation of *Carthamus* combining sequence and microsatellite data. Plant Systematics and Evolution (in press)

Chapman MA, Burke JM (2007) DNA sequence diversity and the origin of cultivated safflower (*Carthamus tinctorius* L.; Asteraceae). Bmc Plant Biology 7: 60.Chapman MA, Hvala J, Strever J, Burke JM (2010) Population genetic analysis of safflower (*Carthamus tinctorius*; Asteraceae) reveals a Near Eastern origin and five centers of diversity. American Journal of Botany 97:831-840.

Delourme R, Eber F, Renard M (1991) Radish cytoplasmic male sterility in rapeseed: breeding restorer lines with a good female fertility. Proceedings 8th International Rapeseed Converence 5:1065.

Delourme R, Eber F (1992) Linkage between an isozyme marker and a restorer gene in radish cytoplasmic male sterility of rapeseed (*Brassica napus* L.) Theoretical Applied Genetics 85:222-228

Delourme R, Foisset N, Horvais R, Barret P, Champagne G, Cheung WY, Landry BS, Renard M (1998) Characterisation of radish introgression carrying the

Rfo restorer gene for the *ogu*-INRA cytoplasmic male sterility in rapeseed (*Brassica napus* L.). Theoretical Applied Genetics 97:129-134.

Deshpande RB (1952) Wild safflower (*Carthamus oxyacantha* Bieb.) - a possible oilseed crop for the desert and arid regions. Indian Journal of Genetic Plant Breeding 12: 10-14.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539-563.

Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: A survey of developments over the last 20 years. Euphytica 156:1-13.

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148:479-495.

Heyn FW (1976) Transfer of restorer genes from *Raphanus* to cytoplasmic male sterile *Brassica napus*. Eucarpia Cruciferae Newslett 1:15-16.

Imrie BC, Knowles PF (1970) Inheritance studies in interspecific hybrids between *Carthamus flavescens* and *C. tinctorius*. Crop Science 10: 349-352.

Johnston AM, Tanaka DL, Miller PR, Brandt SA, Mielsen DC, Lafond GP, Riveland NR (2002) Oilseed crops for semiarid cropping systems in the Northern Great Plains. Agronomy Journal 94:231-240.

Keil DJ (2006) *Carthamus*. In: Flora of North America. Edited by Barkley TM,

Brouillet L, Jeude H, Strother JL, Gandhi K, Kiger RW, Yatskievych K, Sarucchi JL. New York, New York, Vol.19 pp178-181

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

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

McPherson MA, Good AG, Topinka AKC, Hall LM (2004) Theoretical hybridization potential of transgenic safflower (*Carthamus tinctorius* L.) with weedy relatives in the New World. Canadian Journal of Plant Science 84: 923-934.

McPherson MA, Yang RC, Good AG, Nielson RL, Hall LM (2009) Potential for seed-mediated gene flow in agroecosystems from transgenic safflower (*Carthamus tinctorius* L.) intended for plant molecular farming. Transgenic Research 18:281-299.

Mündel HH, Blacksaw RE, Byers JR, Huang HC, Johnson DL, Keon R, Kubik J, McKenzie R, Otto B, Roth B, Stanford K (2004) Safflower production on the Canadian prairies: revisited in 2004. Agriculture and Agri-Food Canada, Lethbridge Research Center, Alberta. Pg 1.

Rieseberg (2009) Evolution: replacing genes and traits through hybridization. Current Biology 19: 119 – 122.

Sanchez Olguin ER, Espinoza-Esquivel A, Lobo JA, Arrieta-Espinoza G (2009) Assessment of gene flow from a herbicide-resistant indica rice (*Oryza sativa* L.) to the Costa Rican weedy rice (*Oryza sativa*) in Tropical America: Factors affecting hybrization rates and characterizatio of F1 hybrids. Transgenic Research 18:633-647.

Seefeldt SS, Zementra R, Young FL, Jones SS (1998) Production of herbicide-resistant jointed goatgrass (*Aegilops cylindrical*) x wheat (*Triticum aestivum*) hybrids in the field by natural hybridization. Weed Science 46:632-634.

Shivrain VK, Burgos NR, Anders MM, Rajguru SN, Moore J, Sales MA (2007) Gene flow between Clearfield (TM) rice and red rice. Crop Protocols 26:349-356.

Smith JF, Sytsma KJ (1994) Molecules and morphology: Congruence of data in Columnea (Gesneriaceae). Plant Systematics and Evolution 193:37-52.

Snow AA, Andow DA, Gepts P, Hallerman EM, Power A, Tiedje JM, Wolfenbarger LL (2005) Genetically engineered organisms and the environment: Current status and recommendations. Ecology Applications 15:377-404.

Specht CD (2006) Systematics and evolution of the tropical monocot family Costaceae (Zingiberales): A multiple dataset approach. Systematic Botany 31:98-106.

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Brown P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato

genomes. Genetics 132:1141-1160.

Vilatersana R, Garnatje T, Susanna A, Garcia-Jacas N (2005) Taxonomic problems in *Carthamus* (Asteraceae): RAPD markers and sectional classification. Botanical Journal of the Linnean Society 147: 375-383.

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

Warwick SI, Legere A, Simard MJ, James T (2003) Do escaped transgenes persist in nature? The case of an herbicide resitance transgene in a weedy *Brassica rapa* population. Molecular Ecology 17:1387-1395.