7th International Safflower Conference
WAGGA WAGGA AUSTRALIA

Development of Microsatellite Markers for investigation of relationships among closely related *Carthamus* species.

Victoria Bowles[1], Jocelyn C. Hall, Allen Good
Department of Biological Sciences, University of Alberta, Edmonton, AB, T6E 2E2, Canada,
[1]vikki.bowles@ualberta.ca

Abstract
Carthamus is a diverse group of plants that are of economic interest due to the commercial growth of one member, *C. tinctorius* (safflower). Recent attention in using safflower as a crop for pharmaceutical production necessitates a better understanding of the relationships between this species and its wild relatives. To this end, molecular markers need to be developed that have appropriate variation and that can be utilized across a range of species. The goal of this research was to isolate microsatellite makers that were able to reliably amplify across the genus. Initial screens of safflower resulted in 48 promising markers, of which eight were selected to study in eight *Carthamus* species. Phylogenetic analyses of this subset of markers reveal potential introgression and hybridization among species. In conclusion, microsatellite data, combined with sequence data, has the potential to resolve species relationships within *Carthamus*.

Key Words Safflower – microsatellite – phylogenetics - library

Introduction

Safflower (*Carthamus tinctorius*) is a commercial oilseed crop grown in approximately 60 countries and is currently being evaluated as a crop platform for molecular farming. Before releasing novel crops, the risk of potential movement of transgenes to wild populations must be addressed (Raybould and Gray 1993). A critical first step in assessing the possible transgenes movement is to elucidate relationships between crop species and their wild relatives. There are several weedy relatives to safflower, in the areas where these crops are grown, that have the potential to cross with the safflower crops and produce fertile offspring.

Safflower belongs to the genus *Carthamus* along with 14 other species. Taxonomy of this genus has been historically problematic. Lopez-Gonzalez (1989) places species in three groups or sections that correspond with chromosome number: sect. *Carthamus* (n=12), sect. *Odonthagnathis* (n=10, 11), and sect. *Atractylis* (n=22, 32). In contrast, recent studies based on variation in DNA sequences indicate the species are divided into two natural groups, sect. *Carthamus* and sect, *Atractylis* (Vilatersana et al. 2000, 2005). However most of these studies did not include safflower (e.g., Vilatersana et al. 2005, 2007) or had limited sampling if safflower was included (e.g., Vilatersana et al. 2000; Chapman and Burke, 2007).

Since sequence data has limited use in discerning species relationships in *Carthamus*, alternative makers need to be developed. Microsatellites are found to show relatively large amounts of variability between individuals, and as such are good candidates for examining diversity within species as well as among closely related taxa. They consist of tandem repeats, with the repeat unit consisting of 2-5bps. The current model is that variation within these tandem repeats results from slippage of the polymerase during replication, resulting in the duplication or deletion of repeat unit(s), altering the length of the overall microsatellite (Griffiths et al., 2005 and Hearne et al., 1992). It is the difference in size that can be visualized by designing primers on either side of the microsatellite, and amplifying the region using PCR (Griffiths, 2005 and Hearne, 1992). Often the differences in product size are only 2bps (one repeat unit) and as such great resolution is required to enable differences in allele size to be detected.

In this research we produced a microsatellite-enriched library from *C. tinctorius* to use for the exploration of microsatellite loci as markers across species of *Carthamus*. These markers were

tested for amplification and variation in eight species of *Carthamus*. Preliminary analyses were also conducted to examine relationships between the different species of the *Carthamus* genus.

Material and methods:
A genomic library was created from *Carthamus tinctorious* S317 DNA. The resulting library was then enriched for microsatellite-containing clones using repeat probes labeled with biotin and a streptavidin column. Probes were for simple di and tetra nucleotide repeats (Hamilton et al., 1999).

Inserts from clones were sequenced using the T7 primer from the pBSII SK+ vector that contained the insert and BigDye Ready Reaction Mix (Applied Biosystems). Sequences were then manually scanned for the presence of a repeat. Primers were designed to the sequence flanking the repeat using Primer3 (Rozen and Skaletsky, 2000).

Using M13 tagged primers and a labeled M13 primer, bands were visualized with a capillary sequencer (Scheulke et al, 2001). Amplification of primer sets were tested on 36 plants from eight species of *Carthamus.* Three species were tested from sect. *Carthamus*: *C. oxyacanthus, C. palaestinus,* and *C. tinctorius*. This sampling includes three cultivars of safflower including, Centennial, S317 and NP12. Five species from sect. *Atractylis* were also examined: *C. beaticus* (= C. *creticus*), *C. glaucus, C. lanatus, C. leucocaulos,* and *C. turkestanicus*.

All microsatellite loci were viewed and sized using Genemapper (Applied Biosystems). To insure accurate results, the band sizes were repeated multiple times. A band size was not accepted as true until it had been repeated at least twice. Band sizes were scored by sizing the tallest of the bands, assuming the smaller bands were stutter bands. When two or more separate sets of bands were present, each set was sized independently.

Genetic distances were calculated using MICROSAT version 1.5 (Goldstein et al, 1995) and were then used to produce neighbors joining trees in PAUP*V4.1 (Swofford, 2000). Distances were constructed using both Nei's genetic distance (Nei, 1972) and delta mu-squared (Goldstein et al, 1995). Nei's genetic distance is based on a ratio of proportion of alleles that are like each other within and between populations, while delta mu-squared is based on variations in the number of repeat units measured as allele length (Alvarez, 2001).

Results:
Dinucleotide repeat libraries from *C. tinctorius* (safflower) were successful upon initial screening, whereas the tetranucleotide repeat libraries were unsuccessful. No tetranucleotide repeats were found in the initial screening of libraries and, as such, were not considered further. The screening of the dinucleotide repeat libraries led to the sequencing of 196 clones. Of these, 48 clones contained microsatellite repeats with flanking sequencing on both ends sufficient for designing primers.

Initial amplification of the markers found that all 48 primer sets produced a product from plants of the three *C. tinctorius* lines, and 33 of these amplified in all eight species tested. Primer sets that created bands that were easy to decipher and amplified in all tested species were then repeated again. A total of 16 primer sets were re-amplified, and 8 of those were screened on all individuals. Consensus sizes were determined for the eight primer sets (Table 1) from the 36 plant samples examined.

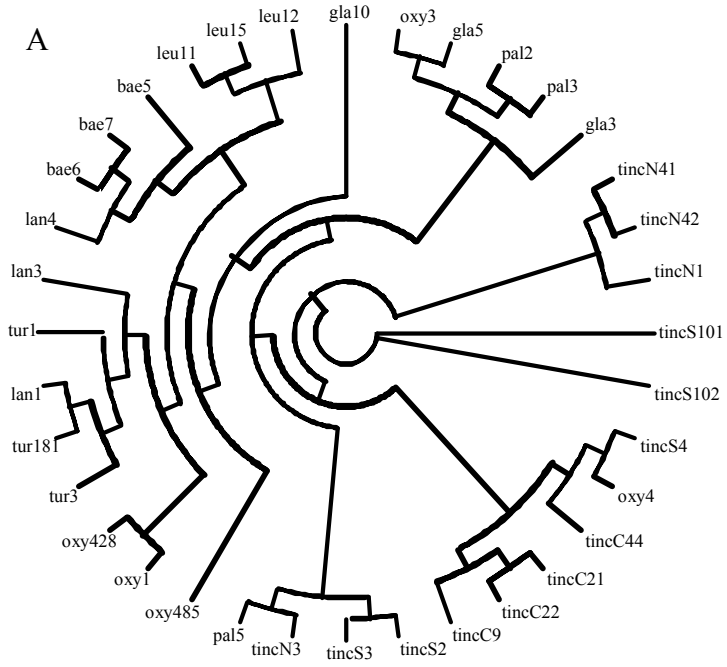
Table 1: Allele summary of the 8 primer sets used for subsequent analyses.

Primer Set	Core Sequence	Allele size range	No. of Alleles	No. of specific alleles	Accessions with specific alleles	Species with specific alleles
VL001	(CT)6C(CT)6(CA)8	130-173	16	4	426477, 407602, 426427	*C. leucocaulos*
VL003	(GT)11	165-175	6	2		*C. leucocaulos, C. glaucus*
VL021	(CA)12	163-176	11	4	426426, 426180, 426428	*C. palaestinus*
VL023	(GT)11	160-173	8	1		*C. oxyacanthus*
VL031	(GT)14	111-174	11	3		*C. creaticus*
VL037	(CT)10	236	1	0		
VL040	(CA)12 - imperfect	255-275	8	4	426426, 426428, 235668	*C. leucocaulos*
VL046	(GA)10	294-337	13	7	426428, 426180, 426426, 235668, 326364	*C. leucocaulos, C. glaucus*

Summary of alleles for loci examined in 36 plants from 8 species of *Carthamus*. Core sequences are known from sequencing of the clones for primer design. Allele size range is given as the size of smallest and largest observed alleles in bp. Specific alleles were seen either only in one species (multiple accessions) or within a single accession of a species. Accessions with specific alleles are identified by the PI number for the accession.

Trees were constructed using both Nei's genetic distance (Fig 1a) and delta mu-squared (Fig 1b), a distance measure specific to microsatellite data. Although there are some minor differences between the two topologies, similar patterns emerge from these analyses. The two sections of *Carthamus* do not form distinct groups. In addition, exemplars from the same species are not clustering together, as would be expected. Two notable exceptions are *C. leucocaulos* in the Nei's tree (Fig. 1a) and *C. beaticus* (= *C. creticus*) in the delta mu-squared tree (Fig. 1b). Samples from *C. lanatus, C. oxyacanthus*, and *C. tinctorius* are dispersed across the tree, regardless of genetic distance used. In contrast, *C. beaticus* (= *C. creticus*) and *C. leucocaulos* are found close to each other, but with exemplars from other species nested within the same cluster.

Figure 1: Phylogenetic trees of *Carthamus* based on microsatellite data.



B

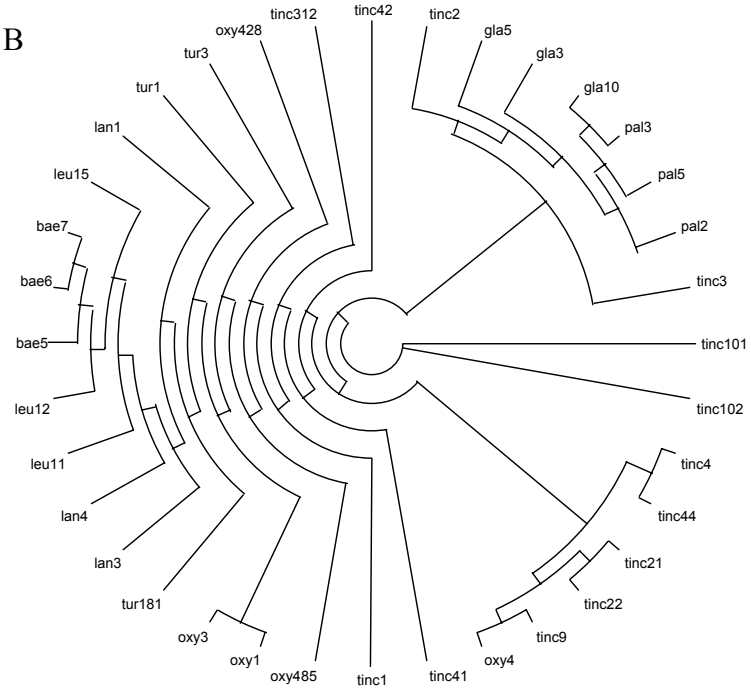


Figure 1a: Nei's genetic distance Neighbour joining phylogenetic tree.
Figure 1b: Delta mu Neighbour joining phylogenetic tree.
Taxon designations are the first three letters of the specific epithet followed by a plant identification number. The abbreviations tinC, tinS and tinN are all cultivars of *C. tinctorius,* corresponding to Centennial, S317 and NP12, respectively. The rest of the abbreviations are as follows: oxy = *C. oxyacanthus*, tur = *C. turkestanicus,* leu = *C. leucocaulos*, lan = *C. lanatus*, bea = *C. creaticus (C. baeticus)*, gla = *C. glaucus* and pal =*C. palaestinus*.

Discussion

The enriched library provided a fast way of identifying clones containing microsatellites available for primer design. From the clones sequenced, 24% resulted in loci available for primer design. These primers showed good amplification across species with 69% amplifying in all tested species.

High polymorphism of the amplified loci was observed with seven of the eight primer sets showing multiple alleles. There was an average of 10.4 alleles per polymorphic loci, with seven loci showing species or accession specific alleles. The average number of alleles per species per locus was 2.3, indicating within-species diversity. The diversity found with these microsatellites is comparable to the amount of diversity found in other groups within the same family (e.g., Tang and Knapp, 2003).

Although the markers are polymorphic, the species do not group together when analyzed. This pattern indicates problems with species boundaries and the need for a more in-depth analysis. Others have noted the taxonomic difficulties in this genus (e.g. Vilatersana, 2005). The nesting of species within species may be explained by introgression, as some of the species have been shown to cross in greenhouse conditions (McPherson, et al. 2004). The results presented here support a hybrid origin of some species within the genus. Three species from sect. *Carthamus* - *C. lanatus, C. leucocaulos* and *C. creticus* (*C. beaticus*) - cluster together in the Nei's genetic distance tree (Fig. 1a), which is consistent with data from RAPD markers (Vilatersana, 2005). This pattern supports the hypothesis that *C. beaticus* (= *C. creticus*) is the result of a cross between *C. lanatus* and *C. leucocaulous* (Khidir and Knowles, 1970). These plants have been seen to overlap in distribution in both the native Mediterranean habitat and in the state of California. The chromosome numbers also are consistent with this model with *C. creticus* (n=32) being the sum of the chromosomal numbers from *C. lanatus* (n=22) and *C. leucocaulos* (n=10).

Since not all species are of hybrid origin in the genus, alternative explanations need to be explored to explain patterns observed in the phylogenetic trees (Fig. 1). Plants of the same species not clustering together may be due to the low number of loci analyzed and the occurrence of convergence of the allele sizes. There may be a high noise-to-signal ratio within the data, as it appears for some of the loci that nearly all possible allele sizes are seen (25.6bp average size difference, with an average of 10.4 alleles per polymorphic loci). It is also possible that microsatellites are too variable to examine relationships across sections (e.g. too much noise), but will provide appropriate variation to discern species relationships within sections.

In summary, the microsatellites generated in this study are highly variable and have uncovered species patterns consistent with hybridization. However, some of the phylogenetic results may be the result of high noise to signal ratio. Continuing work will use sequence data from both the chloroplast and nuclear genomes, along with new microsatellites to elucidate species relationships surrounding cultivated safflower. This sequencing effort will expand on previous taxon sampling and will test the support of the two sections identified by Vilatersana (2005). Additional microsatellites loci will be used within subgroups to increase resolution of relationships among species. Reduction in the noise to signal ratio will hopefully be achieved by using the microsatellites within a subgroup, rather than across the genus. The additional markers will also sort out the possibility that too few markers were used.

Acknowledgements
The research was funded by SemBioSys Genetics Inc. and a NSERC CRD grant to AGG. Marion Mayerhofer generously made available the DNA material utilized in this study. Cory Davis provided valuable logistical and technical assistance in creating the library and initial screening of microsatellites. We would also like to thank the Microbiology Service Unit, University of Alberta, for support.

References

Alvarex, A.E., Van de Wiel, C.C.M., and Smulders, M.J.M. 2001 Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theoretical Applied Genetics 103:1283-1292.

Chapman, M.A., and Burk, J.M. 2007 DNA sequence diversity and the origin of cultivated safflower (*Carthamus tinctorius* L.; Asteraceae). BMC Plant Biology 7:60-68

Griffiths, A., Wessler, S., Lewontin, R., Gelbart, W., Suzuki, D. T., and Miller, J. H. 2005 Introduction to Genetic Analysis, 8th edition. W.H. Freeman and Company, New York. Pp400-401

Goldstein, D.G., Ruiz-Linares, A., Cavalli-Sforza, L.L. and Feldman, M.W. 1995 An evaluation of genetic distances for use with microsatellite loci. Genetics 139:436-471

Hamilton, M. B., Pincus E.P., Fiore, D.F., and Fleischer, R.C. 1999 Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. BioTechniques 27:500-507.

Hearne, C.M., Ghosh, S., and Todd, J.A. 1992 Microsatellites for linkage analysis of genetic traits. Trends in Genetics 8:288-294

Khidir, M.O. and Knowles, P.F. 1970 Cytogenetic studies of *Carthamus* species (Compositae) with 32 pairs of chromosomes. I. Intrasectional Hybridization. American Journal of Botany 57:123-129

Lopez-Gonzalez, G 1989 Acerca del la classification natural del genero *Carthamus* L. s.l. Anales del Jardin Botanico de Madrid 47:11-34

McPherson, MA. Good, AG, Keith, A. Topinka, C. 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

Nei, M. 1972 Genetic distance between populations. American Naturalist 106:283-292

Raybould, A. and Gray, A.J. 1993 Genetically modified crops and hybridization with wild relatives: A UK perspective. Journal Applied Ecology 30:199-219.

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

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

Swofford, D.L. 2000 PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts, USA.

Tang, S. and Knapp, SJ 2003 Microsatellites uncover extraordinary diversity in native American land races and wild populations of cultivated sunflower. Theoretical Applied Genetics 106:990-1003

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

Vilatersana, R., Garnatje T., Susanna, A. and 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., Byrsting, A.K. and Brochmann, C. 2007 Molecular evidence for hybrid origins of the invasive polyploinds *Carthamus creticus* and *C. turkestanicus* (Cardueae, Asteraceae). Molecular Phylogenetics and Evolution 44:610-621