AN ABSTRACT OF THE THESIS OF

Ju-Kyung Yu for the degree of Doctor of Philosophy in Crop Science presented on September 7, 2001. Title: Simple Sequence Repeat Marker Development and Mapping in Cultivated Sunflower, *Helianthus annuus* L.

Abstract approved: Redacted for privacy

Steven J. Knapp

The cultivated sunflower (*Helianthus annuus* L., x = 17) is one of the most important annual oilseed crops in the world. There are very few publicly shared sequence-based DNA markers and genetic maps in sunflower, even though molecular DNA markers and genetic maps have become widely used in all areas of genetic research and breeding in plant species. The objectives of this study were to develop sequence-based molecular markers and utilize the markers for genetic analyses and constructing maps in the cultivated sunflower. A total of 131 functional simple sequence repeat (SSR) markers were developed for 16 elite inbred lines using a small insert genomic library enriched for short simple sequence repeats. The polymorphism information content (PIC) estimated from 74 polymorphic SSR markers ranged from 0.0 to 0.93 with mean value of 0.55. Tetranucleotide repeats were significantly more polymorphic than dinucleotide and trinucleotide repeats, and no obvious correlation was found between repeat numbers and PIC scores. Genetic distance among 16 inbred

lines, estimated from 74 polymorphic SSR markers ranged 0.175 to 0.543. Principal coordinate and cluster analyses of the genetic distance matrix well explained the difference between oilseed lines and confectionery lines, and sterility maintainer lines and fertility restorer lines. A total of 1,090 SSR markers were screened for polymorphism between the parents of two mapping populations. The two genetic maps were constructed by genotyping 94 recombinant inbred lines from a cross between PHA and PHB (276 SSR loci covering 1377.4 cM with mean distance of 4.99 cM), and 94 $F_2$ progeny from a cross between HA370 and HA372 (122 SSR loci integrated into the existing RFLP framework map covering 1348.0 cM with mean distance of 6.77 cM). Ninety-three percent of the SSR markers were mapped to single loci and 56.5 % of the loci were co-dominant. Clustering of SSR loci was observed near centromeric regions and most of the distorted loci were mapped to centromeric or distal regions. A concerted effort to develop SSR markers and generate high-resolution SSR maps will enhance future fingerprinting analyses, fine-scale genome analyses and molecular breeding in the cultivated sunflower.

# Simple Sequence Repeat Marker Development and Mapping in Cultivated Sunflower, *Helianthus annuus* L.

by

Ju-Kyung Yu

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 7, 2001
Commencement June 2002

Doctor of Philosophy thesis of Ju-Kyung Yu presented on September 7, 2001

APPROVED:

Redacted for privacy

Major Professor, representing Crop Science

Redacted for privacy

Chair of the Department of Crop and Soil Science

Redacted for privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for privacy

Ju-Kyung Yu, Author

## ACKNOWLEDGMENTS

There are many people I need to acknowledge who have been a very important part of my research and personal life at Oregon State University.

I wish to express my sincere gratitude and indebtedness to my major professor, Dr. Steven J. Knapp, for his financial support, initiating the projects, his advising and guidance from the beginning of the study to preparation of the manuscripts. Especially, I want to thank him for his patience during difficult research circumstances in the spring, 2000. I am thankful to members of my graduate committee, Dr. Lynda Ciuffetti, Dr. Patrick M. Hayes, Dr. Oscar Riera-Lizarazu, and Dr. Robert P. Griffiths for their time and guidance.

I am very indebted to Dr. Mary B. Slabaugh, who taught me all aspects of molecular techniques and was always generous with her time and even gave me her shoulder when I was in trouble. Her encouragement, support, concern, and warm heart will be forever appreciated. Grateful acknowledgment is made to Dr. Melaku Gedil who taught me how to operate the "PCR" machine and all aspects of lab techniques. Special thanks and gratitude are extended to Dr. Sureeporn Katengam, Dr. Vipa Hongtrakul, Dr. Judy Kolkman and Venkatakrishnakishore for their unreserved help, and interesting academic discussions. Thanks are due to the individuals in the lab: Jimmie, Judy, Shunxue, Pablo, Kishore, Sonali, Felix, Cate and Adam for their warm smile, kindness and friendship throughout my study period. The faculty and staff of the Department of Crop Science who directly or indirectly contributed to the success of this work are gratefully acknowledged.

I wish to thank Pioneer Hi-bred International, INC for providing mapping progeny and performing disease resistant screening, and Caprice Rosato who assisted in automated genotyping of SSR markers using GeneScan and Genotyper software.

I must express my gratitude to Dr. Clara Jarman who has been a mother to me during my stay at OSU. She gave me endless love and showed me her tireless patience during proofreading this manuscript. Many thanks to my former advisors of the Department of Agronomy, Kyung-Pook Nat'l University, Korea for their encouragement and to my friends, especially Min-Jung Jin for her friendship and love.

My deepest appreciation goes to my family and my soul mate for their unending prayers, encouragement and love. Without their support this study could not has been completed. In addition, I hope this study encourages my brother and sister who are studying in graduate school now. Last but not least, I am thankful to my God, Jesus Christ and Saint Mary.

## CONTRIBUTION OF AUTHORS

Professor Steven J. Knapp initiated, advised, and supervised all aspects of the projects. Drs. Lucy Thompson and Keith J. Edwards were involved in genomic library construction for the SSR project. Ms. Jodie Mangor helped in SSR screening and designing primers for the SSR project. Dr. Mary B. Slabaugh contributed and helped in all aspects of molecular techniques and analyses. Dr. Shunxue Tang contributed in the development of SSR markers for the SSR mapping project.

## TABLE OF CONTENTS

## LIST OF FIGURES

**Figure** **Page**

## LIST OF TABLES

**Table** **Page**

# Simple Sequence Repeat Marker Development and Mapping in Cultivated Sunflower, *Helianthus annuus* L.

## CHAPTER 1

## INTRODUCTION

Sunflowers, both the cultivated form of *Helianthus annuus* L. and the 50 species of annual and perennial *Helianthus,* are native to North America (Schilling and Heiser 1981). They include diploids, tetraploids, and hexaploids, all with the base chromosome number of x = 17. Cultivated sunflower belongs to section *Helianthus,* along with ten other species that are primarily annual diploids. Sunflowers were widely cultivated at the time European explorers came to America and are presently grown throughout most of the world (Putt 1997). Sunflowers were first grown as a source of edible oil in nineteenth century Russia. The first significant breeding program to improve sunflower was initiated in Russia in the early twentieth century. In 1950 the U.S Department of Agriculture initiated sunflower research in the U.S in Texas.

Three types of sunflower oils are available (National Sunflower Association; http://www.sunflowernsa.com). High linoleic sunflower oil contains 69 % polyunsaturated fatty acid and is used as cooking oil. High oleic sunflower oil contains 82 % monounsaturated fatty acid and is used for food or industrial applications. NuSun is a mid-oleic sunflower oil developed by hybrid procedures and is new to the market. Breeding programs have three major objectives: i) improvement

of seed and oil yield, ii) introduction of resistance to biotic or abiotic stress, and iii) the development of new genetic resources (Gentzbittel et al. 1992; Fick and Miller 1997).

Molecular markers and marker mapping are part of the promising "new genetics" used in all areas of modern biology from genomics to breeding (Jones et al. 1997). DNA-based markers demonstrate sites of variation at the DNA sequence level. Unlike morphological markers, these variations do not usually show themselves in the phenotype or disturb the physiology of the organism. Compared to morphological markers they are much more abundant (Jones et al. 1997). A variety of molecular marker types have been developed recently in plants. The utility of these markers depends on properties such as reproducibility, information content, cost, and convenience (Powell et al. 1996). In a breeding program, molecular markers are powerful tools for marker-assisted selection (MAS), map-based cloning, germplasm fingerprinting and identification, and quantitative trait locus (QTL) analysis in crop species (Lee 1995).

In sunflower, several molecular markers have been developed. These include rapid amplification polymorphic DNA (RAPD, Lawson et al.1994; Arias and Rieseberg 1995), restriction fragment length polymorphism (RFLP, Gentzbittel et al. 1992, 1994; Berry et al. 1994, 1996, 1997; Zhang et al. 1995; Jan et al. 1998), amplification fragment length polymorphism (AFLP, Hongtrakul et al. 1997; Gedil et al. 2001b), intron fragment length polymorphism (IFLP, Hongtrakul et al. 1998b) and cleavage amplification polymorphism (CAPS, Gedil et al. 2001a). These have been used successfully in construction of genetic maps (Rieseberg et al. 1993; Berry et al.

1995, 1999; Gentzbittel et al. 1995; Jan et al. 1998; Gedil et al. 2001b), mapping and characterization of genes controlling qualitative traits such as disease resistance and fatty acid content (Besnard et al. 1997; Vear et al. 1997; Hongtrakul et al. 1998a, 1998b; Lawson et al. 1998; Gedil et al. 2001a), dissecting quantitative traits such as oil content and disease resistance (Leon et al. 1995; Gentzbittel et al. 1998; Mestries et al. 1998) and study of genetic diversity (Arias and Rieseberg 1995). A large number of RFLP markers were developed using cDNA clones as probes (Gentzbittel et al. 1992, 1994; Berry et al. 1994, 1996, 1997; Zhang et al. 1995; Jan et al. 1998). The first RFLP fingerprinting study in sunflower was carried out by Gentzbittel et al. (1992) on 44 *Helianthus* lines, followed by Berry et al. (1994), Gentzbittel et al. (1994) and Zhang et al. (1995) in cultivated sunflower. These results demonstrated that RFLPs could be useful descriptors for sunflower inbred lines. However, the percentage of polymorphic RFLP markers between pairs of inbred genotypes is relatively low (10 % - 30 %), and hybridization-based markers present practical disadvantages. Moreover, most RFLP markers were developed using independent sets of DNA probes from proprietary domains. Access to probes for RFLP markers and sharing of marker information has been limited, leading to serious drawbacks for the progress of sunflower genetic research. To overcome these problems, the sunflower research community needed to develop sequence-based DNA markers that would be in the public domain.

Microsatellites are regions of DNA that contain tandemly repeated short sequence motifs (< 6 bp) flanked by unique and conserved sequences. Simple sequence repeat (SSR) markers are microsatellite loci that can be amplified by the

polymerase chain reaction (PCR) using primer pairs positioned in the unique flanking sequences. Polymorphism is based on variation in the number of repeats in different genotypes. SSR markers have several advantages over molecular markers that were previously developed in sunflower. Specifically, they are i) highly informative, ii) exhibit co-dominant inheritance, iii) are locus-specific, and iv) lend themselves to automation for high throughput genotyping. Additionally, SSR databases can be easily maintained and information readily exchanged (Tautz 1989; Morgante and Oliviera 1993; Powell et al. 1996; Cho et al. 2000; Ramsay et al. 2000; Temnykn et al. 2000). SSR markers are useful in a wide range of applications such as genetic mapping, diversity analysis, pedigree analysis, genotype identification and variety protection, germplasm conservation, gene and quantitative trait locus analysis, and marker-assisted breeding (Chen et al. 1997; McCough et al. 1997; Roder et al. 1998; Ramsay et al. 2000).

A few SSR markers have been developed for sunflower by searching DNA sequences in public databases (Brunel 1994; Whitton et al. 1997; Hongtrakul et al. 1998b) or by screening small-insert genomic DNA libraries with repeat-containing oligonucleotide probes (Gedil 1999). These methods of SSR marker development are limited by lack of a large public sequence database and the tedious work involved in screening libraries. Edwards et al. (1996) developed a protocol for enriching genomic DNA libraries for short tandem repeats and this protocol has significantly impacted marker development in crop plants (Roder et al. 1998; Cho et al. 2000; Ramsay et al. 2000).

Our goals (Chapter 2) were i) development of public databases for sequence-based, highly polymorphic SSR markers using a small insert genomic library enriched for a variety of repeat motifs, ii) assessment of the efficiency of SSRs as genetic markers, and iii) analyses of the genetic relationships among sunflower inbred lines using the newly developed SSR markers.

Mapping puts markers in order, indicates the relative genetic distance between markers, and assigns them to linkage groups on the basis of the recombination frequencies from all pairwise combinations (Jones et al. 1997). Mapping methodologies have been intensively developed for crop plants using several different DNA marker types and various mapping populations, e.g. $F_2$, backcross (BC), doubled haploids (DH), and recombinant inbred lines (RILs). Once genetic maps are constructed, they are readily applied to genetic analyses of both qualitatively and quantitatively inherited agronomic traits or practical crop breeding programs.

The first sunflower genetic map was constructed using RAPD markers for *H. anomalus*, a diploid species descended from *H. annuus* x *H. petiolaris.* This map provided broad coverage of the sunflower genome (Rieseberg et al. 1993). RAPD maps have been used for studies of sunflower genome reorganization and gene flow between wild and cultivated sunflower (Arias and Rieseberg 1995; Rieseberg et al. 1995; Linder et al. 1998; Ungere et al. 1998).

The first genetic maps for cultivated sunflower (intraspecific crosses) were constructed using proprietary RFLP markers (Berry et al. 1995; Genzbittel et al. 1995). Since publication of the first RFLP maps, significant numbers of RFLP markers and high-density RFLP maps have been developed (Berry et al. 1996; Jan et

al. 1998; Gentzbittel et al. 1999). The most comprehensive RFLP map was developed by Berry et al. (1996) and integrated 635 RFLP marker loci across nine $F_2$ mapping populations. The map covered 1650 cM and the mean distance between markers was 2.3 cM. Jan et al. (1998) mapped 269 RFLP marker loci from an $F_2$ population, which integrated 20 linkage groups covering 1164 cM. A composite RFLP map across seven $F_2$ segregating populations was developed by Gentzbittel et al. (1999). This map had a mean density of 7 cM and comprised 23 linkage groups with 1573 cM genome coverage. In addition, it incorporated morphological loci and known functional genes related to disease or stress resistance. The three maps were comprised of 1141 RFLP marker loci developed using independent sets of probes from each research group. Restrictions on access to markers and sharing of information among public and private laboratories however, hindered further research and prevented cross-referencing and integration of linkage groups. Berry et al. (1997) released 81 selected cDNA probes distributed across the 17 linkage groups for public use and these RFLP markers were integrated into selected RFLP markers developed by Jan et al. (1998) on a HA370 x HA372 $F_2$ population (Gedil et al. 2001b). The integrated map was 961 cM long with 96 RFLP marker loci dispersed into 17 linkage groups. It was the first and only map that integrated RFLP marker loci from different laboratories. However, the incomplete genome coverage of the integrated map was one factor that led us to develop public SSR genetic linkage maps. One advantage of SSRs is their ability to identify homologous loci across genotypes, thus facilitating cross-referencing of the genetic maps.

Our aims (Chapter 3) were i) the development of a map based only on SSRs, using a recombinant inbred line population, ii) the development of a saturated public SSR map integrated into an RFLP framework map, using an $F_2$ population, and iii) comparison of the efficiency of the two different maps based on marker systems and mapping populations.

As mentioned earlier, one of the main goals of sunflower breeding is the improvement of biotic stress resistance, because sunflowers can be infected by fungal, bacterial and viral pathogens (Gulya et al. 1997). One of the most serious diseases in sunflower is caused by *Sclerotinia sclerotiorum* (Lib.) de Bary. In North America losses can be severe, near 100 % in parts of fields or even entire fields under extreme circumstances (Sackston 1992). The ultimate goal of this thesis research was the development of a sunflower genetic map to analyze quantitatively inherited *Sclerotinia* disease resistance. Although we were not able to complete this analysis due to lack of sufficient markers on the chromosome corresponding to presumed disease resistance loci and lack of time to screen sufficient progeny in the RIL population, QTL analyses of *Sclerotinia* resistance will be greatly aided by the SSR map reported in Chapter 3.

# CHAPTER 2

## ALLELIC DIVERSITY OF SIMPLE SEQUENCE REPEAT MARKERS AMONG ELITE INBRED LINES IN CULTIVATED SUNFLOWER

Ju-Kyung Yu[1], Jodie Mangor[1], Lucy Thompson[2], Keith J. Edwards[2], Mary B. Slabaugh[1], and Steven J. Knapp[1]

[1]Department of Crop and Soil Science, Oregon State University, Corvallis, Oregon, 97331, USA and [2]ICAR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Long Ashton, Bristol, BS18 9AF, UK

**Abstract**

Sequence-based, high-throughput DNA markers are needed for molecular breeding and genome mapping research in cultivated sunflower (*Helianthus annuus L.*), a species with 17 chromosomes. Although several proprietary RFLP and AFLP maps have been developed for sunflower, a dense public map has not been developed and very few sequence-based markers have been developed or mapped. The aims of this research were to develop and assess the polymorphism of simple sequence repeat (SSR) markers in sunflower. We sequenced 970 clones from a genomic DNA library enriched for dinucleotide, trinucleotide and tetranucleotide repeat motifs. This library yielded 176 unique dinucleotide, 60 unique trinucleotide, and 23 unique tetranucleotide repeats. We developed and tested primers for 171 repeats. One hundred and thirty-one primer pairs cleanly amplified SSRs from genomic DNA and 74 SSRs were polymorphic among 16 elite inbred lines. Polymorphic information contents (PIC) ranged from 0.0 to 0.93 with a mean of 0.53 for dinucleotide, 0.53 for trinucleotide, and 0.83 for tetranucleotide repeats. Repeat numbers varied from 4 to 62. PIC and repeat numbers were uncorrelated. Genetic distances among the 16 inbred lines ranged from 0.175 to 0.543. Principal coordinate and cluster analysis of the genetic distance matrix separated lines into three major groups: oilseed B-lines, oilseed R-lines, and confectionery lines. Eighty-eight percent of the polymorphic markers have been mapped in three mapping populations.

**Key Words:** SSR, molecular marker, sunflower (*Helianthus annuus L.*), PIC

## Introduction

Simple sequence repeats (SSRs), or microsatellites, are regions of DNA that contain short (<6 bp) tandemly repeated sequence motifs flanked by unique and conserved sequences. SSRs were first studied and developed as genetic markers in the human (Hamada et al. 1982). Hamada et al. (1982) discovered a long dinucleotide repeat, $(TG)_{25}$, in the intron of cardiac muscle actin gene, λHA-25, in humans (*Homo sapiens L.*). This discovery demonstrated that repetitive sequence motifs are highly repeated in the human genome, and scattered throughout a diverse evolutionary spectrum of the eukaryotic genome (Hamada and Kakunaga 1982; Tautz and Renz 1984). These results brought much attention to the research and development of SSRs as molecular markers in a wide array of organisms, including plants.

Plant SSRs were broadly surveyed by Wang et al. (1994) based on EMBL and GeneBank DNA sequence databases. This survey showed that SSRs were fairly abundant in plants although they seemed less abundant in plants as compared to vertebrates (Lagercrantz et al. 1993). (AT) was the most abundant repeat, followed by (A)/(T) and (GA)/(CT) repeats. (AAT) was the most frequently observed repeat among trinucleotide repeats in plants.

In crop plants, SSRs were first developed as genetic markers in soybeans, which revealed that they are highly polymorphic as a result of variation in the number of repeats (Akkaya et al. 1992; Morgante and Olivieri 1993). Length variation is presumably caused by DNA replication slippage or unequal crossing-over between sister-chromatids (Levinson and Gutman 1987; Jeffreys et al. 1994). Based on various published studies, it has been shown that SSR markers are more informative and

polymorphic than restriction fragment length polymorphism (RFLP) markers in wheat (Roder et al. 1995), barley (Struss and Plieske 1998), rice (Ishii and McCough 2000) and potato (Ashkenazi et al. 2001).

There are important technical reasons that make SSRs ideal molecular markers. Because primers for SSR markers can be designed in flanking sequences, which are unique and conserved sequences, these markers can be assayed with the PCR (Weber and May 1989). Second, SSR markers can be maintained and shared easily because only sequence data are needed to describe a marker. Third, automation and multiplexing by employing several primer pairs simultaneously are possible (Mansfield et al. 1994). These outstanding features make SSRs the molecular markers of choice for a wide range of applications. High allelic variation of SSRs makes these markers a powerful tool for germplasm fingerprinting of closely related genotypes in soybean (Wang et al. 1998), barley (Struss and Plieske 1998), and potato (Ashkenazi et al. 2001). In addition, SSR linkage maps have been rapidly developed and integrated into existing the amplified fragment length polymorphism (AFLP) framework map in barley (Ramsay et al. 2000), or existing RFLP framework maps in rice (Cho et al. 1998), sugar beet (Rae et al. 2000) and wheat (Roder et al. 1998). Once assigned to genetic linkage maps, SSRs are markers of choice for marker-assisted selection because of their ease and efficiency of assay.

SSR markers have been developed by several methods. The development of markers from public sequence database searches is limited by the amount of sequence data available and thus does not permit rapid development of new markers (Brunel 1994). A hybridization-based technique called oligo-fingerprinting, although

reproducible, was not practicable for screening large number of genotypes owing to its time and labor requirement and the frequency of false-positive hybridization signals (Dehmer and Friedt 1998). The polymerase chain reaction (PCR)-aided enriched library strategy, which has been recently used widely to develop SSRs, has the potential to develop the greatest returns in the shortest time (Edwards et al. 1996).

Dinucleotide motifs are frequently used to enrich SSR libraries because of their abundance in genomes and high allelic variation. Dinucleotide repeats appear to have mutation rates 1.5 to 2 times higher than tetranucleotide repeats, and non-disease-causing trinucleotide repeats have mutation rates intermediate between di- and tetranucleotides in humans (Chakraborty et al. 1997). Previous studies have found that dinucleotides are more polymorphic than trinucleotides in several plant species, such as wheat (Roder et al. 1995), barley (Struss and Plieske 1998), and rice (Ishii and McCouch 2000).

The library utilized in this study was enriched for (CT) and (CA) sequences even though (AT) and (A)/(T) repeats are the most abundant in plants. The self-annealing property of (AT) repeats can lead to the formation of secondary structures in DNA molecules and failure of PCR reactions.

The cultivated sunflower (*Helianthus annuus L.*), an annual diploid crop with x = 17, is one of the major edible oil seed crops in the world, ranking second in importance after soybean. Several molecular markers have been developed in sunflower for genetic studies and practical breeding programs. However, compared to most major crops the application of markers to sunflower has been extremely limited because molecular breeding resources are much less developed. A relatively large

number of RFLPs have been generated (Gentzbittel et al. 1992, 1994; Berry et al. 1994; Zhang et al. 1995; Jan et al. 1998). However, there is a significant limitation in assessing the utility of RFLPs because most RFLP markers were developed using independent sets of DNA probes from proprietary domains. Therefore, access to probes for RFLP markers or sharing of marker information has been restricted, slowing the progress of sunflower genetic research. To overcome these drawbacks, the sunflower research community needed to develop a critical mass of public domain DNA markers. A few SSRs as sequence-based markers have been developed in sunflower based on cDNA sequences in public databases (Brunel 1994; Whitton et al. 1997; Hongtrakul et al. 1998) and sequences of clones isolated by screening small-insert genomic DNA libraries (Gedil 1999). They found an abundance of repeat sequences and different levels of allelic variation in sunflowers. In sum, previous methods of SSR marker development were limited by the lack of public sequence database and the tedious work involved in screening DNA libraries. In this study, the enriched small insert genomic library developed by Edwards et al. (1996) was used to develop SSR markers for genetic analyses of cultivated sunflowers. The objectives of this study were as follows i) to develop a public collection of high-throughput and highly polymorphic SSR markers, ii) to assess the efficiency of SSRs as genetic markers in sunflower, and iii) to determine the genetic relationship among sunflower elite inbred lines using SSR markers.

# Materials and Methods

## Library Construction and Screening

Genomic DNA was isolated from the sunflower cultivar HA89. DNA libraries were produced from genomic DNA selectively enriched for CA, CT, CAA, CATA, and GATA repeats using methods described by Edwards et al. (1996). One μg of genomic DNA was digested with *Ssp*I or *Rsa*I and hybridized to filter-immobilized CA, CT, CAA, CATA, and GATA oligonucleotides to enrich for genomic fragments harboring SSRs. The selected fragments were amplified by PCR and ligated into a modified pUC19 vector (pJV1). Plasmids were transformed into the *E. coli* strain DH10B$^{TM}$ (Life Technologies, Gaithersburg, MD, USA) and plated on L-agar plates containing 100 μg/mL ampicillin. The bacteria were incubated overnight at 37°C. Single colonies were picked and transferred to 96-well microtitre plates. The picked colonies were incubated overnight at 37°C. Glycerol was added to a final concentration of 25% and the plates were stored at -70°C.

The genomic libraries were screened for the presence of SSRs by spotting the clones onto nylon membranes and probing with radiolabeled oligonucleotides (CA, GA, CAA, CATA, and GATA) (Sambrook et al. 1989). A Beckmann Biomek 2000 was used to replicate 1,536 colonies onto an 8 x 12 cm membrane, Hybond$^{R}$ N$^{+}$ (Amersham Pharmacia Biotech, Picastaway, NJ, USA). The membrane was placed on L-agar containing 100 μg/mL ampicillin and incubated for 16 h at 37°C. The membranes were transferred onto 3 mm Whatman paper, soaked in denaturation

buffer (1.5 M NaCl, 0.5 M NaOH) for 5 min, and then twice onto paper soaked in neutralization buffer (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.5) for 3 min. The membranes were washed vigorously in 2x SSC (0.3 M NaCl, 0.03 M trisodium citrate) for 5 min. After air-drying, DNA was fixed by crosslinking to the membrane using the optimal cross-link function of a Spectrolinker XL-1000 UV Crosslinker (Spectronics Corporation,Westbury, NY, USA). $(CA)_{15}$, $(CT)_{15}$, $(CATA)_{10}$, $(CAA)_{10}$, and $(GATA)_{10}$, were each end-labeled using [$\alpha$-$^{32}$P] dATP (DuPont-NEN, Boston, MA, USA) and T4 polynucleotide kinase (Amersham Pharmacia Biotech, Picastaway, NJ, USA). Membranes were pre-washed with hybridization buffer (6x SSC, 0.25% dried milk powder, and 0.01% SDS) and incubated in a OV5 (Biometra, Gottingen, Germany) rotary oven at 50°C for 1 h. The wash buffer was replaced by 25 mL hybridization buffer containing 100 ng of the radiolabeled oligonucleotide and incubated at 50°C for 16 h. Membranes were washed 4 times for 5 min at 50°C with 200 mL of 2x SSC containing 0.1% SDS, then air dried and exposed to X-ray film overnight.

**SSR Marker Development and Screening**

Amplified PCR products produced with universal M13 primers from individual colonies were purified using a QIAquick PCR purification kit (Qiagen, Valencia, CA, USA) and sequenced on an ABI 377 automated DNA sequencer (Applied Biosystem, Perkin-Elmer, Foster City, CA, USA) in the Center for Gene Research and Biotechnology (CGRB) at Oregon State University. The sequences

were aligned with the PileUp function of the GCG software package (Madison, WI, USA) to detect possible duplicates.

Primers were designed to sequences flanking repeats using PRIMER (Lincoln et al. 1991) or PRIMER PREMIER (Premier Biosoft International, Palo Alto, CA, USA) software. To establish a high-throughput system, all primers were chosen under the same criteria: 57°C – 61°C melting temperature, 40% - 60% GC content, 18 –22 bp primer length and 90 bp to 500 bp PCR product size. Also, forward primers were labeled with either blue (6FAM), green (TET), or yellow (HEX) fluorescent tags (Applied Biosystem, Perkin-Elmer, Foster City, CA, USA). Primers amplifying over-lapping fragment sizes were labeled with different fluorescent dyes. The sequence, 5'GTGTCTT3', was added to the 5'end of reverse primers to enhance the non-templated addition of adenine at the 3' end of PCR products (Brownstein et al. 1996).

PCR reactions were carried out in a 20 uL solution containing 20 ng of genomic DNA template, 5.5 pMol of each primer, 2.5 mM $MgCl_2$, 0.125 mM of each dNTP, 1X Qiagen reaction buffer, 0.8 unit of Qiagen Taq polymerase and 0.1% Tween 20. After one denaturing step of 3 min at 94°C, a touchdown amplification program was performed (Don et al. 1991) on a Thermal Cycler 9600 (Perkin-Elmer, Foster City, CA, USA). This profile included a denaturing step of 30 s at 94°C and an extension step of 45 s at 72°C. The initial annealing step was 30 s at 64°C for one cycle and subsequently was dropped by 1°C for every cycle until a final temperature of 54°C was reached. The annealing temperature of 54°C was employed for the last 30 cycles of the amplification, followed by one cycle of 72°C for 10 min. PCR products were visualized on 1.5% agarose gels.

SSRs were screened for length polymorphisms among 16 sunflower lines: five public oilseed sterility maintainer (B) lines (HA89, HA821, HA370, HA372, and HA383), three public oilseed fertility restorer (R) lines (RHA377, RHA274, and RHA 801), one public confectionery B-line (HA292), one public confectionery R- line (RHA280), one proprietary oilseed B-line (PH-D), and five proprietary oilseed R-lines (PH-A, PH-B, PH-C, PH-E, and PH-F). DNA was isolated from the leaves and stems of three-week old greenhouse grown seedlings using a modified CTAB method (Webb and Knapp 1990). DNA concentrations were estimated using a Versafluor$^{TM}$ fluorometer (Bio-Rad Laboratories, Hercules, CA, USA).

For multiplexing, 2 uL of PCR product labeled with 6FAM or TET, and 6 uL of product labeled with HEX were mixed in a final volume of 50 uL with distilled water. A total of 1.3 uL master mix including 0.8 uL formamide, 0.3 uL blue dextran dye (loading buffer) and 0.2 uL size standard (GeneScan 500 TAMRA) were mixed with 0.5 uL multiplexed sample. One point five micro-liter out of 1.8 uL pooled samples were loaded and separated on the polyacrylamide gel using filter set C on the ABI Prism 377 DNA Sequencer (Applied Biosystem, Perkin-Elmer, Foster City, CA, USA) in the Center for Gene Research and Biotechnology (CGRB) at Oregon State University. GeneScan (version 2.1) and Genotyper (version 2.0) software (Applied Biosystem, Perkin- Elmer, Foster City, CA, USA) were used for automated data collection and, accurate visualization of the alleles and computation of allele size, respectively.

**Data Analyses**

Polymorphism information content (PIC) scores were estimated for each locus using the estimator described by Botstein (1992). Because we fingerprinted inbred lines, PIC estimates the probability of observing a polymorphism between two sampled inbred lines randomly drawn from the sample of 16 inbred lines (Anderson 1993). The mean number of alleles per locus was estimated with and without monomorphic markers (A and $A_P$, respectively). Mean PIC scores were estimated with and without monomorphic markers (PIC and $PIC_P$, respectively). Genetic distances were estimated using the the allele sharing estimator (*D*) described by Bowcock et al. (1994) and Goldstein et al. (1995). *D* was estimated using the computer program MICROSAT (Minch et al. 1997; http://hpgl.stanford.edu/projects/microsat/) and polymorphic loci only.

Principal component analysis was performed on the genetic distance matrix using the PROC PRINCOMP routine of SAS (version 8.0). Principal scores were estimated for the first three principal components for each inbred line and plotted. Cluster analysis was performed on the genetic distance matrix using the UPGMA clustering algorithm of NTSYSpc (version 2.0). The dendrogram was produced using the NTSYSpc (version 2.0) TREE PLOT function. The goodness-of-fit of the estimated distance matrix (for clusters) to the original genetic distance matrix was tested using the Mantel (1967) Z-test. The test was performed using the NTSYSpc program MXCOPM.

# Results

## SSR Marker Development and Screening

The inserts from 1,342 genomic DNA clones were amplified by PCR; 970 clones yielded single, discrete PCR products and were sequenced. SSRs were present in 632 sequences (65.2% of clones sequenced). Sequences harboring common motifs were aligned to identify duplicate sequences, identifying 259 unique sequences (26.7% of the clones sequenced). The DNA sequences of clones harboring unique SSRs have been deposited as supplemental data (www.nrc.ca/genome). The complete set of DNA sequences has been deposited in the Sunflower Genome Database, sunDB (www.css.orst.edu/knapp-lab/sunflower).

We recovered 171 perfect repeats, e.g. $(CA)_n$; 36 simple imperfect repeats, e.g., $(CA)_nN(CA)_n$; and 52 compound repeats, e.g. $(CA)_nN(GA)_n$ $(CA)_nN(GA)_n$. Of the total recovered, 170 were dinucleotide, 60 were trinucleotide, and 23 were tetranucleotide repeats. $(CA)_n$, $(CT)_n$, $(AT)_n$, and $(GC)_n$ were recovered. Ninety-five percent of the dinucleotide repeats were $(CA)_n$ and $(CT)_n$ dinucleotide repeat motifs used to enrich the libraries. Six trinucleotide repeat motifs were recovered: (CGG), (CAA), (CCG), (AAG), (ACC), and (AGG). The library was enriched for (CAA). One tetranucleotide repeat motifs were recovered: (CATA). The library was enriched for (CATA) and (GATA).

Flanking oligonucleotide primer pairs were designed for 171 SSRs (66% of the unique SSRs) and screened for functionality on agarose. Primers could not be

designed for 98 unique SSRs because of insufficient flanking sequences. Of the total, 131 primer pairs produced single, discrete amplicons and were screened for length polymorphisms on polyacrylamide gels using fluorescently labelled amplicons (Table 2.2). The 40 primer pairs that failed either produced no amplicons or non-specific or complex products.

Seventy four SSR markers (56.5%) were polymorphic among the 16 inbred lines (Table 2.1). The most polymorphic marker, a tetranucleotide repeat (ORS256), had 15 alleles and a PIC score of 0.93. The mean number of alleles per locus among the polymorphic SSR markers ($A_P$) was 3.7 for dinucleotide, 3.6 for trinucleotide, and 9.5 for tetranucleotide repeat markers. Tetranucleotide repeats were significantly more polymorphic than dinucleotide and trinucleotide repeats (Fig. 2.1 and Table 2.1). The $PIC_P$ score ranges for polymorphic markers were 0.18 to 0.88 for dinucleotide, 0.18 to 0.70 for trinucleotide, and 0.62 to 0.93 for tetranucleotide repeats (Fig. 2.1). Mean $PIC_P$ scores were 0.53 for dinucleotide, 0.53 for trinucleotide, and 0.83 for tetranucleotide repeats. The $PIC_P$ score for AC repeats (0.58) was significantly greater than the $PIC_P$ score for AG repeats (0.44). Fifty-five SSR markers (42.0%) amplified two or more bands and 12 SSR markers (9.2%) produced null alleles. The number of repeat units ranged from four to 62 in the reference allele sequences (Fig. 2.2). The mean number of repeat units was 13.4 for dinucleotides, 7.7 for trinucleotides, and 25.8 for tetranucleotides (Fig. 2.2). Among SSR markers developed from reference alleles longer than four repeat units; the mean repeat lengths were 26.8 bp for dinucleotides, 23.1 bp for trinucleotides, and 103.2 bp for tetranucleotides.

**Table 2.1** The number and percentage of SSR markers developed and polymorphic SSR markers, polymorphic information content scores for monomorphic and polymorphic SSR markers (PIC) and polymorphic SSR markers only ($PIC_P$), and the mean number of alleles per SSR marker (A) and per polymorphic SSR marker ($A_P$) for 16 inbred lines of sunflower for different repeat classes. The percentage of functional markers was calculated by dividing the number of primers tested by the number of SSR markers developed. The percentage of polymorphic markers was calculated by dividing the number of polymorphic SSR markers by the number of SSR markers.

| Class | Primer Tested | SSR Markers Developed | | Polymorphic SSR Markers | | PIC | $PIC_P$ | A | $A_P$ |
|---|---|---|---|---|---|---|---|---|---|
| | | Number | % | Number | % | | | | |
| Compound | 40 | 31 | 77.5 | 17 | 54.8 | 0.26 | 0.48 | 2.45 | 3.41 |
| Imperfect | 24 | 17 | 70.8 | 12 | 70.6 | 0.37 | 0.52 | 3.05 | 3.67 |
| Perfect | 107 | 83 | 77.6 | 45 | 54.2 | 0.32 | 0.59 | 3.02 | 4.29 |
| AC | 53 | 45 | 84.9 | 34 | 75.5 | 0.44 | 0.58 | 3.40 | 3.86 |
| AG | 21 | 15 | 71.4 | 5 | 33.3 | 0.15 | 0.44 | 1.87 | 3.00 |
| Dinucleotide | 114 | 91 | 79.8 | 54 | 59.3 | 0.31 | 0.53 | 2.78 | 3.67 |
| Trinucleotide | 43 | 31 | 72.1 | 14 | 45.2 | 0.24 | 0.53 | 2.32 | 3.57 |
| Tetranucleotide | 11 | 7 | 63.6 | 4 | 57.1 | 0.47 | 0.83 | 6.43 | 9.50 |
| Total | 171 | 131 | 76.6 | 74 | 56.4 | 0.31 | 0.55 | 2.89 | 3.99 |

**Table 2.2** SSR marker names, repeat classes, size of the reference allele (HA89), and oligonucleotide primer sequences for amplifying SSRs.

| Marker | Repeat | Size | Forward Primer | Reverse Primer |
|---|---|---|---|---|
| ORS-121 | (ca)13 | 241 | caatctctgatttcccggaa | caaaaatagccggtgaagga |
| ORS-123 | (tc)17 | 147 | gaaaacccatgcaggcatac | acatccatcacagtccattttg |
| ORS-124 | (ac)14 | 252 | aatcgccataccactccatc | gatatcaccccacgataacatg |
| ORS-125 | (gt)5 | 252 | accgaaccacgtaaatctcg | gagcaaagctgcgaaactg |
| ORS-126 | (gt)22 | 293 | cactgtcccttctggtagttcc | ttcccacgcaaacttcaattcc |
| ORS-127 | (gt)15 | 159 | tcccggcccatatttacata | gttcgcacagacgggatc |
| ORS-128 | (ca)13 | 378 | gaccgtccacgtgtcagc | caaaatagcgttgacgagca |
| ORS-131 | (ct)13 | 465 | ccctcatcacatcccagc | ctcggtgtctagggtttcga |
| ORS-132 | (ca)15 | 416 | ctccgcagttataacaacctcc | gaaaacataccccgatgcac |
| ORS-133 | (tg)13 | 145 | aaagctcgctttttgtgcat | tgatgatgatccagaagaccc |
| ORS-134 | (ac)16 | 295 | taaagatgttgaggtcctgaatcg | tacgtgtacgtatgcatatgag |
| ORS-135 | (ac)4 N1(ca)7 | 184 | caaaatggagaacaaagctccc | tattgccggtccgacatcga |
| ORS-138 | (ac)4 N2 (tg)22 | 192 | tgagcctcgattcatccttc | acatgagggaccaaagtggg |
| ORS-139 | (tc)18 N6 (ac)9 | 149 | ccgagtgtccacgaaccacc | acaccccagatccctctttagc |
| ORS-140 | (at)7 (ac)14 | 321 | tgcgttatctagatcacttggg | agtgtcagagttcatggcttcg |
| ORS-141 | (ag)7 | 282 | gtcttctctcctccggcag | acgaagtcaacagacattccg |
| ORS-142 | (ca)13 (ta)6 (ag)6 | 235 | atgtgggatcccatttcaaa | gcgtacgcaacgtacgttaa |
| ORS-143 | (ag)8 N5 (ag)8 N7 (tg)18 | 274 | tcagggcaatagtagggctg | tatgcaggcatctgcaagag |
| ORS-144 | (ac)4 N2 (ac)14 (at)4 N2 (at)3 | 312 | ctcgaggtggcaatcaaggc | ctgacggctgagatgatcgc |
| ORS-145 | (ca)11 N2 (ca)3 | 171 | tttgtataggcagtttgtgttg | ctgcttccagtgagaccct |
| ORS-146 | (gt)4 N2 (gt)17 | 242 | cccatttttgttcttagggagg | ttcaagctttaacatgacatgc |
| ORS-147 | (cg)6 N2 (gt)27 | 294 | ggttccgcttggttgggtag | gggcctgcgaattgttcc |
| ORS-148 | (ac)14 | 256 | atccgtatattgccttgttgag | agaagccacgcaattgattatc |
| ORS-149 | (ca)15 | 137 | gctctctatctcccttgactcg | tgctctaagatctcaggcgtgc |
| ORS-151 | (tg)7 N2 (tg)7 | 454 | tcacatccatcagccataaca | tgtgttcgagtttcggtttg |

**Table 2.2** continued

| | | | | |
|---|---|---|---|---|
| ORS-153 | (gt)14 | 145 | ttgtccgtacaccacccac | ttgagttggaagaatcacatgc |
| ORS-154 | (at)5 (gt) 7 N2 (gt)9 | 203 | gcacctttggtgaggagata | tgcatcagtagctattgtctat |
| ORS-157 | (gt)22 (gc)5 | 258 | gccgcaaccggctgtaaatc | tgtgcagggattaaactcgtgg |
| ORS-158 | (ca)12 | 247 | tttggttttagccagaatttca | gcacgcccgatgatacctc |
| ORS-159 | (ac)8 N8 (ac)6 | 351 | caaagcttgccctaatctgtgc | gttcctgtcgggtaactgtagc |
| ORS-160 | (ac)6 N2 (ac)11 | 177 | tcccttcctttcatcgtctgct | tggcaatttgccaaggacc |
| ORS-161 | (ac)12 | 164 | ccaggtacgctcgatgggatag | ccactaggtgcgttgggttg |
| ORS-163 | (ca)4 N2 (ca)20 | 386 | atggttaagttacacacccgct | gctatcaacgccaccaccac |
| ORS-165 | (ct)8 N9 (ac)14 | 238 | aattgacctagcgagttcttcc | cagggaggctccaaaggg |
| ORS-166 | (ac)17 | 344 | cagccacatgccctctgac | tgttaagaaccgcgacaactgc |
| ORS-167 | (ca)14 | 365 | cgagccttctaataaaagcagc | agagcaaacagattgggacaac |
| ORS-168 | (ag)6 | 290 | ctgtgtcatttccgatctgttc | tatgcggtcttcgttctactgc |
| ORS-169 | (ca)8 N34 (ca)8 | 198 | tggaactgtaaatggacccaag | gcactgcaccatttatgagaag |
| ORS-170 | (tc)7 | 116 | gcagtgccactcatttctgttg | tgtgatgattgcaggattggag |
| ORS-171 | (gt)12 | 412 | ggggattaaatgatgagggact | cattagttgacatgcaagtggg |
| ORS-172 | (ta)8 (tg)12 | 201 | catgcgtgattatggttttgag | aaattctggatcggtattaggc |
| ORS-173 | (ac)14 | 186 | aacaaatcaccacggaaccaac | gctaattcggtttggtggcttg |
| ORS-174 | (gt)16 | 233 | cacacacatctggtgttttgg | ttgggttgaaacccactttc |
| ORS-176 | (tg)16 | 462 | ccctaactggttttctgaccc | aacttttgtttgtttgtccagg |
| ORS-177 | (ca)8 N10 (ca)7 | 333 | agggccttcaaagtggaaat | tgtaacgtgtgtgtgtgtaacg |
| ORS-178 | (ca)11 | 328 | gctctgatgttgatgaatcagc | gagcccaacattccaaaaaa |
| ORS-179 | (tg)15 | 122 | ggaaggtggtggggtaatg | tacacacacacacacacacaca |
| ORS-180 | (ac)6 N10 (ct)14 | 180 | gagcaaacactgtcacgtga | gagcgtgctaccaggacaat |
| ORS-183 | (ta)5 | 153 | aacgtgggattctaagttaaga | tacatacgtgcacacactttcg |
| ORS-184 | (ta)5 (tg)7 N2 (tg)7 N19 (gt)5 | 158 | cgactgagatcgcgtacatattg | ttagattaacgacaatggctc |
| ORS-185 | (gt)4 (ga)4 N1 (ag)4 N1 (ga)10 | 306 | agccgctcctaacttgaacca | tcacccttaaacatcacccacc |
| ORS-186 | (tg)12 (ta)6 | 103 | ctgaggctttatttttgggg | aattttgtaaccgaaccccac |

**Table 2.2** continued

| | | | | |
|---|---|---|---|---|
| ORS-187 | (ta)8 (ca)9 (ta)3 (ca)9 | 267 | ccgtaatgtacgtcaaccttcc | acacacccacaccatttagagg |
| ORS-188 | (gt)18 | 139 | cttcgtagccaactcccacc | caatggttgacaatgggtttgc |
| ORS-189 | (ac)14 | 342 | ccttgtgctatcatttgatatg | taccgaaagtaacatgagcttg |
| ORS-190 | (ac)11 | 377 | ccgcatcaagccaccaag | agaaggattaagcatcaggcac |
| ORS-191 | (ca)12 N6 (ca)9 | 200 | actgcgtttgtgattactggtg | catgcactgaagacatacaccc |
| ORS-192 | (gt)7 N12 (gt)4 N3 (gt)5 | 149 | agtcgatcaccggagctttg | cccaacaaaccaaaccctacct |
| ORS-193 | (tc)5 N2 (tc)5 N24 (ac)7 | 167 | aagggctttgagtctttggc | gctctcctggtccacccac |
| ORS-194 | (ac)15 (at)7 | 398 | ttaattgttcgattgtcttagc | tacatgtccgtcacaagaactc |
| ORS-195 | (ac)7 (atac)5 (ac)22 (atac)10 | 335 | cgttgtgttgtgtttgttagtc | acttaaagttcttgagccactg |
| ORS-197 | (ac)22 | 315 | ttcacgaggtcgtctgacaagg | aacagagggtactccaagaacc |
| ORS-199 | (ac)12 | 101 | tcgcacacttaaacgacctctc | atgtgggtgaatcagctggcag |
| ORS-200 | (ca)19 | 332 | cgcactgccttaaaccctc | actcttgattgaatgatgctcc |
| ORS-201 | (at)5 N2 (gt)7 N3 (tg)10 | 238 | ccagaaagcgatgttgagtttg | cgatcgaccgactagggag |
| ORS-202 | (ca)17 | 417 | aatcttccgcaggtttgattg | tgtgaacggtaaatatgatgct |
| ORS-203 | (ac)4 N11 (ca)5 N2 (ca)5 | 264 | gcccaagatgtgaagcgaatg | gtcagaacaggaccgaaccact |
| ORS-204 | (gt)17 | 312 | cgtctggcattatgaaatcgtc | ccgcataacagcaatggtcaac |
| ORS-209 | (ta)4 (tg)11 | 358 | cacatgcttgatgaagatagcg | aacattgcaacaatcgcttatg |
| ORS-210 | (ta)7 (tg)18 | 195 | tgatttcatgggtcctaacttg | gttctcacatgaacctaccctc |
| ORS-211 | (ct)6 N1 (tc)8 N4 (ac)14 | 199 | aaagtaggaacaaccgaaatgc | tggtaatggtagcgcagacg |
| ORS-212 | (tc)6 | 132 | ccgtctgcgctaccattacc | ataaacttccagcagcggttag |
| ORS-214 | (gt)18 | 315 | catccctaaccgccaccac | gcttccgctgaacttaaacatg |
| ORS-215 | (ca)14 N2 (ca)7 (ta)5 | 350 | tggttctcaccagcagtttagg | cctctgctgattgaatggattg |
| ORS-216 | (ct)12 N7 (tc)4 N2 (tc)6 | 174 | cttcctccaccctcaagcg | tccctaatgtaccaccaccatc |
| ORS-218 | (ct)5 | 258 | ttgggatgaaacaaagaccacc | ttttgctcagaggaagagtccg |
| ORS-219 | (ag)7 N18 (ga)18 | 154 | cacaaaactcgaaaccggactc | gagagtgttggcagcccatc |
| ORS-220 | (ag)10 | 329 | tcaaacacaaatggaggtcagg | ataagatagcaccgccgccac |
| ORS-221 | (tg)8 N4 (tg)8 N4 (tg)6 N1 (ag)5 | 423 | aacatgcgtgattatggttttc | attctgcacttttccgttaagc |

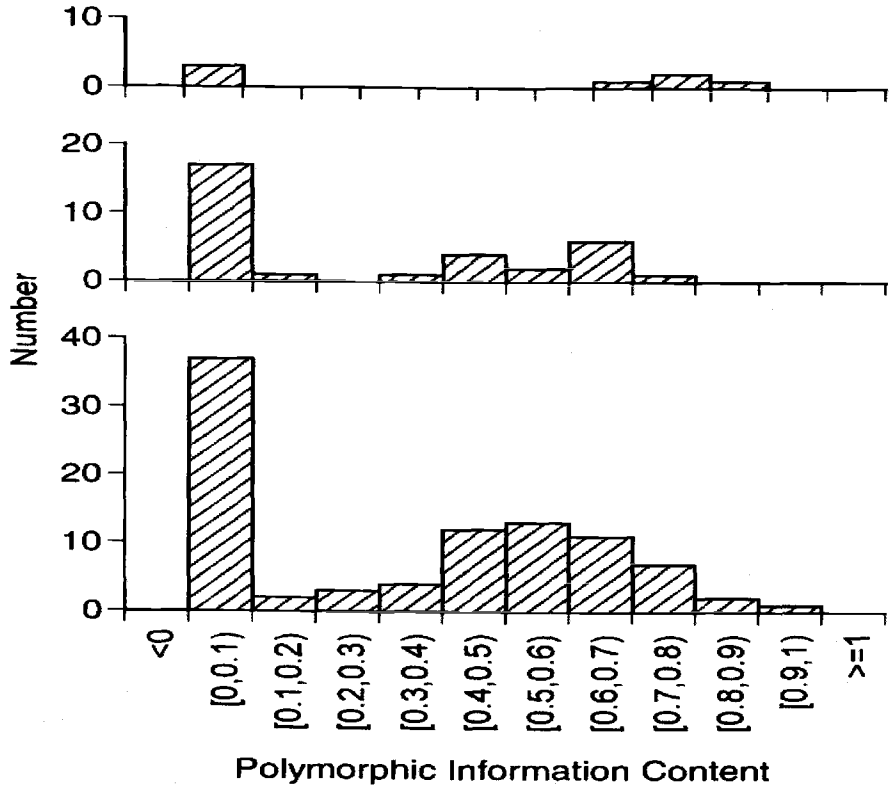
**Table 2.2** continued

| | | | | |
|---|---|---|---|---|
| ORS-222 | (ca)13 | 340 | aattgagcttcaatttggtgga | atccgtgcgaattaaccatcag |
| ORS-223 | (ta)4 N2 (ac)6 N2 (ac)7 | 312 | caaaggtgagaatcatttgcg | aatgaccattcttgcccttcc |
| ORS-224 | (ga)8 N2 (ga)13 | 136 | aaccaaagcgctgaagaaatc | tggactaactaccagaagctac |
| ORS-225 | (ac)4 N1 (ac)11 N4 (ct)4 | 198 | tctgcaggagaagatgtggacc | tgctttaaggtggaggaggtgg |
| ORS-226 | (tc)8 | 197 | tcgccaccacctcctccac | gaagtccggctggtttacatg |
| ORS-227 | (cgg)6 | 417 | gggttttgaatccagtggca | aacccacaactggacaatgacc |
| ORS-228 | (cgg)5 | 154 | tcatcggcggtgttggag | cgatctccaggaccgaacc |
| ORS-229 | (caa)12 | 176 | tccgacccgaatcttatgaacc | gacccgaatgagacccaaactg |
| ORS-230 | (cgg)5 | 359 | gccaccgctacccgttac | tccggtgttgtttatgtgaagc |
| ORS-232 | (gcc)5 | 314 | cataatcctccgccaccacc | gccctctgctcctgagttcttc |
| ORS-233 | (ggc)6 | 396 | gtagtggtcgtggtagtggtcg | ggcctgctgtttgctgtctg |
| ORS-235 | (caa)9 | 138 | aagcaactgccgctcccac | agcgacagctgtgacaatgc |
| ORS-237 | (gtt)22 | 192 | caaggtctgtctacatcccacc | gctgtaaagcctgcatatcctc |
| ORS-238 | (gcc)5 | 260 | gctgcagcagccaatggaac | gattctgctgcccttgtggag |
| ORS-239 | (cac)6 | 179 | atggaccaccgagacctatg | gtcactcagttcagttactggg |
| ORS-240 | (gcg)6 | 259 | ggtgatgatggaggagcaactg | cactcaaccattgttctcccac |
| ORS-242 | (ggc)5 | 123 | atccgtcagcacgagttcttg | cctaacaccaccacaatccacc |
| ORS-243 | (ggt)7 | 170 | gggatgacgtgcgtttgg | accaccatttctaccgtttctc |
| ORS-244 | (ggc)5 | 233 | aggtgaatcaacgagtgaatgg | caccaccaccgccgtctc |
| ORS-245 | (tcc)7 | 343 | atcaccatcaccaaccctcatc | tacaatcgtaaccgccgacttc |
| ORS-246 | (aac)12 | 292 | caagcaatcgagtcgaacc | cgaaaagtgtaatttacccttg |
| ORS-247 | (atgt)42 | ~600 | atggaatccataggtaagcatg | cctacttcgtgcattcgtacac |
| ORS-249 | (cata)6 | 129 | cctaattacatccattcctacc | gtaggtttgtgtgaaggcag |
| ORS-252 | (gtat)43 | ~600 | gcgcatacgcacactactcaac | tttgcgttttggtgggctagac |
| ORS-253 | (tgta)6 N1 (gtat)3 | 166 | tgtaggtttgtgtgaaggcatg | catacctacgtacgtgcatacc |
| ORS-254 | (taca)25 | 386 | aaatcccacttcatacaaacgt | ccttcagtgctcatgcagtg |
| ORS-256 | (tgta)11 | 183 | gattaggcatgtatgagttagg | ctacgtataaactgccttctcg |

**Table 2.2** continued

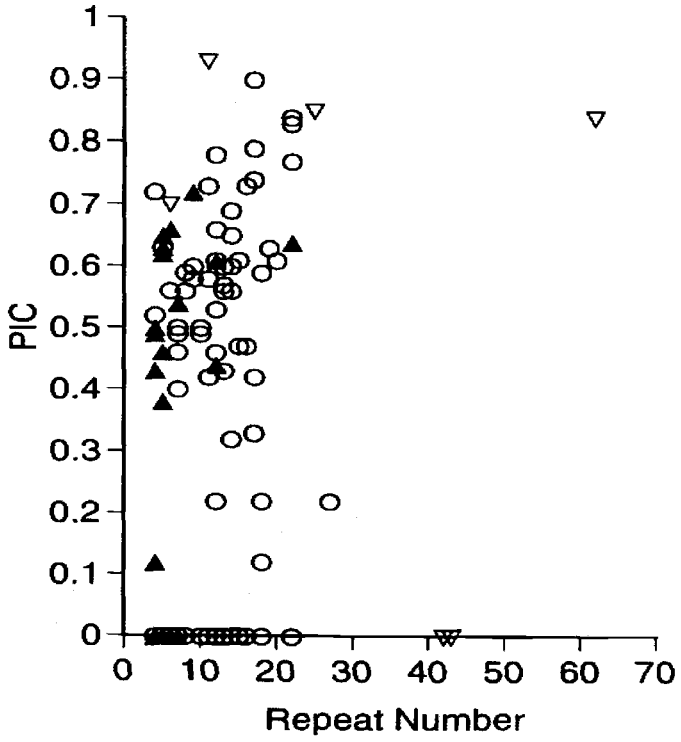
| | | | | |
|---|---|---|---|---|
| ORS-257 | (tgta)62 | ~650 | gtgactacgttatggatgcatg | gcctttgcttgcatatctacg |
| ORS-258 | (ct)4 N52 (ca)4 | 332 | ggcccgattacaagataacg | ttgcgtccgatgctgttc |
| ORS-260 | (ag)4 | 250 | ttgcctacgtcaatcaagttcg | caccgtgaatcacctttctctc |
| ORS-261 | (ggt)3 N18 (cg)4 | 199 | ctgttccgttcgtcagaaactc | agcgaaaggatcgagaatcatc |
| ORS-263 | (ta)4 | 272 | acccatcctacgccacttgtc | cagcagcggaagggtgttatg |
| ORS-266 | (ga)4 N56 (ga)3 N14 (ag)4 N36 (ga)4 | 284 | tgtcgacgtaacggagagc | tctccaattaatgacaccagag |
| ORS-267 | (ct)4 | 183 | cgtcggcaaccacattcg | gaagatttggcgacactacctg |
| ORS-268 | (ga)4 | 320 | ccagtctgcgtcagtgtagagc | ctggcggaaactaggttgattg |
| ORS-270 | (ac)3 N9 (ca)4 | 417 | tggatcacccatttcatcactc | attcaagacccacctccatctc |
| ORS-271 | (caa)4 | 498 | cgaacatgtcgtgcgatactg | gcaggagctgcttggttacc |
| ORS-272 | (cgg)4 | 126 | gtggccggaatctagctgacc | gccgcaccaaatgcacaac |
| ORS-274 | (gcc)4 | 164 | ggaaacctagtgtcgcttccg | aaccgactgttagggtttggac |
| ORS-275 | (aag)4 | 202 | aggaggagttgcgagcagtgg | gtcaagagcccatccctgacg |
| ORS-277 | (ggc)4 | 101 | gatgctgcccggctgaag | tccttctccaccacatctcctc |
| ORS-278 | (gga)4 | 312 | ggaggtggaggacgagagtg | ccaaatgccctacctcaccac |
| ORS-279 | (ccg)4 | 428 | ctattgggcttgtctacttggg | cttgagaccagcaccagctttc |
| ORS-281 | (cgc)4 | 386 | ttaacccggctaaccttctgc | actccacctcgcacatacatcc |
| ORS-285 | (cgg)4 | 231 | agcacggcggagggaatc | cgcattctagcatcacccatcc |
| ORS-286 | (gcc)4 | 189 | catcggtgtcttcgtaatggtc | cggcgagatccgtacatagttc |
| ORS-288 | (gcg)4 | 104 | ggttggaaaccctaatctgg | cgcagccacctaacccag |
| ORS-289 | (ccg)4 | 265 | acacttgaccttcaccaccatc | tatgaccatgattacgccaagc |
| ORS-291 | (cgg)4 | 417 | ccttggggaccgaacgaaac | taatcagattccatcggcggag |
| ORS-292 | (ccg)4 | 110 | ccagatctaacggttgctactg | cggtggttacggtgggag |
| ORS-293 | (cca)4 | 285 | ggctgttcaggctatggctc | ttcggaagtaaggcggtagatg |

**Figure 2.1** Distribution of polymorphic information contents for 92 dinucleotide (bottom), 32 trinucleotide (middle), and 7 tetranucleotide (upper) markers genotyped for length polymorphisms among 16 inbred lines of sunflower.




The correlation between the number of repeat units in the reference allele sequence and PIC score ($r = 0.20$) was not significantly different from 0.00 ($p = 0.09$) (Fig. 2.2). The weak correlation prompted us to design and test primers for 23 unique dinucleotide and trinucleotide repeats comprised of four repeat units based on the reference allele sequence. (We originally restricted marker development to repeats comprised of a minimum of five repeat units in the reference allele sequence.). Seventeen of the 23 $n = 4$ markers tested were monomorphic; thus, while functional markers were developed for 100% of the short repeat SSRs, the yield of polymorphic SSR markers from such sequences was low (26%).

**Figure 2.2** The number of repeat units in the reference allele sequence versus the polymorphic information content for 92 dinucleotide (open circles), 32 trinucleotide (solid triangles), and 7 tetranucleotide (inverted open triangles) markers genotyped for length polymorphisms among 16 inbred lines of sunflower




Despite this, the mean $PIC_P$ score (0.46) for the six polymorphic SSR markers in the $n$ = 4 set was only 0.09 less than the mean $PIC_P$ score (0.55) for the complete set of polymorphic SSR markers. The PIC scores for the six $n$ = 4 SSR markers ranged from 0.12 to 0.72.

Several tetranucleotide repeats were exceptionally long (Fig. 2.2), marker development attrition rates were greater for tetranucleotide repeats than for shorter repeat motifs. We identified 23 unique tetranucleotide repeats, but could only design primers for 11 (47.8%) because of short flanking sequences. One tetranucleotide

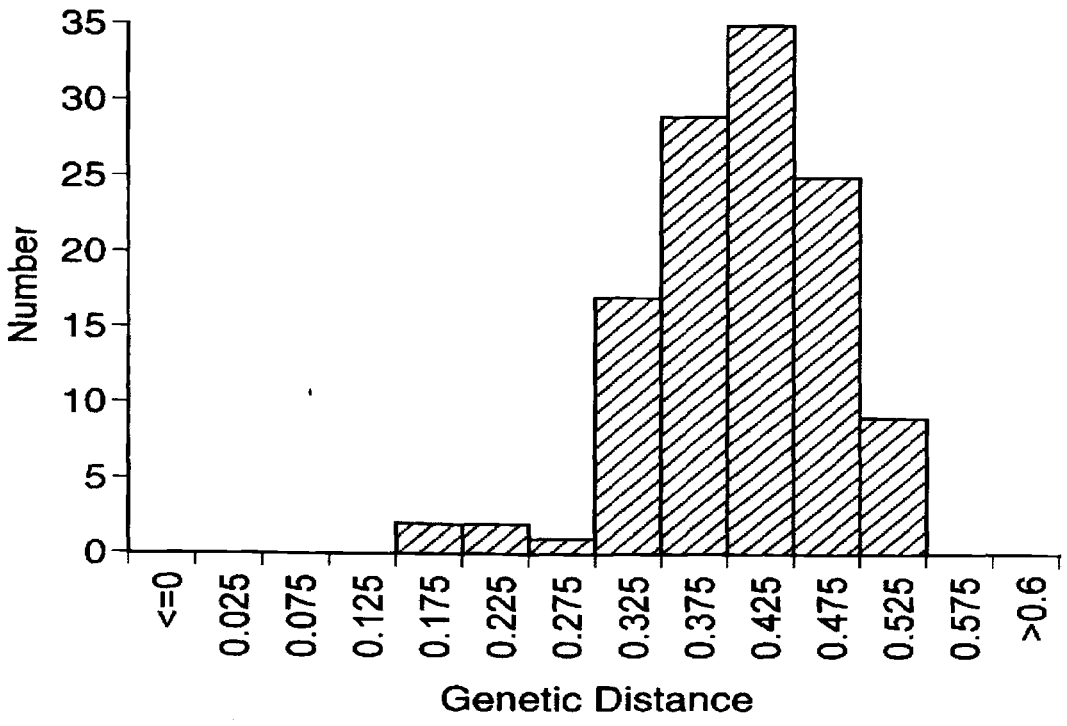
repeat spanned the entire length of the cloned DNA fragment (~500 bp). By comparison, primers were designed for 157 out of 230 unique dinucleotide and trinucleotide repeats (68.2%). Additionally, the primer failure rate for tetranucleotide repeats (36.4%) was greater than for dinucleotide and trinucleotide repeats (22.8%).

**Genetic Diversity Among Elite Inbred Lines**

Genetic distances among the 16 inbred lines ranged from 0.175 for PH-A x RHA801 to 0.543 for HA821 x RHA280 (Fig.2.3). The mean genetic distance between a particular line and the other 15 lines in the sample ($D$) ranged from 0.37 to 0.49 (Fig. 2.4). The two most diverse inbred lines with respect to the other lines sampled were RHA280 ($D = 0.49$) and HA372 ($D = 0.45$). Similarly, the two least diverse lines with respect to the other lines sampled were RHA377 ($D = 0.36$) and RHA801 ($D = 0.37$).

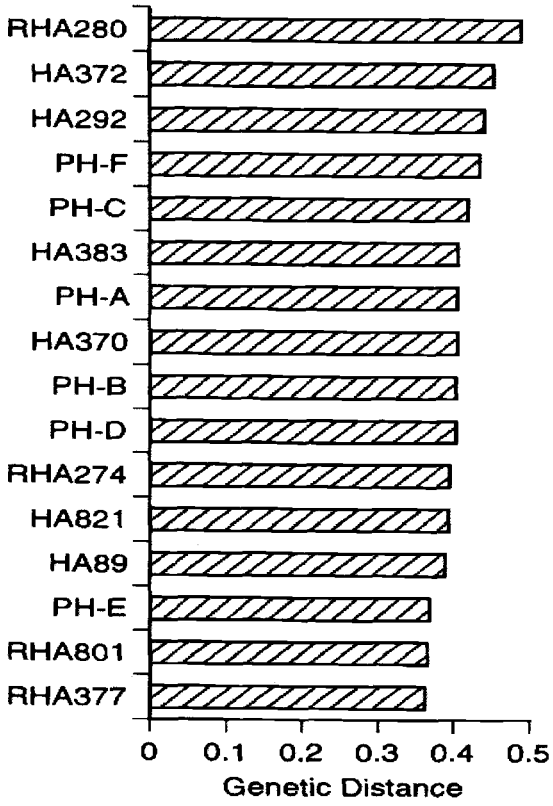
Principal component and cluster analyses of the genetic distance matrix estimated from 74 polymorphic SSR marker loci, uncovered patterns of genetic diversity (Fig. 2.5-6) similar to patterns uncovered with RFLPs (Berry et al. 1994; Gentzbittel et al. 1994), and AFLPs (Hongtrakul et al. 1997). The first three principal components were associated with 23.3, 21.1, and 11.3% of the variance of the genetic distance matrix (55.7% total). The plot of the first and second principal scores showed that, for the most part, lines from different market (oilseed and confectionery) and fertility restorer (maintainer and restorer) classes fell into separate quandrants (Fig. 2.5).

**Figure 2.3** Distribution of genetic distances among 16 inbred lines of sunflower estimated from the genotypes of 74 polymorphic SSR marker loci.



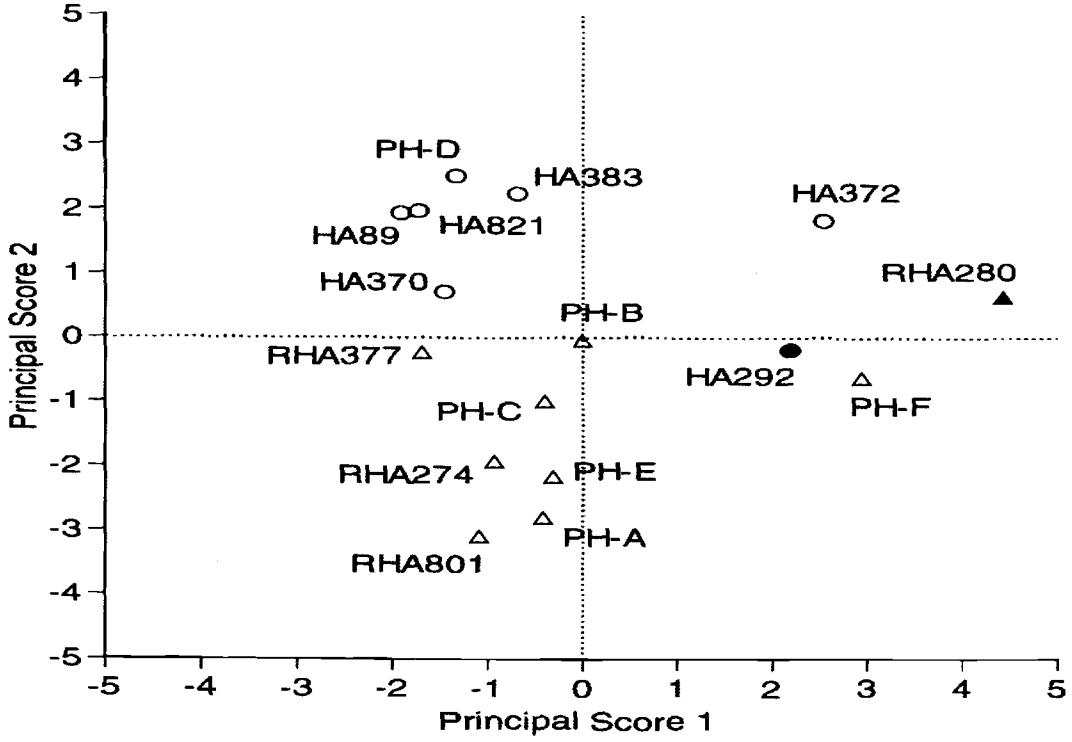

The lower left quandrant was solely comprised of oilseed R-lines, the upper left quandrant was solely comprised of oilseed B-lines, the upper right quadrant was comprised of the lone confectionery R-line (RHA280) and an oilseed B-line (HA372), and the lower right quadrant was comprised of the lone confectionery B-line (HA292) and an oilseed R-line (PH-F). The first principal component primarily separated lines along market class boundaries (with two outliers, HA372 and PH-F), whereas the second principal component primarily separated oilseed inbred lines along fertility restorer (R) and fertility maintainer (B) boundaries.

**Figure 2.4** The mean genetic distance between each inbred line and the other 15 inbred lines of sunflower in the sample of 16.



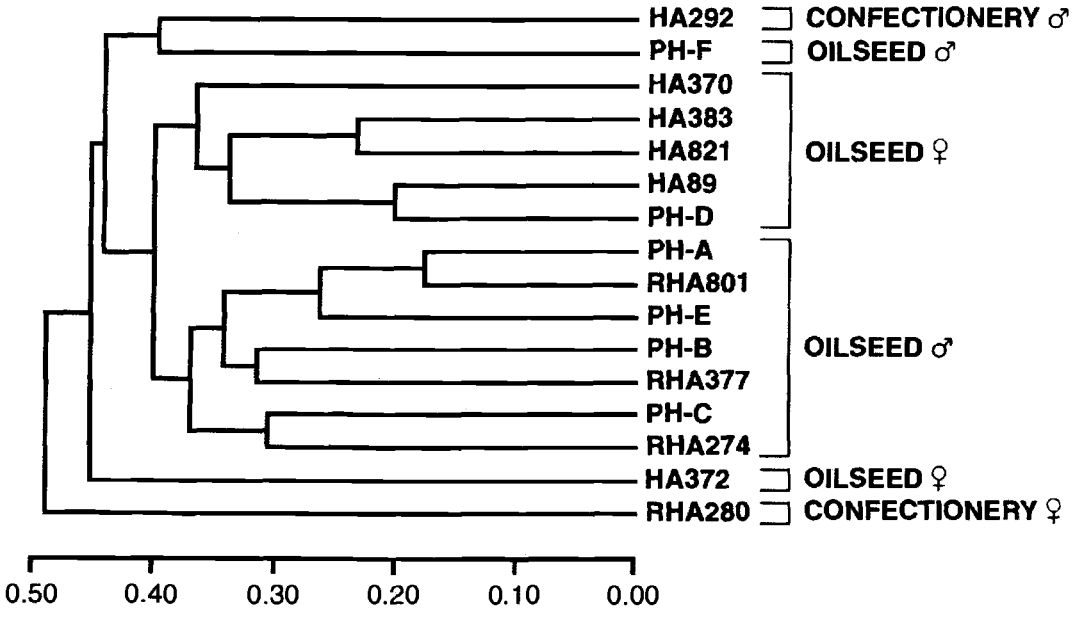

The dendogram produced by UPGMA cluster analysis of the genetic distance matrix shows the same pattern (Fig. 2.6). The cophenetic correlation (Mantel Z-statistic) between the tree and original genetic distance matrices was 0.84. The probability that random Z-statistics, estimated from 500 permutations, were greater than the observed Z-statistic, was 0.004; thus, the fit between the estimated tree matrix and the original genetic distance matrix was excellent. Oilseed B- and R-lines, apart from two outliers (HA372 and PH-F), formed two diverse but distinct clusters

**Figure 2.5** The first two principal component scores produced by an analysis of the genetic distance matrix for six oilseed fertility maintainer (B) lines (open circles), one confectionery B-line (filled circle), eight oilseed fertility restorer (R) lines (open triangles), and one confectionery R-line (filled triangle) of sunflower estimated from the genotypes of 74 polymorphic SSR marker loci.




separated by a genetic distance of 0.40 (Fig. 2.6). The mean genetic distance among the 16 inbred lines was 0.41 (Fig. 2.3). HA292 (a confectionery B-line) and PH-F formed a third, albeit weak, cluster that merged with the giant cluster formed by the merger of the oilseed B- and R-line clusters. The genetic distance between HA292 and PH-F (0.41) was as great as the mean distance between the oilseed B and R clusters (0.40).

**Figure 2.6** Dendrogram constructed by UPGMA cluster analysis of the genetic distance matrix for 16 inbred lines of sunflower estimated from the genotypes of 74 polymorphic SSR marker loci.




HA372 (an oilseed B-line) clustered second to last and RHA280 (a confectionery R-line) clustered last. Apart from merging with the global cluster, neither clustered with other lines in the sample of 16. The genetic distance between HA292 and RHA280 (0.48), the two confectionery inbred lines, was greater than the mean genetic distance between the oilseed B and R clusters and all but four oilseed inbred line hybrid combinations, specifically hybrids between HA372 and four R-lines (PH-A, PH-C, RHA274, and RHA801) (Fig. 2.6).

## Discussion

The SSR markers described herein complement the set of 72 SSR markers developed by Gedil (1999) and, with the addition of markers described by Brunel (1994), Whitton et al. (1997), and Hongtrakul et al. (1998b), increase the total number of SSR markers for sunflower to ~210. Primer sequences, polymorphism databases, and other data for both sets of SSR markers have been deposited in public databases (http://www.css.orst.edu/knapp-lab and http://compositdb.ucdavis.edu). The percentage of polymorphic markers for the new set (56.6%) was greater than that for set developed by Gedil (1999) (38.6%); however, the mean $PIC_P$ scores for the two sets were not significantly different. The PIC and $PIC_P$ scores for the Gedil (1999) SSR markers were 0.17 and 0.49, respectively, while the PIC and $PIC_P$ scores for the new set were 0.31 and 0.55, respectively.

The complexity of the sunflower genome and diversity of methods used to assess and report marker diversity complicates comparisons among DNA markers. The most meaningful and logical comparisons for our purposes were between RFLP and SSR markers. Both are locus specific, individually typeable, and typically multiallelic. SSR markers can be multiplexed, and depending on the assay system and degree of multiplexing, have multiplex ratios two to 16-fold greater than individually typed, single copy RFLP markers (see below). Furthermore, because single-copy RFLP and SSR marker loci are homologous across genotypes, they are widely used for intellectual property protection (IPP), inbred, cultivar, and population identification (Plaschke et al. 1995; Smith et al. 1997, 2000), and are ideal for genetic mapping. Berry et al. (1994) reported that the percentage of single copy probes in

randomly selected samples of genomic and cDNA probes ranged from 15 to 45% in sunflower. Fifty-five of the SSR markers described herein (42% of the total) amplified two or more loci, while 76 (58% of the total) amplified one band per inbred; thus, slightly less than half of the genomic regions sampled in our study were duplicated.

Hypervariability and multiallelism have fueled the development of SSR markers in crop plants; 20% of the polymorphic SSR markers (15 out of 74) had $PIC_P$ scores ranging from 0.70 and 0.93 (Fig. 2.1). SSR markers have been shown to be more polymorphic than RFLP markers in soybean (Akkaya et al. 1992; Rongwen et al. 1995), wheat (Plaschke et al. 1995; Roder et al. 1995), rice (Wu and Tanksley 1993; Olufowote et al. 1997), barley (Saghai Maroof et al. 1994), and rapeseed (Kresovich et al. 1995; Charters et al. 1996). Maize (*Zea mays* L.) is one of the few species where SSR markers do not seem to be more polymorphic than RFLP markers (Smith et al. 1997).

The polymorphism rates and PIC scores for RFLP and SSR markers in sunflower do not seem to be dramatically different. Berry et al. (1994) screened a sample of 57 RFLP probes for polymorphisms among 24 elite inbred lines, Gentzbittel et al. (1994) screened a sample of 181 RFLP probes for polymorphisms among 17 elite inbred lines, and Zhang et al. (1995) screened a sample of 51 RFLP probes for polymorphisms among 26 elite inbred lines of cultivated sunflower. The percentage of polymorphic RFLP probes ranged from 28.6 to 40.3% and genetic dissimilarities ranged from 0.04 to 0.69. Berry et al. (1994) reported a mean PIC score of 0.49 for polymorphic RFLP markers, Gentzbittel et al. (1994) reported a mean PIC score of

0.21 for polymorphic and monomorphic RFLP markers, and Zhang et al. (1995) reported a mean gene diversity of 0.59 for polymorphic RFLP markers. While the percentage of polymorphic markers was greater for SSRs (56.6%) than RFLPs, the power for identifying and discriminating between genotypes was virtually identical. The mean PIC and $PIC_P$ scores for SSR markers (0.31 and 0.55, respectively) (Table 2.1) were within the ranges reported for RFLP markers.

Tetranucleotide SSR markers (Fig.2.1 and Table 2.1) seem to be significantly more polymorphic than RFLP and shorter motif SSR markers. The hypervariability of tetranucleotide repeats tends to be offset by greater genotyping noise and greater rates of attrition in marker development. Tetranucleotide SSR markers, as a whole, produced more complex genotyping patterns than dinucleotide and trinucleotide SSR markers among the 16 elite inbred lines. Developing tetranucleotide SSR markers is more difficult than developing dinucleotide and trinucleotide SSR markers, primarily because tetranucleotide repeats tend to be long and difficult to amplify and short insert clones often lack sufficient flanking sequence for primer design. The problem of short flanking sequences can be overcome by cloning longer DNA fragments; however, enrichment protocols that target longer DNA fragments, e.g., 1,000 to 1,500 bp, have not been tested.

Dinucleotide repeats have been shown to be more polymorphic than trinucleotide repeats in several plant species: barley (Struss and Plieske 1998), rice (Akagi et al. 1997), wheat (Plaschke et al. 1995), maize (Smith et al. 1997), and soybean (Rongwen et al. 1995). We found both classes of SSRs to be equally polymorphic in sunflower. The $PIC_P$ scores for dinucleotide and trinucleotide SSRs

were identical (0.53), and the number of alleles per locus ($A_P$) were virtually identical (3.67 for dinucleotide and 3.57 for trinucleotide repeats) (Table 2.1). If this pattern holds up, then greater focus should be placed on isolating and developing trinucleotide SSR markers in sunflower, particularly since dinucleotide and tetranucleotide repeats are primarily located in non-coding regions, whereas trinucleotide repeats, particularly GC-rich trinucleotide repeat motifs, frequent coding regions and regulatory elements (Akagi et al. 1997; Broun and Tanksley 1996; Cho et al. 2000; Temnykn et al. 2000). Moreover, trinucleotide repeats, as a whole, tend to produce cleaner genotyping patterns than dinucleotide and tetranucleotide repeats, with or without the addition of 'pigtails' designed to reduce noise caused by partially adenylated amplicons (Brownstein et al. 1996).

SSR markers are significantly more polymorphic than RAPDs (Arias and Rieseberg 1995), AFLPs (Hongtrakul et al. 1997), and allozymes (Cronn et al. 1997) in sunflower. The polymorphism differences between SSRs, RAPDs, and AFLPs primarily stem from the bi-allelic nature of the latter two marker systems. Hongtrakul et al. (1997) reported a mean PIC of 0.14 for AFLP markers among elite inbred lines, roughly half that of the mean PIC for SSR markers (Table 2.1). Similarly, Cronn et al. (1997) reported a mean heterozygosity (equal to PIC when estimated from inbred lines) of 0.13 for 30 allozyme loci among 'domesticated' accessions in sunflower. The SSR markers described herein had twice as many alleles as allozyme markers. Cronn et al. (1997) reported A and $A_P$ estimates of 1.39 and 2.05, respectively, for allozyme markers, while we reported A and $A_P$ estimates of 2.89 and 3.99, respectively, for SSR markers (Table 2.1).

Several factors apart from hypervariability have spurred the development of SSR markers in sunflower and other crop plants. Very high-throughput, semi-automated assay systems have been developed, SSR markers can be multiplexed, typically by pooling separately produced amplicons, and SSR markers can be electronically transmitted between laboratories. The multiplexing capacity of some DNA fragment length assay systems greatly increases genotyping throughput. We performed SSR genotyping on a semi-automated, high-throughput system (the ABI377) and designed SSR primers to facilitate multiplexing by dividing SSR markers into length and color bins. Reference allele amplicon lengths were uniformly sorted into 50 to 100 bp bins spanning ~100 to ~500 bp, and the three fluorophores (6FAM, TET, and HEX) were uniformly distributed among the length bins. The final distribution among length bins was not completely uniform; the bias was towards shorter amplicons because of the difficulty of identifying optimum primers in flanking sequences, particularly for long or non-centrally positioned repeats. Because the allele lengths for the screening panel genotypes were not known *a priori*, we multiplexed by color only in the primer and polymorphism screening process; multiplexing was done by pooling three separately produced amplicons, each labelled with a different fluorophore. The length and color bins we used for primer design routinely permit assays of 10 to 12 SSR markers per lane on ABI377, ABI3700, or other multicolor, semi-automated SSR genotyping platforms. Ten to 13- locus multiplexes (amplicon pools) can be designed for mapping or marker-assisted selection in segregating populations where the allele lengths are known *a priori*.

The genetic relationships among inbred lines were assessed principal component analysis (PCA) (Fig.2.5) and UPGMA cluster analysis (Fig.2.6). Both methods clearly separated oilseed B-lines, oilseed R-lines and confectionery lines. That different heterotic groups (oilseed B- vs. R-line) of cultivated sunflower are distinct has been shown by RFLP (Berry et al. 1994; Gentzbittle et al. 1994; Zhang et al. 1995), AFLP (Hongtrakul et al. 1997), and SSR fingerprint analyses from several different public and proprietary germplasm collections. In our study, the first three principal coordinates explained 56% of the genetic variance in the SSR data. This was much higher than the genetic variance explained using AFLP markers (33%), and the PCA scales were 1.3 times wider than those generated from AFLP marker scores. Additionally, SSR data showed R-line germplasm pools to be more diverse than B-lines, in agreement with RFLP fingerprint analysis.

Some published data indicate a high degree of correlation between DNA fingerprinting analysis and pedigree analysis (Smith et al. 1990), while others indicate no correlation between them (Ahnert et al. 1996). In our study, two apparently anomalous results are supported by coancestry analysis (Cheres et al. 1998). HA372, an oilseed B-line, however, did not cluster with other oilseed B lines. This result, however, is in agreement with Cheres et al. (1998), which demonstrated by coancestry analysis that HA372 failed to group with other oilseed B-lines. The genetic distance between HA89 and HA821 based on SSR data was only 0.319 and the first two principal coordinates indicated very tight grouping. However, cluster analysis separated them into two different subgroups. Likewise, pedigree analysis (Cheres et al. 1998) indicated that HA89 and HA821 belong to heterotic subgroups OB-B and

OB-C, respectively. Due to the limited numbers of inbreds surveyed, the present SSR fingerprint study was not a comprehensive analysis in the cultivated sunflower. Therefore, SSR fingerprint analysis from broad germplasm resources will be needed to compare the genetic distances estimated by coancestry with genetic distances estimated by molecular markers.

We have developed 131 new SSR markers for the public domain and assessed the efficiency of markers among 16 sunflower elite inbred lines. This study shows that SSR markers will be powerful tools for DNA fingerprinting, evaluating genetic diversity, producing genetic maps, and for molecular breeding in cultivated sunflower.

## References

Ahnert, D., Lee, M., Austin, D.F., Livini, C., Woodman, W.L., Openshaw, S.J., Smith, J.S.C., Porter, K., and Dalton, G. (1996) Genetic diversity among elite sorghum inbred lines assessed with DNA markers and pedigree information. Crop Sci. 36: 1358-1392.

Akagi, H., Yokozeki, Y., Inagaki, A., and Fujimura, T. (1997) Highly polymorphic microsatellites of rice consists of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor. Appl. Genet. 94: 61-67.

Akkaya, M.S., Bhagwat, A.A., and Cregan, P.B. (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132: 1131-1139.

Anderson, J.A., Churchill, G.A., Sutrique, J.E., Tanksley, S.D., and Sorrels, M.E. (1993) Optimizing parental selection for genetic linkage maps. Genome 36: 181-186.

Arias, D.M., and Rieseberg, L.H. (1995) Genetic relationships among domesticated and wild sunflowers (*Helianthus annuus*, *Asteraceae*). Econ. Bot. 49: 239 248.

Ashkenazi, V., Chani, E., Lavi, U., Levy, D., Hillel, J., and Velleux, R.E. (2001) Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. Genome 44: 50-62.

Berry, S.T., Allen, R.J., Barnes, S.R., and Caligari, P.D.S. (1994) Molecular marker analysis of *Helianthus annuus* L. 1.Restriction fragment length polymorphism between inbred lines of cultivated sunflower. Theor. Appl. Genet. 89: 435-441.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphims. Am. J. Hum. Genet. 32: 314-331.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, A., Minch, E., Kidd, J.R., and Cavalli-Sforza, L.L. (1994) High resolution human evolutionary threes from polymorphic microsatellites. Nature 368: 455-457.

Broun, P., and Tanksley, S.D. (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol. Gen. Genet. 250: 39-49.

Brownstein, M.J., Carpten, J.D., and Smith J.R. (1996) Modulation of non-template nucleotide addition by Taq DNA Polymerase: primer modifications that facilitate genotyping. BioTechniques 20: 1004-1010.

Brunel, D. (1994) A microsatellite marker in *Helianthus annuus* L. Plant Mol. Biol. 24: 397-400.

Chakraborty, R., Kimmel, M., Stivers, D.N., Davison, L.J., and Deka, R. (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA 94: 1041-1046.

Charters, Y.M., Robertson, A., Wilkinson, M.J., and Ramsay, G. (1996) PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. oleifera) using 5'-anchored simple sequence repeat (SSR) primers. Theor. Appl. Genet. 92: 442-447.

Cheres, M.T., and Knapp, S.J. (1998) Ancestral origins and genetic diversity of cultivated sunflower: analysis of the pedigrees of public germplasm sources. Crop Sci. 38: 1476-1482.

Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCough, S.R., Park, W.D., Ayres, N. and Cartinhour, S. (2000) Diversity of microsatellites derived from genomic libraries and GeneBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100: 713-722

Cronn, R., Brothers, M., Klier, K., Bretting, P.K., and Wendel, J.F. (1997) Allozyme variation in domesticated annual sunflower and its wild relatives. Theor. Appl. Genet. 95: 532-545.

Dehmer, K.J., and Friedt, W. (1998) Evaluation of different microsatellite motifs for analysing genetic relationships in cultivated sunflower (*Helianthus annuus* L.). Plant Breed. 117: 45-48.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., and Mattick, J.I. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 19: 4008.

Edwards, K.J., Barker, J.H.A., Daly, A., Jones, C., and Karp, A. (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20: 758-760.

Gedil, M.A., (1999) Marker development, genome mapping, and cloning of candidate disease resistance genes in sunflower, *Helianthus annuus* L. Ph.D. Thesis, Oregon State Univ., Corvallis.

Gentzbittel, L., Zhang, Y.X., Vear, F., Griveau, B., and Nicolas, P. (1992) RFLP studies of genetic relationships among inbred lines of the cultivated sunflower,

*Helianthus annuus* L.: evidence for distinct restorer and maintainer germplasm pools. Theor. Appl. Genet. 89: 419-425.

Gentzbittel, L., Zhang, Y.X., Vear, F., and Griveau, B. (1994) RFLP studies of genetic relationships among inbred lines of the cultivated sunflower, *Helianthus annuus* L.: evidence for distinct restorer and maintainer germplasm pools. Theor. Appl. Genet. 89: 419-425.

Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L., and Feldman, M.W. (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463-471.

Hamada, H., and Kakunaga, T. (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298: 396-398.

Hamada, H., Petrino, M.G., and Kakunaga, T. (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA 79: 6465-6469.

Hongtrakul, V., Huestis, G.M., and Knapp, S.J. (1997) Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theor. Appl. Genet. 95: 400-407.

Hongtrakul, V., Slabaugh, M.B., and Knapp, S.J. (1998) DFLP, SSCP, and SSR markers for D9 stearoyl-acyl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are hypervariable among elite inbred lines. Mol. Breed. 4: 195-203.

Ishii, T., and McCough, S.R. (2000) Microsatellites and microsynteny in the chloroplate genome of *Oryza* and eight other *Gramineae* species. Theor. Appl. Genet. 1000: 1257-1266.

Jan, C.C., Vick, B.A., Miller, J.F., Kahler, A.L., and Butler III, E.T. (1998) Construction of an RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. 96: 15-22.

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L., and Armour, J.A.L. (1994) Complex gene conversion events in germline mutation at human minisatellites. Nature Genetics 6: 136-145.

Kresovich, S., Szewc-McFadden, A.K., and Blick, S.M. (1995) Abundance and characterization of simple sequence repeats (SSRs) isolated from a size-

fractionated genomic library of *Brassica napus* L.(rapeseed). Theor. Appl. Genet. 91: 206-211.

Lagercrantz, U., Ellegren, H., and Anderson, L. (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Research 21:1111-1115.

Levison, G., and Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evo. 4: 203-221.

Lincoln, S.E., Daly, M.J. and Lander, E.S. (1991) PRIMER: A computer program for automatically selecting PCR primers. MIT. Cambridge, MA.

Mansfield, D.C., Brown, A.F., Green, D.K., Carothers, A.D., Morris, S.W., Evans, H.J., and Wright, A.F. (1994) Automation of genetic linkage analysis using fluorescent microsatellite markers. Genomics 24: 225-233.

Mantel, N.A. (1967) The detection of disease clustering and a generalized regression approach. Can. Res. 27: 209-220.

Minch, E., Ruiz-Linares, A., Goldstein, D., Feldman, M. and Cavalli-Sforza, L.L. (1997) MICROSAT: A computer program for calculating various statistics on microsatellite allele data. (http://crick.stanford.edu/microsat/microsat.html)

Morgante, M., and Olivieri, A.M. (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J. 3:175-182.

Olufowote, J.O., Xu, Y., Chen, X., Park, W.D., Beachell, H.M., Dilday, R.H., Goto, M., and McCouch, S.R. (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40: 370-378.

Plaschke, J., Ganal, M.W., and Roder, M.S. (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor. Appl. Genet. 91: 1001-1007.

Rae, S.J., Aldam, C., Dominguez, I., Hoebrechts, M., Barnes, S.R., and Edwards, K.J. (2000) Development and incorporation of microsatellite markers into the linkage map of sugar beet (*Beta vulgaris* spp). Theo. Appl. Genet. 1000: 1240-1248.

Ramsay, L., Macaulay, M., Ivanissevich, S.D., Maclean, K., Cardle, L., Fuller, J., Edwards, K.J., Tuvesson, S., Morgante, M., Massari, A., Maetri, E.,

Marmiroli, M., Sjakste, T., Ganal, M., Powell, W., and Waugh, R. (2000) A simple sequence repeat-based linkage map of barley. Genetics 156: 1997-2005.

Roder, M.S., Korzum, V., Wendehake, K., Plaschke, J., Tixier, M.-H., Leroy, P., and Ganal, M.W. (1998) A microsatellite map of wheat. Genetics 149: 2007-2023.

Roder, M.S., Plaschke, J, Konig, S.U., Borner, A., Sorrels, M.E., Tanksley, S.D., and Ganal, M.W. (1995) Abundance, variability and chromosomal location of microsatellite in wheat. Mol. Gen. Genet. 246: 327-333.

Rongwen, J., Akkaya, M.S., Bhagwat, A.A., Lavi, U., and Cregan, P.B. (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90: 43-48.

Saghai Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q., and Allard, R.W. (1994) Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. USA 91: 5466-5470.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: A laboratory manual. New York: Cold Spring Harbor.

Smith, J.S., Chin, E.C.L., Shu, H., Smith, O.S., Wall, S.J., Senior, M.L., Mitchell, S.E., Kresovich, S., and Ziegle, J. (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95: 163-173.

Smith, J.S.C., Kresovich, S., Hopkins, M.S., Mitchell, S.E., Dean, R.E., Woodman, W.L., Lee, M., and Porter, K. (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequences repeats. Crop Sci. 40:226-232.

Smith, O.S., Smith, J.S.C., Bowen, S.L., Tenborg, R.A., and Wall, S.J. (1990) Similarities among a group of elite maize inbreds as measured by pedigree, $F_1$ heterosis, and RFLPs. Theor. Appl. Genet. 80: 833-840.

Struss, D., and Plieske, J. (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor. Appl. Genet. 97: 308-315.

Tautz, D., and Renz, M. (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research 12: 4127-4137.

Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, T. G., Ishii, T., and McCouch, S.R. (2000) Mapping and genome organization of

microsatellite sequences in rice *Oryza sativa* L. Theo. Appl. Genet. 100: 697-712.

Wang, G., Mahalingam, R., and Knap, H.T. (1998) (C-A) and (G-A) anchored simple sequence repeats (ASSRs) generated polymorphism in soybean, *Glycine max* (L.) Merr. Theor. Appl. Genet. 96: 1086-1096.

Wang, Z., Weber, J.L., Zhong, G., and Tanksley, S.D. (1994) Survey of plant short tandem DNA repeats. Theor. Appl. Genet. 88: 1-6.

Webb, D.M., and Knapp, S.J. (1990) DNA extraction from a previously recalcitrant plant genus. Mol. Biol. Rep. 8: 180-185.

Weber, J.L., and May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388-396.

Whitton, J., Rieseberg, L.H., and Ungerer, M.C. (1997) Microsatellite loci are not conserved across the *Asteraceae*. Mol. Biol. Evol. 14. 204-209.

Wu, K.S., and Tanksley, S.D. (1993). Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet. 241: 225-235.

Zhang, Y.X., Gentzbittel, L., and Nicolas, P. (1995) Assessment of inter- and intra-inbred line variability in sunflower (*Helianthus annuus* L.) by RFLPs. Genome 38: 1040-1048.

# CHAPTER 3

## THE DEVELOPMENT OF SIMPLE SEQUENCE REPEAT MAPS FOR SUNFLOWER

Ju-Kyung Yu[1], Shunxue Tang[1], Mary B. Slabaugh[1], Glenn Cole[2] and Steven J. Knapp[1]

[1]Department of Crop and Soil Science, Oregon State University, Corvallis, Oregon, 97331, USA and [2]Pioneer Hi-Bred International, Woodland, California, 95695, USA

**Abstract**

The development of a dense genetic map supplies a critical mass of sequence-based DNA markers for constructing and cross-referencing genetic maps for cultivated sunflower (*Helianthus annuus* L.). Of the 1,090 public SSR markers described thus far for sunflower, roughly two-thirds are known to be polymorphic among elite oilseed inbred lines. Our aims were to add 192 previously unmapped SSR marker loci to the genetic map, integrate SSR markers into a preexisting framework of 77 RFLP markers using 94 $F_2$ progeny produced from a cross between two oilseed sterility maintainer lines (HA370 x HA372), and construct a dense SSR map using 94 recombinant inbred lines (RILs) produced from a cross between two oilseed fertility restorer lines (PHA x PHB). The HA370 x HA372 RFLP-SSR map was 1348.0 cM long, had a mean density of 6.77 cM, and was comprised of 77 RFLP and 122 SSR marker loci. The PHA x PHB SSR map was 1377.4 cM long, had a mean density of 4.99 cM, and was comprised of 276 SSR marker loci. Sixty common and 278 unique SSR marker loci were mapped in two mapping populations. The new maps create a dense framework of SSR markers for constructing and cross-referencing genetic maps and assigning new loci to the genetic map of sunflower.

**Key Words:** microsatellite, simple sequence repeat, restriction fragment length polymorphism, *Helianthus*, sunflower

## Introduction

Cultivated sunflower (*Helianthus annuus* L.) is one of 12 annual diploid species (2n = 2x =34) belonging to the subtribe Helianthinae of the Compositae (Asteraceae) family (Schilling and Heiser 1981). Sunflower was domesticated by native Americans from wild *Helianthus annuus* L. Sunflowers were first grown as a source of edible oil in nineteenth century Russia and are currently the second largest oilseed crop in the world, exceeded only by soybean. Common cultivated sunflower seed produces an oil which is rich in unsaturated fatty acids such as oleic and linoleic acids, and contains a large amount of vitamin E.

Several genetic maps have been developed for cultivated sunflower. Restriction fragment length polymorphism (RFLP) maps were developed using intraspecific crosses (Berry et al. 1995, 1996; Gentzbittel et al. 1995, 1999; Jan et al. 1998). The most comprehensive RFLP map was developed by Berry et al. (1996) and integrated 635 RFLP marker loci across nine $F_2$ mapping populations covering 1650 cM with 2.3 cM of mean distance. Jan et al. (1998) mapped 269 RFLP marker loci to 20 linkage groups covering 1164 cM using an $F_2$ population. A composite RFLP map constructed using seven $F_2$ segregating populations, was developed by Gentzbittel et al. (1999). This map had a mean density of 7 cM and was comprised of 23 linkage groups covering 1573 cM. In addition, it incorporated morphological loci and known functional genes related to disease and stress resistance. Thus, a total of 1141 RFLP loci have been mapped by three research groups using independent probes derived from genomic and cDNA clones. However, only 81 cDNA probes for RFLP loci dispersed across the 17 linkage groups have been released for public use (Berry et al.

1995). This situation makes it difficult to i) identify homologous loci across genotypes, ii) assign linkage group polarities, iii) generate linkage group assemblages, and iv) cross-reference maps.

Simple sequence repeats (SSRs) have become important genetic markers in many plant genomes because of the high level of polymorphism (Wang et al. 1994). In addition, the analysis of SSR markers is based on the polymerase chain reaction (PCR), facilitating the screening of genotypes as compared to RFLP markers. Genetic maps based on SSR markers have been developed in wheat (Roder et al. 1998), rice (Temnykh et al. 2000), barley (Ramsay et al. 2000), and other crop plants.

Yu et al. (2001) developed 131 SSR markers using a small insert genomic library of sunflower enriched for a variety of repeat motifs. A total of 56.5 % of the markers were polymorphic and the discrimination power of the markers was demonstrated using 16 inbred lines. The percentage of polymorphic SSR markers per cross ranged from 19% to 42% (Yu et al. 2001). This study also demonstrated several practical advantage of SSRs such as high multiplex ratio and high throughput genotyping. Tang et al. (2001) developed a total of 887 additional SSR markers using genomic DNA libraries enriched for AC and AG repeats. In that study 587 SSR markers were polymorphic among four inbred lines (RHA280, RHA801, PHA and PHB). A total of 456 SSR marker loci were mapped on 94 recombinant inbred lines from a cross between RHA280 (confectionery line) and RHA801 (oilseed line). The map was comprised of 17 linkage groups and covered 1,567 cM, which was estimated to be 95% of the sunflower genome. This created a framework of publicly shared, sequence-based DNA markers for cross-referencing genetic maps and adding new loci

to the genetic map of sunflower.

Herein, we report the development of two sunflower genetic maps based on SSR markers. These markers were drawn from 1,090 SSR markers developed in our laboratory (Gedil 1999; Tang et al. 2001; Yu et al. 2001).

## Materials and Methods

### DNA Marker Development and Screening

We screened 1,090 SSR markers (Gedil 1999; Tang et al. 2001; Yu et al. 2001) and two sequence characterized amplified region (SCAR) markers, SCTO6151 and SCX20600 (Lawson et al. 1998) for polymorphisms among four sunflower inbred lines, HA370, HA372, PHA, and PHB. HA370 and HA372 are public oilseed sterility maintainer lines developed by Miller and Gulya (1990). PHA and PHB are proprietary oilseed fertility restorer lines developed by Pioneer Hi-Bred International, Woodland, CA, USA. SSR genotyping assays were performed essentially as described by Yu et al. (2001), on an ABI Prism 377 DNA Sequencer (Applied Biosystem, Perkin- Elmer, Foster City, CA, USA) using polyacrylamide gels and fluorescently labelled amplicons. Filter set C and the GeneScan 500 TAMRA internal standard were used for assays performed with 6FAM, TET and HEX labelled amplicons. Filter set A and the GeneScan 500 ROX internal standard were used for assays performed with 6FAM, HEX, and NED labelled amplicons. The amplicons for each SSR markers were separately produced, pooled post-PCR, and loaded into each lane; each amplicon in each pool (lane) was labelled with a different fluorophore (e.g., 6FAM, HEX, and TET). Three-color multiplexes were used so that the DNA

fragments produced by each SSR primer pair could be unequivocally identified. Genotypes (SSR allele lengths) were recorded using GeneScan (version 2.1) and Genotyper (version 2.0) software (Applied Biosystem, Perkin- Elmer, Foster City, CA), and manually checked. Genotypes were only recorded for samples with fluorescent signal strengths greater than 100 fluorescence intensity units. The individual loci amplified by multilocus SSR markers were labelled with consecutive letters (e.g., A, B, and C). SCAR markers were assayed as described by Lawson et al. (1998).

**Genetic Map Development**

Genetic maps were constructed by genotyping 94 $F_2$ progeny from a cross between HA370 and HA372 (Gedil et al. 2001b) and 94 $F_6$ recombinant inbred lines (RILs) from a cross between PHA and PHB. Genotyping of HA370 x HA372 $F_2$ progeny was performed on DNA samples isolated from the progeny described by Gedil et al. (2001b); 116 SSR markers (Gedil 1999; Tang et al. 2001; Yu et al. 2001), and the SCTO6151 SCAR marker (Lawson et al. 1998) were genotyped. SSR genotyping assays were performed by multiplexing three to 13 SSR markers per lane as described by Tang et al. (2001) and Yu et al. (2001). SSR markers were selected for multiplexing on the basis of allele lengths and primer fluorophores; amplicons were separately produced for each SSR primer pair and pooled for gel loading. Genotypes for the SSR and SCAR marker loci were merged with genotypes for 77 RFLP marker loci, a resistance gene candidate marker locus (*RGC1A*), and $Pl_1$ from the map of Gedil et al. (2001a, 2001b); the locus *RGC1A* was previously named HR-4W2.

The PHA x PHB RILs (265 total) were developed by single seed descent between 1994 and 1998 in Woodland, CA, USA. Progeny were advanced through the $F_6$ generation by bagging one head per lineage, bulk harvesting the seed, and advancing one individual per lineage (bagged head). DNA was isolated using a modified CTAB method (Webb and Knapp 1990) from three-week-old seedling leaves harvested from three to 10 $F_6$ individuals per RIL; the leaves from each RIL were bulked, lyophilized, and ground to fine powder. DNA concentrations were measured using a Versafluor fluorometer (Bio-Rad Laboratories, Hercules, CA, USA). Genotyping was performed on DNA samples from 94 randomly selected RILs from the sample of 265 RILs; 250 SSR markers (Gedil 1999; Tang et al. 2001; Yu et al. 2001), and the SCTO6151 and SCX20600 SCAR markers (Lawson et al. 1998) were genotyped. A morphological locus, branching trait, was scored by Pioneer Hi-Bred International, Woodland, CA, USA.

Statistical analyses were performed and maps were constructed using G-MENDEL (Holloway and Knapp 1993). Tests for segregation distortion were performed for each locus. Loci were grouped in the HA370 x HA372 $F_2$ map using a likelihood odds (LOD) threshold of 3.0 and a recombination frequency threshold of 0.25. Loci were initially grouped in the PHA x PHB RIL map using a LOD threshold of 7.0 and a recombination frequency threshold of 0.25. If loci coalesced into more than one linkage group corresponding to single known linkage group, then groups and orders were re-estimated using a LOD threshold of 3.0.

Because dominant marker loci linked in repulsion often cannot be properly ordered in $F_2$ populations (Knapp et al. 1995), the $F_2$ map was constructed using

dominant markers linked in coupling and unlinked dominant markers (independent of linkage phase) only. Map distances (cM) were calculated using the Kosambi (1944) mapping function. RIL map distances were calculated using recombination frequency estimates corrected for multiple meioses under selfing ($r$), where $r = R / (2\text{-}2R)$, $R = n_r / (n_r + n_n)$ is the recombination frequency estimated from RIL genotypes *per se*, and $n_r$ and $n_n$ are the number of recombinant and non-recombinant genotypes, respectively, between two loci (Haldane and Waddington 1931). The reproducibilities of locus orders were checked by performing 100 repeat runs of the locus ordering algorithm (via the MONTE function) and estimating the mean locus order and the concordance (W) among the 100 locus order estimates (Holloway and Knapp 1993). The statistical significance of W was tested as described by Kendall and Gibbons (1990). If W = 1, then there is complete concordance between the locus order estimates. Similarly, if W = 0, then there is no concordance between the locus order estimates.

## Results

HA370, HA372, PHA, and PHB were screened for polymorphisms using 1,090 SSR markers (Gedil 1999; Tang et al. 2001; Yu et al. 2001); 285 SSR markers (26.1%) were polymorphic between HA370 and HA372 and 362 SSR markers (33.2%) were polymorphic between PHA and PHB (Table 3.1). Of the total, 148 SSR markers were polymorphic in both mapping populations and 351 SSR markers were polymorphic in one or the other. One hundred twenty-two marker loci were genotyped in the HA370 x HA372 $F_2$ population and 276 SSR marker loci were genotyped in the PHA x PHB RIL population. Of the 366 SSR primer pairs used for genotyping, 341 (93.2%) amplified a single locus and 25 (6.8%) amplified two to three loci. The loci amplified by multilocus SSR primers mapped to different linkage groups except for ORS 670 and 1203 on PHA x PHB and ORS 781 and ORS 1068 on HA370 x HA372.

The HA370 x HA372 and PHA x PHB maps (Fig. 3.1) were developed in parallel with the RHA280 x RHA801 RIL map described by Tang et al. (2001). SSR marker loci were added to a pre-existing backbone of RFLP marker loci on the HA370 x HA372 map spanning the 17 known linkage groups (Gedil et al. 2001b). SSR marker loci that mapped in two or three mapping populations ('common' SSR marker loci) were used to number and orient linkage groups using the nomenclature of Berry et al. (1997) and Gedil et al. (2001b). In addition, the polarities of LG6, LG7, LG12 and LG15 on the HA370 x HA372 map and LG1, LG6, LG8 and LG14 on the PHA x PHB map were determined using the RHA 280 x RHA 801 map (Tang et al. 2001).

**Table 3.1** The summary of SSR marker development for mapping in two populations

| | HA370 x HA372 ($F_2$) | PHA x PHB (RIL) |
|---|---|---|
| Polymorphic markers | 285 | 362 |
| Mapped markers | 116 | 250 |
| Mapped single-locus | 112 | 229 |
| Mapped multi-locus | 4 | 21 |
| Mapped loci | 122 | 276 |
| Codominant loci | 83 (68.0 %) | 142 (51.5 %) |

The HA370 x HA372 map was constructed using 77 RFLP and 122 SSR marker loci (Table 3.2); RFLP marker prefixes were ZVG and UB and the SSR marker prefix was ORS. The map was 1,348.0 cM long and had a mean density of 6.77 cM. The linkage group assignments and locus orders for RFLP markers on the reference RFLP map (Berry et al. 1997) and HA370 x HA372 RFLP and RFLP-SSR maps were identical (Gedil et al. 2001b). The number of SSR marker loci per linkage group ranged from two on LG6 and LG8 to LG16 on LG17. The number of RFLP marker loci per linkage group ranged from one on LG12 and LG15 to seven on LG3 and LG17 (Table 3.2). The density of markers ranged from 3.89 cM on LG7 to 13.78 cM on LG12. The SSR markers placed on the HA370 x HA372 map were selected for end-to-end genome coverage rather than marker density *per se*; 100% of the SSR markers mapped to one of the 17 known linkage groups. The disease resistance-related markers (*RGA1A*, *Pl*1 and SCTO6151) were closely linked to ORS166 on LG8

within 5.8 cM (Fig. 3.1). The segregation ratios for three RFLP and six SSR marker loci were significantly distorted. The distorted loci were not clustered and the frequency of distorted loci (9/116 = 0.078) was close to the nominal frequency of false positive errors.

The PHA x PHB map was constructed using 276 SSR marker loci (Table 3.2 and Fig. 3.1). The map was 1,377.4 cM long and had a mean density of 4.99 cM. The SSR marker loci coalesced into 23 linkage groups; 8 SSR marker loci (ORS197, 374, 388, 457, 628, 769, 976A and 1042) coalesced into three linkage groups that could not be assigned to any of the known linkage groups (data not shown). LG8, LG9, and LG16 were comprised of two linkage subgroups. The subgroups were positioned and oriented with known linkage groups using common SSR markers mapped in RHA280 x RHA801 (Tang et al. 2001). The widest gaps on the map, and the only gaps longer than 20 cM, were between ORS1093 and ORS178 on LG1 (21.1 cM), ORS963 and ORS785 on LG4 (27.8 cM), ORS626 and ORS852 on LG5 (25.3 cM), and ORS204 and ORS1051 on LG17 (24.0 cM) (Fig. 3.1). SCAR DNA marker loci (SCTO6151, SCX20600) which are linked to rust resistance traits, were located on LG8 and LG13, respectively (Fig. 3.1). A branching trait locus was mapped on LG10 (Fig. 3.1). The segregation ratios for 76 SSR marker loci were significantly distorted in the PHA x PHB population; 47 had an excess of PHA homozygotes, four had an excess of PHB homozygotes, and 25 had an excess of heterozygotes. Segregation distortion, whether caused by an excess of heterozygotes or gametic selection (an excess of one allele or the other), has no effect on the estimation of recombination frequencies (Bailey 1960) or locus orders.

**Table 3.2** The number of SSR and RFLP loci, total and mean distance (cM) for SSR map in the PHA x PHB population and SSR-RFLP integrated map in the HA370 x HA372 population of sunflower.

| | HA370 x HA372 | | | | PHA x PHB | | |
|---|---|---|---|---|---|---|---|
| LG | SSR | RFLP | Size | Mean | SSR | Size | Mean |
| 1 | 8 | 6 | 100.3 | 7.16 | 4 | 30.0 | 7.50 |
| 2 | 6 | 5 | 45.7 | 4.16 | 15 | 70.4 | 4.69 |
| 3 | 9 | 7 | 117.1 | 7.32 | 15 | 58.1 | 3.87 |
| 4 | 5 | 5 | 67.8 | 6.78 | 10 | 82.6 | 8.26 |
| 5 | 13 | 4 | 85.5 | 5.03 | 27 | 122.7 | 4.54 |
| 6 | 2 | 2 | 15.7 | 3.93 | 3 | 23.0 | 7.67 |
| 7 | 11 | 3 | 54.4 | 3.89 | 22 | 93.8 | 4.26 |
| 8 | 2 | 5 | 78.1 | 11.16 | 15 | 110.7 | 7.38 |
| 9 | 6 | 6 | 76.8 | 6.40 | 25 | 102.2 | 4.09 |
| 10 | 10 | 5 | 105.5 | 7.03 | 26 | 139.5 | 5.37 |
| 11 | 4 | 5 | 63.0 | 7.00 | 14 | 93.6 | 6.69 |
| 12 | 3 | 1 | 55.1 | 13.78 | 12 | 41.8 | 3.48 |
| 13 | 6 | 6 | 123.4 | 10.28 | 12 | 65.8 | 5.48 |
| 14 | 4 | 3 | 77.5 | 11.07 | 11 | 55.4 | 5.04 |
| 15 | 4 | 1 | 32.1 | 6.42 | 17 | 57.4 | 3.38 |
| 16 | 13 | 6 | 133.6 | 7.03 | 23 | 98.9 | 4.30 |
| 17 | 16 | 7 | 116.4 | 5.06 | 17 | 131.5 | 7.74 |
| Total | 122 | 77 | 1348.0 | 6.77[§] | 276[†] | 1377.4[‡] | 4.99[§] |

[†] Estimated number of SSR loci including eight loci from three undesignated linkage groups.

[‡] Estimated total distance (cM) of linkage groups including three undesignated linkage groups.

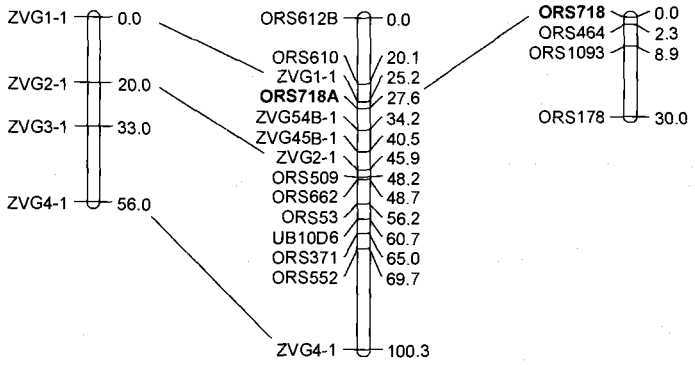
[§] Calculated mean distance of linkage groups with only SSR and RFLP loci.

The mean frequency of heterozygotes across SSR marker loci was 0.0387 among the PHA x PHB RILs. This was slightly but not significantly greater than the expected frequency of heterozygotes for $F_6$ RILs (0.0313).

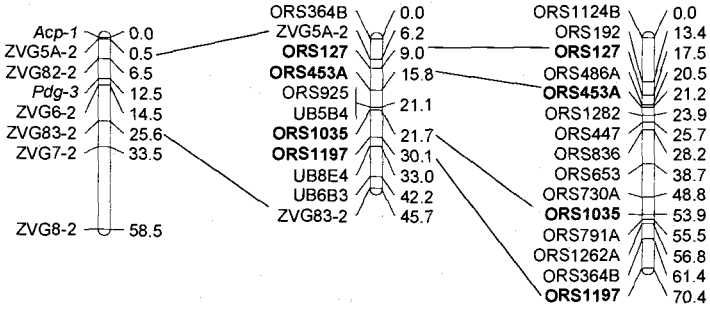
Of 398 SSR marker loci mapped in HA370 x HA372 and PHA x PHB, 98 were previously unmapped, 60 were placed on both maps, 278 were placed on one map or the other. Of 732 SSR marker loci mapped in RHA280 x RHA801 and PHA x PHB, 155 were placed on both maps and 422 were placed on one map or the other. Of 578 SSR marker loci mapped in RHA280 x RHA801 and HA370 x HA372, 84 were on both maps and 410 were placed on one map or the other. Ninety-eight previously unmapped SSR marker loci could be added to the HA370 x HA372 or PHA x PHB maps; 100% of the SSR markers found to be polymorphic in RHA280 x RHA801 have been mapped (Tang et al. 2001). Thus far, 554 different SSR marker loci have been mapped and 52 SSR marker loci are common to the three maps.

Most of the newly mapped SSR marker loci fell within the boundaries of the previously mapped RFLP marker loci; however, several mapped distal to RFLP markers and lengthened the genetic map of sunflower. Specifically, ORS612B and ORS610 on LG1, ORS364B on LG2, several markers on the upper ends of LG3, ORS795 and ORS1265 on LG9, ORS200, and ORS316 on LG13 and several markers on the lower end of LG17, mapped distal to RFLP markers. Several SSR marker loci on the upper ends of LG7, LG12, and LG14 seemed to be positioned distal to RFLP. This could not be shown unequivocally because key SSR or RFLP loci (three SSR markers distal to ORS1041 on LG7 and ZVG54-12 and ZVG62-14 on LG12 and LG14, respectively) were not polymorphic and thus have not been mapped in

1

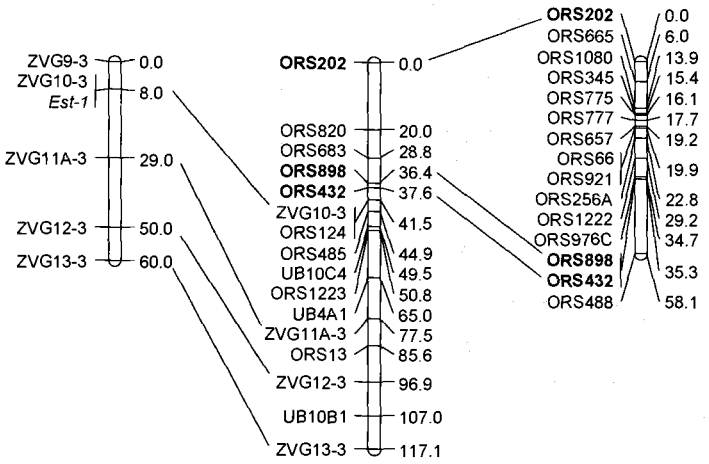



2

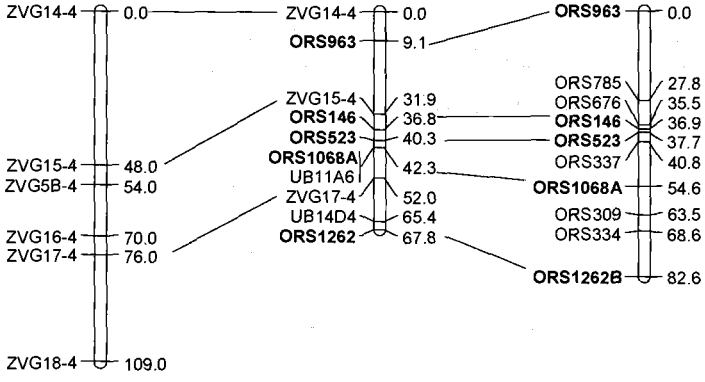



Continued

**3**

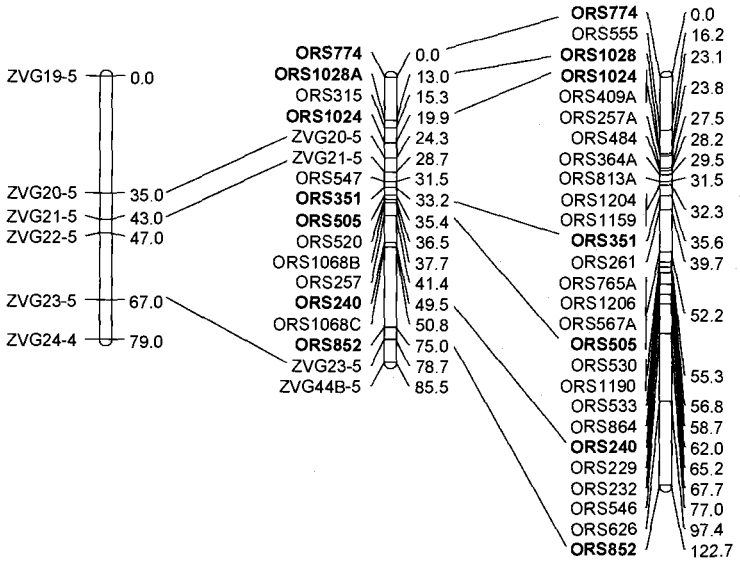



**4**

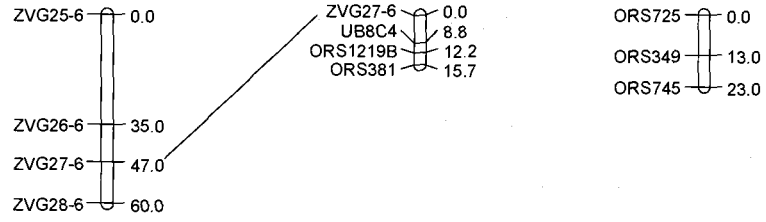



Continued

**5**

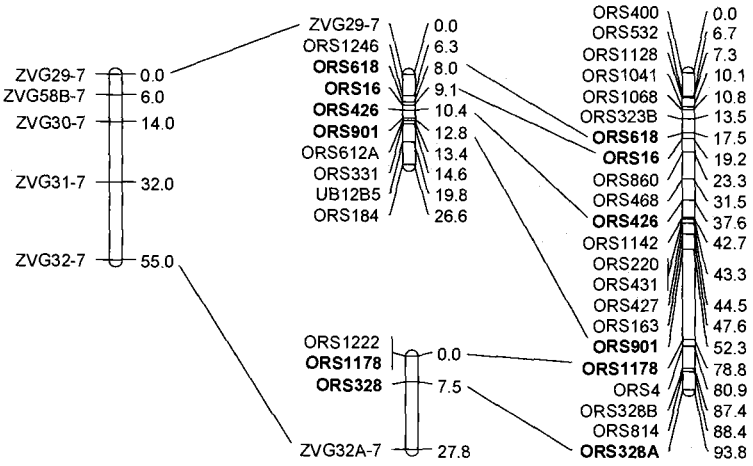



**6**

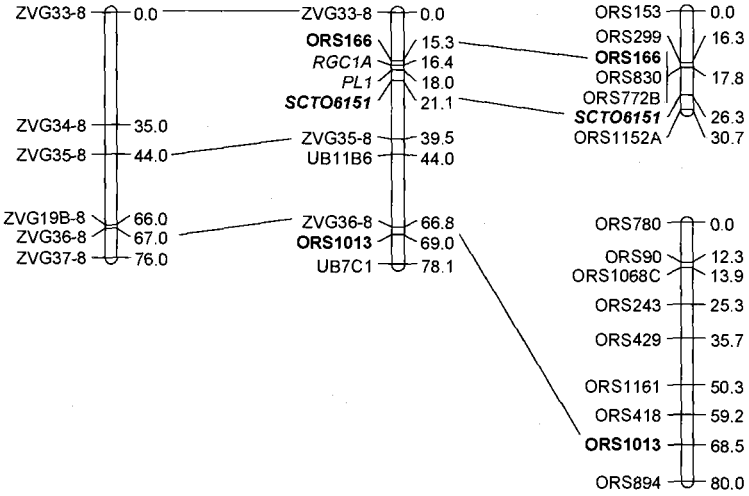



Continued

**7**

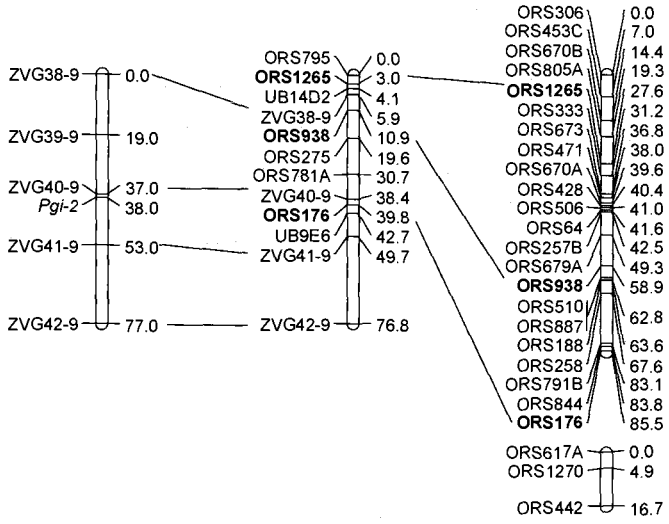



**8**

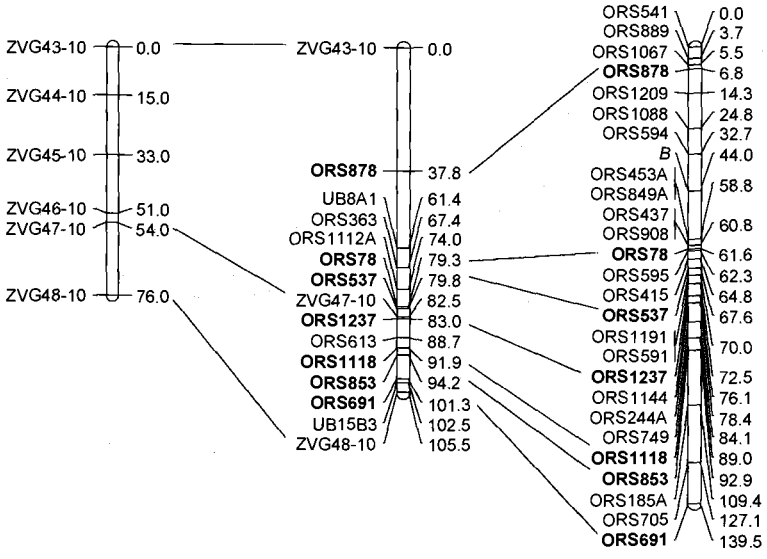



Continued

**9**

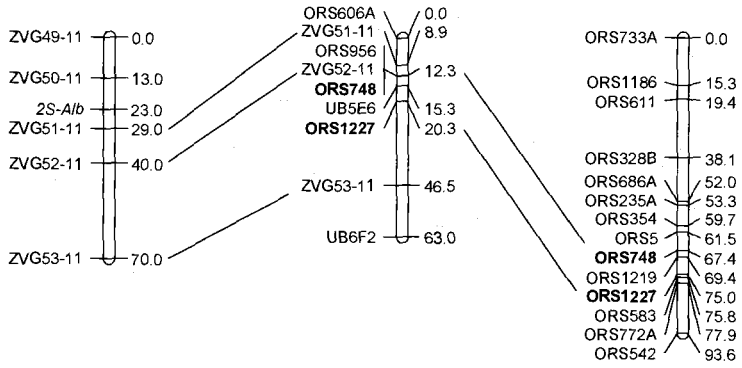



**10**

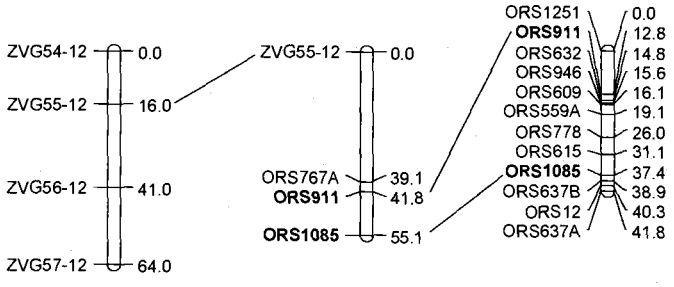



Continued

**11**

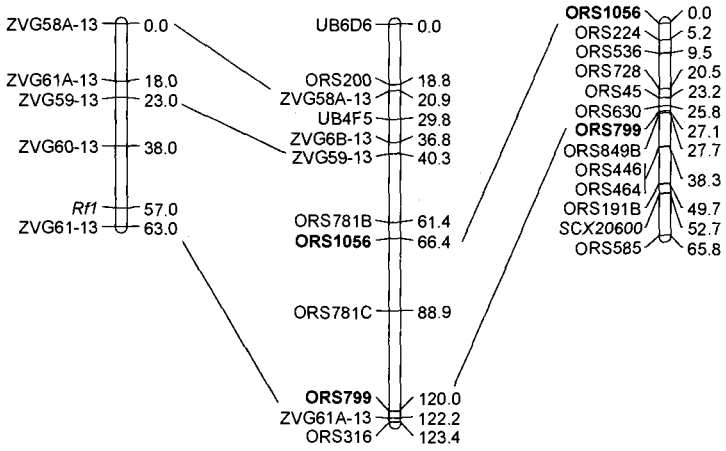



**12**

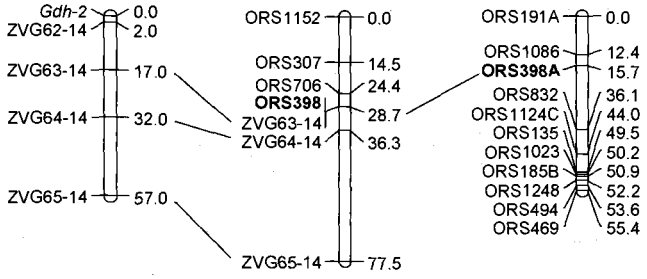



Continued

**13**

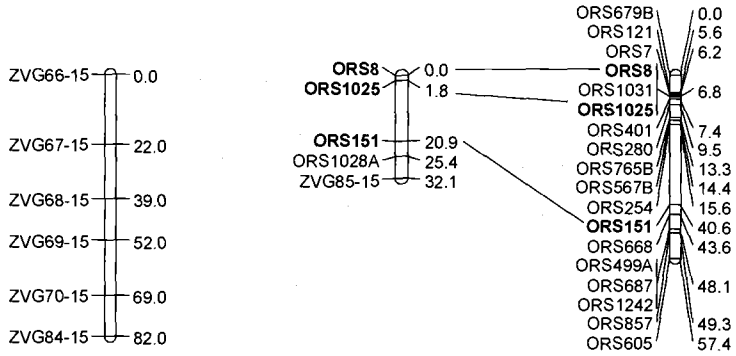



**14**

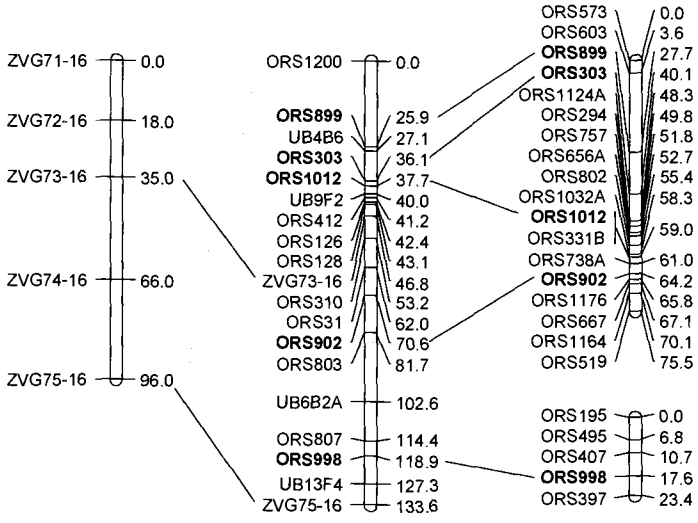



Continued

**15**




**16**




Continued

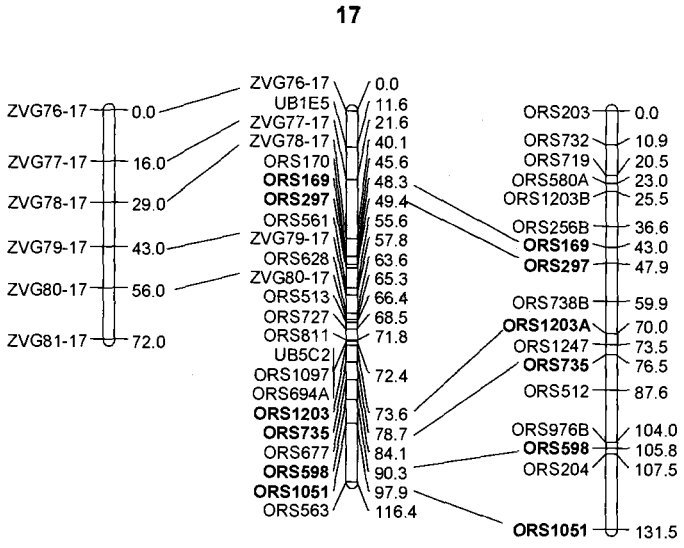



**Figure. 3.1** SSR maps for cultivated sunflower (*Helianthus annuus* L.) from HA370 x HA372 ($F_2$ population) and PHA x PHB (RIL population). Each linkage group includes three chromosomes. The ZVG RFLP map is on the left, the HA370 x HA372 SSR-RFLP map is in the center, and the PHA x PHB SSR map is on the right.

HA370 x HA372 (Fig. 3.1; Gedil et al. 2001b). A total of 177.7 cM were extended between RFLP marker loci adding by including SSR marker loci. This may be due to genotyping or sampling errors. Especially, mistyping tends to introduce spurious recombinants thereby inflating map distance estimates (Lincoln and Lander 1992).

Two linkage groups on the HA370 x HA372 map (LG6 and 15) and two linkage groups on the PHA x PHB map (LG1 and 6) were sparsely populated with SSR markers (Fig. 3.1). LG6 has been short in several maps and seems to be one of the least polymorphic chromosomes in sunflower (Berry et al. 1995, 1997; Gedil et al.

2000b). The unmapped regions on LG15 in HA370 x HA372 and LG1 in PHA x PHB may be artifacts of coancestry. PHA and PHB are both fertility restorer lines, while HA370 and HA372 are both sterility maintainer lines. LG1, LG6, and LG15 were well covered with SSR markers in the RHA280 x RHA801 map (Tang et al. 2001).

The HA370 x HA372 RFLP-SSR and PHA x PHB SSR maps were shorter (1,348.0 and 1,377.4 cM, respectively) than the RHA280 x RHA801 SSR map (1,567.3 cM) (Tang et al. 2001). The latter is the longest map produced thus far for cultivated sunflower using a single mapping population and individually typeable, locus-specific DNA markers (e.g. RFLPs and SSRs). RHA280 x RHA801, a cross between confectionery and oilseed fertility restorer lines, is more polymorphic than crosses between elite inbred lines (Yu et al. 2001).

The three maps together supply a dense framework of SSR markers for constructing new maps. The sunflower genome is well covered with SSR marker loci, although some regions are less dense than others and some gaps persist in the map. The two new maps were shorter than the predicted length range of the genetic map of sunflower (1,650 to 1,800 cM) (Gentzbittel et al. 1995), but cover 81.7 to 83.5 % of the genome. Collectively, the RHA280 x RHA801, HA370 x HA372, and PHA x PHB maps seem to cover 95.0% or more of the sunflower genome.

## Discussion

We present, herein, two genetic maps for cultivated sunflower; an SSR map for a RIL population and an RFLP-SSR map for an $F_2$ population. A property of recombinant inbred lines is that they undergo several rounds of meiosis before homozygosity is reached. Therefore, recombination between closely linked markers is more readily detected in a RIL population as compared to a conventional segregating population such as $F_2$ or backcross. However, this advantage can cause difficulty in mapping linkage near a telomeric region because of a lack of polymorphic markers in distal regions of the RIL population. There were 13 unlinked markers on the PHA x PHB map (ORS14, 126, 321, 340, 424, 502, 513, 525, 561, 597, 841, 927 and 1230) (data not shown). Of the 13 unlinked markers, ORS502, 525 and 597 were common markers between PHA x PHB and RHA280 x RHA801 population (Tang et al. 2001) and were mapped in distal regions (LG12, LG13 and LG17, respectively) of the RHA280 x RHA801 RIL population. This cross between confectionery and oilseed lines is more polymorphic than the PHA x PHB cross between two oilseed lines (Yu et al. 2001). An additional advantage of RILs is that they constitute a permanent population in which segregation is nearly complete. They can be used indefinitely for mapping, so new data can continuously be added to a pre-existing map.

A significant percentage (43.5%) of the SSR marker loci we mapped were dominant. This is much higher than observed with RFLPs (Berry et al. 1995; Jan et al. 1998; Gentzbittel et al. 1999), 11, 35 and 26 %, respectively. The high percentage of null alleles suggests that primer annealing sites flanking SSRs are often polymorphic in sunflower (Gedil et al. 2001b; Tang et al. 2001). In an $F_2$ population, dominant

marker loci can create major problems such as mis-estimation of recombination frequencies and locus ordering problems in repulsion phase (Knapp et al. 1995). Therefore, for accurate estimation of recombination frequencies and locus orders, "pure-coupling" maps were developed. Dominant markers can be split into two groups: dominant alleles from the female parent and dominant alleles from the male parent. We first developed two "pure-coupling" maps for each linkage group, and then either merged them into one linkage group or selected one "coupling" map for constructing the $F_2$ RFLP-SSR integration map. In only one case, LG4, could two coupling maps be merged into one linkage map without repulsion problems.

In the HA370 x HA372 $F_2$ map, the length of the RFLP framework map (Gedil et al. 2001b) was extended by SSR loci (209.3 cM). SSR markers mapped distal to RFLP marker loci on 13 out of 34 chromosome arms. SSR markers tended to cluster in presumed centromeric regions of some chromosomes in both maps. There are two main hypotheses for the apparent clustering of SSR loci in centromeric regions. One is that recombination is suppressed in centromeric regions; thus linkage maps show short genetic distances in centromeric regions that actually span very large physical distances (Wu and Tanksley 1993). This is a common phenomenon detected in other dense maps of sunflower with different marker systems (Berry et al. 1996; Gedil et al. 2001b). The Gedil et al. (2001b) map had dense clusters of AFLP marker loci on several linkage groups, while the Berry et al. (1996) map had dense clusters of RFLP markers on every linkage group. A second hypothesis is that centromeric regions in higher eukaryotes usually contain massive blocks of pericentric heterochromatin composed of highly repeated DNA, and a high level of diversity is typical of the

centromeric highly repeated DNA families (Summer 1994). Clustering of SSRs has been reported in other plant species such as in wheat (Roder et al. 1998) and barley (Ramsay et al. 2000), but not in rice (Temnykh et al. 2000). However, clustering in the map of wheat is much less than that of barley. Heterochromatic DNA is more highly methylated than euchromatic DNA (Martienssen 1998). One way to improve genome coverage by enriching for gene-rich regions would be to use methylation sensitive enzymes such as *Pst I* in constructing small-insert genomic SSR libraries (Roder et al. 1998). Chen et al. (1997) reported clustering of specific SSR sequences such as (GA), (ATT) and (GATA) near centromeres in the rice genome. However, we could not examine the relationship between clustering and specific repeat sequences or motifs, because the libraries which were screened for polymorphism were primarily enriched for dinucleotides, (GA) and (CA).

The segregation distortion loci on the $F_2$ map (7.4%) was similar to that of previously published $F_2$ maps based on RFLP markers, from Gentzbittel et al. (1995) and Berry et al. (1995). However, the segregation distortion from the RIL population (27.5 %) was higher and the most prominent feature was the trend of favoring transmission of maternal alleles. Segregation distortion, or aberrant Mendelian segregation, is a genetic phenomenon of meiotic drive caused by competition among gametes for preferential fertilization in the pre- and postzygotic phase. This phenomenon occurred in both intra- and interspecific crosses and showed the trend of unidirectional genome transmission (Zamire and Tadmor 1986). The unidirectional deviation in the PHA x PHB population indicated that maternally-derived genomes are more likely to maintain their integrity. These genomes transmit more alleles to the

next generation than expected and consequently the progeny show a greater resemblance to one parent.

The high frequency of duplicated loci led to the hypothesis that sunflower is of amplipolyploid origin (Sossey-Alaoui et al. 1998). An alternative hypothesis is that SSR markers can be derived from moderately repeated DNA sequences, provided that their primer sequences are sufficiently specific to amplify only a single or very few loci (Roder et al. 1998). However, the question of duplicated loci in sunflower is still under study.

The current level of genome coverage already provides sufficient genetic markers to be useful for mapping agronomically important genes. Several agronomically important traits including candidate disease resistance and branching genes have been mapped and closely linked to SSR markers for marker-assisted selection in our laboratory. A concerted effort to generate and map more of these highly informative, PCR-based, and locus-specific markers will improve the resolution of the sunflower map and provide the basis for fine-scale genome analysis, positional cloning, and targeted genetic improvement.

## References

Bailey, N.T.J., (1996) Introduction to the mathematical theory of genetic linkage. Clarendon Press, Oxford.

Berry, S.T., Leon, A. J., Hanfrey, C.C., Challis, P., Burkholz, A., Barness, S., Rufener, G.K., Lee, M., and Caligari, P.D.S. (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. 91: 195-199.

Berry, S.T., Leon, A.J., Challis, P., Livin, C., Jones, R., Hanfrey, C.C., Griffiths, S., and Roberts, A. (1996) Construction of a high density, composite RFLP linkage map for cultivated sunflower *Helianthus annuus*. *In* "Proceedings of the 14th International sunflower conference, Vol. 2, pp. 1150-1160, Beijing, China.

Berry, S.T., Leon, A.J., Peerbolte, R., Challis, C., Livini, C., Jones, R., and Feingold, S. (1997). Presentation of the Advanta sunflower RFLP linkage map for public research. *In* "Proc. 19th Sunflower Res. Workshop", pp. 113-118, Fargo, USA.

Chen, X., Temnykn, S., Xu, Y., Cho, Y.G., and McCouch, S.R. (1997) Development of a microsatellite framework map providing genome-wide coverage in rice *Oryza sativa* L. Theor. Appl. Genet. 95: 553-567.

Gedil, M.A., (1999) Marker development, genome mapping, and cloning of candidate disease resistance genes in sunflower, *Helianthus annuus* L. Ph.D. Thesis, Oregon State Univ., Corvallis.

Gedil, M.A., Slabaugh, M.M., Berry, S., Segers, B., Peleman, J., Michelmore, R., Miller, J.F., Gulya, T., and Knapp, S.J. (2001a) Candidate disease resistance genes in sunflower cloned using conserved nucleotide binding site motifs: genetic mapping and linkage to downy mildew resistance gene *Pl*1 gene. Genome 44: 205-212.

Gedil, M.A., Wye, C., Berry, S., Segers, B., Peleman, J., Jones, R., Leon, A., Slabaugh, M.B., and Knapp, S.J. (2001b) An integrated restriction fragment length polymorphism - amplified fragment length polymorphism linkage map for cultivated sunflower. Genome 44: 213-221.

Gentzbittel, L., Vear, F., Zhang, Y.X., and Berville, A. (1995) Development of a consensus linkage RFLP map of cultivated sunflower "*Helianthus annuus*" L. Theor. Appl. Genet. 90: 1079-1086.

Gentzbittel, L., Mestries, E., Mouzeyar, S., Mazeyrat, F., Badaoui, S., Vear, F., Labrouhe, D.T.D., and Nicolas, P. (1999) A composite map of expressed sequences and phenotypic traits of the sunflower *Helianthus annuus* L. genome. Theor. Appl. Genet 99: 218-234.

Haldane, J.B.S., and Waddington, C.H. (1931) Inbreeding and linkage. Genetics 16: 357-374.

Holloway, J.L., and Knapp, S.J. (1993) Gmendel 3.0 Users Guide. Department of Crop and Soil Science, Oregon State University, OR, USA.

Jan, C.C., Vick, B.A., Miller, J.F., Kahler, A.L., and Butler III, E.T. (1998) Construction of an RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. 96: 15-22.

Kendall, M., and Gribbons, J.D. (1990) "Rank correlation methods," Oxford University. Press, New York.

Knapp, S.J., Holloway, J.L., Bridges, W.C., and Liu, B.H. (1995) Mapping dominant marker using F2 matings. Theor. Appl. Genet. 91: 74-81.

Kosambi, D.D. (1944) The estimation of map distance from recombination values. Ann. Eugen. 12: 172-175.

Lawson, W.R., Goulter, K.C., Henry, R.J., Kong, G.A. and Kochman, J.K. (1998) Marker-assisted selection for two rust resistance genes in sunflower. Mol. Breed. 4: 227-234.

Lincoln, S.E., and Lander, E.S. (1992) Systematic detection of errors in genetic linkage data. Genomics 14: 604-610.

Martienssen, R., (1998) Transposons, DNA methylation and gene control. Trends Genet. 14: 263-264

Miller, J.F, and Gulya, T.J. (1990) Registration of ten oilseed sunflower germplasm lines. Crop Sci. 30: 430-431.

Ramsay, L., Macaulay, M., Ivanissevich, S.D., Maclean, K., Cardle, L., Fuller, J., Edwards, K.J., Tuvesson, S., Morgante, M., Massari, A., Maetri,E., Marmiroli, N., Sjakste, T., Ganal, M., Powell, W., and Waugh, R. (2000) A simple sequence repeat-based linkage map of barley. Genetics 156: 1997-2005.

Roder, M.S., Korzum, V., Wendehake, K., Plaschke, J., Tixier, M.-H., Leroy, P., and Ganal, M.W. (1998) A microsatellite map of wheat. Genetics 149: 2007-2023.

Schilling, E.E., and Heiser, C.B. (1981) An infrageneric classification of *Helianthus*. Taxon 30: 393-403.

Sossey-Alaoui, K., Serieys, H., Tersc, M., Lambert, P., Schilling,E., Griveau, Y., and Berville, A. (1998) Evidence of several genomes in *Helianthus*. Theo. Appl. Genet. 97: 422-430.

Summer, A.T. (1994) Functional aspects of the longitudinal differentiation of chromosomes. Eur. J. Histochem. 38: 91-109.

Tang, S., Yu, J.-K, Slabaugh, M.B., Shintani, D.K., and Knapp, S.J. (2001) Simple sequence repeat map for the sunflower genome. Genetics (submitted)

Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, T. G., Ishii, T., and McCouch, S.R. (2000) Mapping and genome organization of microsatellite sequences in rice *Oryza sativa* L. Theo. Appl. Genet. 100: 697-712.

Yu, J.-K, Tang, S., Slabaugh, M.B., and Knapp, S.J. (2001) Allelic diversity of simple sequence repeat markers among elite inbred lines in cultivated sunflower. Genome (submitted)

Wang, Z., Weber, J.L., Zhong, G., and Tanksley, S.D. (1994) Survey of plant short tandem DNA repeats. Theor. Appl. Genet. 88: 1-6.

Webb, D.M., and Knapp, S.J. (1990) DNA extraction from a previously recalcitrant plant genus. Mol. Biol. Rep. 8:180-185.

Wu, K.S., and Tanksley, S.D. (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet. 241: 225-235.

Zamir, D., and Tadmor, T., (1995) Unequal segregation of nuclear genes in plants. Bot. Gaz. 147: 355-358

# CHAPTER 4

# CONCLUSION

Although sunflower is one of the world's major oilseed crops, the availability of DNA-based markers and dense genetic maps in the public domain is very limited. The objectives of this research were to develop DNA markers and generate genetic maps for various sunflower genetic analyses and molecular breeding.

The first project, the development of simple sequence repeat (SSR) markers, was undertaken to generate sequence based, high through-put markers and analyze genetic relationships among 16 inbred lines using newly developed SSR markers (Chapter 2). Using a small insert library enriched for a variety of repeat motifs, 259 unique SSR sequences were discovered including 176 dinucleotide, 60 trinucleotide and 23 tetranucleotide repeats. Of a total recovered, 171 were perfect repeats, 36 were imperfect repeats and 52 were compound repeats. Repeat length varied from four to 62 in the reference allele sequences and the mean length of tetranucleotide repeats (25.8) was significantly longer than that of dinucleotide (13.4) and trinucleotide (7.7) repeats. From a total of 259 unique SSR sequences, 171 primer pairs were designed. A total of 131 primer pairs produced single, discrete amplicons and were screened for length polymorphism on polyacrylamide gels using fluorescently labeled amplicons. Of the total, 74 markers were polymorphic among 16 inbred lines. The number of alleles ranged from one to 17 and mean number of alleles per locus among the polymorphic SSR markers was 3.7 for dinucleotide, 3.6 for trinucleotide and 9.5 for tetranucleotide repeat markers. Polymorphic information content (PIC) for

polymorphic markers varied from 0 to 0.93 and tetranucleotide markers (0.83) were significantly more polymorphic than shorter motif SSR markers; 0.53 for dinucleotide and 0.53 for trinucleotide markers. SSR markers are significantly more polymorphic than RAPDs (Arias and Rieseberg 1995), AFLPs (Hongtrakul et al. 1997), and allozymes (Cronn et al. 1997) in sunflower, and the mean PIC score was higher, but not significantly, than that of RFLPs from previous reports in cultivated sunflower (Berry et al. 1994; Gentzbittel et al. 1994; Zhang et al. 1995). (CA) repeats were more abundant and more polymorphic than (GA) repeats. The correlation between the number of repeat units in the reference allele sequences and PIC score was not significantly related.

The genetic distance ranged from 0.175 to 0.543 among the 16 elite inbred lines. The genetic relationship was demonstrated by principal component analysis and cluster analysis. Principal component and cluster analyses of the genetic distance matrix, estimated from 74 polymorphic SSR marker loci, uncovered patterns of genetic diversity. The plot of the first and second principal component scores showed that, for the most part, lines from different market (oilseed and confectionery) and fertility restorer (maintainer and restorer) classes fell into separate quadrants. These patterns are similar to those uncovered with RFLPs (Berry et al. 1994; Gentzbittel et al. 1994; Zhang et al. 1995) and AFLPs (Hongtrakul et al. 1997). The dendrogram produced by UPGMA cluster analysis of the genetic distance matrix shows the same pattern as PCA analysis. SSRs are a powerful tool for fingerprinting inbred lines and their development overcomes the previous lack of highly polymorphic DNA markers in cultivated sunflower.

Even though a relatively dense sunflower RFLP map was developed, the number of mapped public DNA markers is still limited, and separate sets of RFLP probes were used in different mapping populations (Berry et al. 1995, 1996; Genztbittel et al. 1994, 1999; Jan et al. 1998). The second project, the development of sunflower genetic maps, was undertaken to construct an integrated public RFLP-SSR map in HA370 x HA372, an $F_2$ population, and an SSR map in PHA x PHB, a recombinant inbred line population (Chapter 3). A total of 1,090 SSR markers were screened for polymorphism between the parents of the two mapping populations. Of the total, 148 SSR markers were polymorphic in both populations and 351 SSR markers were polymorphic in one or the other. The two maps were developed in parallel with the RHA280 x RHA801 RIL map described by Tang et al. (2001). This allowed us to coalesce marker loci into 17 linkage groups corresponding to the 17 chromosomes of cultivated sunflower. A total of 122 SSR loci and three miscellaneous marker loci were integrated into 77 preexisting RFLP marker loci in the HA370 x HA372 $F_2$ population. Of the total SSR loci, 83 (68.0 %) were co-dominant. The map was 1348.0 cM long and had a mean density of 6.77 cM. The number of SSR marker loci per linkage group ranged from two on LG6 and LG8 to 16 on LG17, while the number of RFLP marker loci per linkage group ranged from one on LG12 and LG15 to seven on LG3. SSR marker loci mapped distal to RFLP markers on 13 chromosome arms and extended genome coverage of the RFLP HA370 x HA372 map developed by Gedil et al. (2001) by 21.8%. In addition, SSR marker loci added between RFLP loci increased the length of the same map by 18.5%. The segregation

ratios for 6 out of 122 SSR marker loci were significantly distorted but distorted loci were not clustered in the HA370 x HA372 map.

The SSR map was constructed using an $F_6$ population from a cross between PHA and PHB. The map included 276 SSR loci, two SCAR marker loci, and one morphological locus. The SSR marker loci were mapped into 23 linkage groups; LG8, LG9 and LG16 were each comprised of two linkage group fragments. Eight SSR markers (three linkage groups) could not be assigned to any known linkage groups. The map was 1377.4 cM long and had a mean density of 4.99 cM. A total of 57 loci showed a high rate of heterozygosity and the mean heterozygote frequency across SSR marker loci was 0.0387, which was slightly greater than the expected mean heterozygote frequency for $F_6$ RILs (0.0313). In addition, 25 out of 57 high heterozygosity loci caused segregation distortion. SSR marker loci with significant segregation distortion ratios (76 out of 267) were clustered in centromeric or distal regions and most had an excess of PHA alleles. The new maps created a dense framework of SSR markers for constructing and cross-referencing genetic maps and assigning new loci to the genetic map of sunflower for breeding programs.

This study presented i) the development of SSR markers and estimation of SSR marker utility, and ii) the construction of an SSR map and an integrated RFLP-SSR map for the public sector.

# BIBLIOGRAPHY

Ahnert, D., Lee, M., Austin, D.F., Livini, C., Woodman, W.L., Openshaw, S.J., Smith, J.S.C., Porter, K., and Dalton, G. (1996) Genetic diversity among elite sorgumn inbred lines assessed with DNA markers and pedigree information. Crop Sci. 36: 1358-1392.

Akagi, H., Yokozeki, Y., Inagaki, A., and Fujimura, T. (1997) Highly polymorphic microsatellites of rice consists of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor. Appl. Genet. 94: 61-67.

Akkaya, M.S., Bhagwat, A.A., and Cregan, P.B. (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132: 1131-1139.

Anderson, J.A., Churchill, G.A., Sutrique, J.E., Tanksley, S.D., and Sorrels, M.E. (1993) Optimizing parental selection for genetic linkage maps. Genome 36: 181-186.

Arias, D.M., and Rieseberg, L.H. (1995) Genetic relationships among domesticated and wild sunflowers (*Helianthus annuus*, *Asteraceae*). Econ. Bot. 49: 239 248.

Ashkenazi, V., Chani, E., Lavi, U., Levy, D., Hillel, J., and Velleux, R.E. (2001) Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. Genome 44: 50-62.

Bailey, N.T.J., (1996) Introduction to the mathematical theory of genetic linkage. Clarendon Press, Oxford.

Berry, S.T., Allen, R.J., Barnes, S.R., and Caligari, P.D.S. (1994) Molecular marker analysis of *Helianthus annuus* L. 1.Restriction fragment length polymorphism between inbred lines of cultivated sunflower. Theor. Appl. Genet. 89: 435-441.

Berry, S.T., Leon, A.J., Hanfrey, C.C., Challis, P., Burkholz, A., Barness, S., Rufener, G.K., Lee, M., and Caligari, P.D.S. (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. 91: 195-199.

Berry, S.T., Leon, A.J., Peerbolte, R., Challis, C., Livini, C., Jones, R., and Feingold, S. (1997). Presentation of the Advanta sunflower RFLP linkage map for public research. *In* "Proc. 19th Sunflower Res. Workshop", pp. 113-118, Fargo, USA.

Berry, S.T., Leon, A.J., Challis, P., Livin, C., Jones, R., Hanfrey, C.C., Griffiths, S., and Roberts, A. (1996) Construction of a high density, composite RFLP linkage map for cultivated sunflower *Helianthus annuus*. *In* "Proceedings of the 14th International sunflower conference, Vol. 2, pp. 1150-1160, Beijing, China.

Besnard, G., Griveau, Y., Quillet, M.C., Serieys, H., Lambert, P., Vares, D. and Berville, A. (1997) Specifying the introgressed segments from *H. argophyllus* in cultivated sunflower (*Helianthus annuus* L.) to mark *Phomopsis* resistance genes. Theor. Appl. Genet. 94: 131-138.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32: 314-331.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, A., Minch, E., Kidd, J.R., and Cavalli-Sforza, L.L. (1994) High resolution human evolutionary threes from polymorphic microsatellites. Nature 368: 455-457.

Broun, P., and Tanksley, S.D. (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol. Gen. Genet. 250: 39-49.

Brownstein, M.J., Carpten, J.D., and Smith J.R. (1996) Modulation of non-template nucleotide addition by Taq DNA Polymerase: primer modifications that facilitate genotyping. BioTechniques 20: 1004-1010.

Brunel, D. (1994) A microsatellite marker in *Helianthus annuus* L. Plant Mol. Biol. 24: 397-400.

Chakraborty, R., Kimmel, M., Stivers, D.N., Davison, L.J., and Deka, R. (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA 94: 1041-1046.

Charters, Y.M., Robertson, A., Wilkinson, M.J., and Ramsay, G. (1996) PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. oleifera) using 5'-anchored simple sequence repeat (SSR) primers. Theor. Appl. Genet 92: 442-447.

Chen, X., Temnykn, S., Xu, Y., Cho, Y.G., and McCouch, S.R. (1997) Development of a microsatellite framework map providing genome-wide coverage in rice *Oryza sativa* L. Theor. Appl. Genet. 95: 553-567.

Cheres, M.T., and Knapp, S.J. (1998) Ancestral origins and genetic diversity of cultivated sunflower: analysis of the pedigrees of public germplasm sources. Crop Sci. 38: 1476-1482.

Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCough, S.R., Park, W.D., Ayres, N. and Cartinhour, S. (2000) Diversity of microsatellites derived from genomic libraries and GeneBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100: 713-722

Cronn, R., Brothers, M., Klier, K., Bretting, P.K., and Wendel, J.F. (1997) Allozyme variation in domesticated annual sunflower and its wild relatives. Theor. Appl. Genet. 95: 532-545.

Dehmer, K.J., and Friedt, W. (1998) Evaluation of different microsatellite motifs for analysing genetic relationships in cultivated sunflower (*Helianthus annuus* L.). Plant Breed. 117: 45-48.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., and Mattick, J.I. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 19: 4008.

Edwards, K.J., Barker, J.H. A., Daly, A., Jones, C., and Karp, A. (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20: 758-760.

Fick, G.N. and Miller, J.F. (1997) Sunflower Breeding, *in* Sunflower technology and production, Schneiter, A.A. (ed.). Amer. Soc. Agronomy, Madison, WI. pp 395-439.

Gedil, M.A., Slabaugh, M.M., Berry, S., Segers, B., Peleman, J., Michelmore, R., Miller, J.F., Gulya, T., and Knapp, S.J. (2001a) Candidate disease resistance genes in sunflower cloned using conserved nucleotide binding site motifs: genetic mapping and linkage to downy mildew resistance gene *Pl*1 gene. Genome 44: 205-212.

Gedil, M.A., Wye, C., Berry, S., Segers, B., Peleman, J., Jones, R., Leon, A., Slabaugh, M.B., and Knapp, S.J. (2001b) An integrated restriction fragment length polymorphism - amplified fragment length polymorphism linkage map for cultivated sunflower. Genome 44: 213-221.

Gedil, M.A., (1999) Marker development, genome mapping, and cloning of candidate disease resistance genes in sunflower, *Helianthus annuus* L. Ph.D. Thesis, Oregon State Univ., Corvallis.

Gentzbittel, L., Mestries, E., Mouzeyar, S., Mazeyrat, F., Badaoui, S., Vear, F., Labrouhe, D.T.D., and Nicolas, P. (1999) A composite map of expressed sequences and phenotypic traits of the sunflower *Helianthus annuus* L. genome. Theor. Appl. Genet 99: 218-234.

Gentzbittel, L., Mouzeyar, S., Badaoui, S., Mestries, E., Vear, F., De Labrouhe, D.T., and Nicolas, P. (1998) Cloning of molecular markers for disease resistance in sunflower, "*Helianthus annuus*" L. Theor. Appl. Genet. 96: 519-525.

Gentzbittel, L., Vear, F., Zhang, Y.X., and Berville, A. (1995) Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). Theor. Appl. Genet. 90: 1079-1086.

Gentzbittel, L., Zhang, Y.X., Vear, F., Griveau, B., and Nicolas, P. (1992) RFLP studies of genetic relationships among inbred lines of the cultivated sunflower, *Helianthus annuus* L.: evidence for distinct restorer and maintainer germplasm pools. Theor. Appl. Genet. 89: 419-425.

Gentzbittel, L., Zhang, Y.X., Vear, F., and Griveau, B. (1994) RFLP studies of genetic relationships among inbred lines of the cultivated sunflower, *Helianthus annuus* L.: evidence for distinct restorer and maintainer germplasm pools. Theor. Appl. Genet. 89: 419-425.

Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L., and Feldman, M.W. (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463-471.

Haldane, J.B.S., and Waddington, C.H. (1931). Inbreeding and linkage. Genetics 16: 357-374.

Hamada, H., and Kakunaga, T. (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298: 396-398.

Hamada, H., Petrino, M. G., and Kakunaga, T. (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA 79: 6465-6469.

Holloway, J.L., and Knapp, S.J. (1993) Gmendel 3.0 Users Guide. Department of Crop and Soil Science, Oregon State University, OR, USA.

Hongtrakul, V., Huestis, G.M., and Knapp, S.J. (1997) Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theor. Appl. Genet. 95: 400-407.

Hongtrakul, V., Slabaugh, M.B., and Knapp, S.J. (1998a) A seed specific Delta-12 oleate desaturase gene is duplicated, rearranged, and weakly expressed in high oleic acid sunflower lines. Crop Sci. 38: 1245-1249.

Hongtrakul, V., Slabaugh, M.B., and Knapp, S.J. (1998b) DFLP, SSCP, and SSR markers for D9 stearoyl-acyl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are hypervariable among elite inbred lines. Mol. Breed. 4: 195-203.

Ishii, T., and McCough, S.R. (2000) Microsatellites and microsynteny in the chloroplate genome of *Oryza* and eight other *Gramineae* species. Theor. Appl. Genet. 1000: 1257-1266.

Jan, C.C., Vick, B.A., Miller, J.F., Kahler, A.L., and Butler III, E.T. (1998) Construction of an RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. 96: 15-22.

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L., and Armour, J.A.L. (1994) Complex gene conversion events in germline mutation at human minisatellites. Nature Genetics 6: 136-145.

Jones, N., Ougham, H., and Thomas, H. (1997) Markers and mapping: we are all geneticists now. New Phytologist 137: 165-177.

Kendall, M., and Gribbons, J.D. (1990) "Rank correlation methods," Oxford University. Press, New York.

Knapp, S.J., Holloway, J.L., Bridges, W.C., and Liu, B.H. (1995) Mapping dominant marker using F2 matings. Theor. Appl. Genet. 91: 74-81.

Kosambi, D.D. (1944) The estimation of map distance from recombination values. Ann. Eugen. 12: 172-175.

Kresovich, S., Szewc-McFadden, A.K., and Blick, S.M. (1995) Abundance and characterization of simple sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L.(rapeseed). Theor. Appl. Genet. 91: 206-211.

Lagercrantz, U., Ellegren, H., and Anderson, L. (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Research 21:1111-1115.

Lawson, W.R., Goulter, K.C., Henry, R.J., Kong, G.A. and Kochman, J.K. (1998) Marker-assisted selection for two rust resistance genes in sunflower. Mol. Breed. 4: 227-234.

Lawson, W.R., Henry, R.J., Kochman, J.K., and Kong, G.A. (1994) Genetic diversity in sunflower (*Helianthus annuus* L.) as revealed by Random Amplified Polymorphic DNA analysis. Aust. J. Agric. Res. 45: 1319-1327.

Lee, M. (1995) DNA markers and plant breeding programs. Advances in Agronomy 55: 265-344.

Leon, A.J., Lee, M., Rufener, G.K., Berry S.T. and Mowers, R.P. (1995) Use of RFLP markers for genetic linkage analysis of oil percentage in sunflower seed. Crop Sci. 35: 558-564.

Levison, G., and Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4: 203-221.

Lincoln, S.E., Daly, M.J. and Lander, E.S. (1991) PRIMER: A computer program for automatically selecting PCR primers. MIT. Cambridge, MA.

Lincoln, S.E. and Lander, E.S. (1992) Systematic detection of errors in genetic linkage data. Genomics 14: 604-610.

Mansfield, D.C., Brown, A.F., Green, D.K., Carothers, A.D., Morris, S.W., Evans, H.J., and Wright, A.F. (1994) Automation of genetic linkage analysis using fluorescent microsatellite markers. Genomics 24: 225-233.

Mantel, N.A. (1967) The detection of disease clustering and a generalized regression approach. Can. Res. 27: 209-220.

Martienssen, R., (1998) Transposons, DNA methylation and gene control. Trends Genet. 14: 263-264

McCouch, S.R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y.G., Huang, N., Ishii, T., and Blair, M. (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol. Biol. 35: 89-99.

Mestries, E., Gentzbittel, L., Tourvieille de Labrouhe, D., Vicolas, P., and Vear, F. (1998) Analyses of quantitative trait loci associated with resistance to *Sclerotinia sclerotiorum* in sunflower (*Helianthus annuus* L.). Mol. Breed. 4: 215-226.

Miller, J.F, and Gulya, T.G. (1990) Registration of ten oilseed sunflower germplasm lines. Crop Sci. 30: 430-431.

Minch, E., Ruiz-Linares, A., Goldstein, D., Feldman, M., and Cavalli-Sforza, L.L. (1997). MICROSAT: A computer program for calculating various statistics on microsatellite allele data. (http://crick.stanford.edu/microsat/microsat.html)

Morgante, M., and Olivieri, A.M. (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J. 3:175-182.

Olufowote, J.O., Xu, Y., Chen, X., Park, W.D., Beachell, H.M., Dilday, R.H., Goto, M., and McCouch, S.R. (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40: 370-378.

Plaschke, J., Ganal, M.W., and Roder, M.S. (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor. Appl. Genet. 91: 1001-1007.

Powell, W., Machray, G.C., and Provan, J. (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1: 215-222.

Putt, E.D. (1997) Early history of sunflower, *in* Sunflower technology and production, Schneiter, A.A. (ed.). Amer. Soc. Agronomy, Madison, WI. pp 1-20.

Rae, S.J., Aldam, C., Dominguez, I., Hoebrechts, M., Barnes, S.R., and Edwards, K.J. (2000) Development and incorporation of microsatellite markers into the linkage map of sugar beet (*Beta vulgaris* spp). Theo. Appl. Genet. 1000: 1240-1248.

Ramsay, L., Macaulay, M., Ivanissevich, S.D., Maclean, K., Cardle, L., Fuller, J., Edwards, K.J., Tuvesson, S., Morgante, M., Massari, A., E.Maetri, N.Marmiroli, Sjakste, T., Ganal, M., Powell, W., and Waugh, R. (2000) A simple sequence repeat-based linkage map of barley. Genetics 156: 1997-2005.

Rieseberg, L.H., Choi, H., Chan, R., and Spore, C. (1993) Genomic map of a diploid hybrid species. Heredity 70: 285-293.

Roder, M.S., Korzum, V., Wendehake, K., Plaschke, J., Tixier, M.-H., Leroy, P., and Ganal, M.W. (1998) A microsatellite map of wheat. Genetics 149: 2007-2023.

Roder, M.S., Plaschke, J., Konig, S.U., Borner, A., Sorrels, M.E., Tanksley, S.D., and Ganal, M.W. (1995) Abundance, variability and chromosomal location of microsatellite in wheat. Mol. Gen. Genet. 246: 327-333.

Rongwen, J., Akkaya, M.S., Bhagwat, A.A., Lavi, U., and Cregan, P.B. (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90: 43-48.

Sackston, W.E. (1992) On a treadmill: Breeding sunflowers for resistance to disease. Ann. Rev. Phytopathol. 30: 529-551.

Saghai Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q., and Allard, R.W. (1994) Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. USA 91: 5466-5470.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: A laboratory manual. New York: Cold Spring Harbor.

Schilling, E.E., and Heiser, C.B. (1981) An infrageneric classification of *Helianthus*. Taxon 30: 393-403.

Smith, J.S., Chin, E.C.L., Shu, H., Smith, O.S., Wall, S.J., Senior, M.L., Mitchell, S.E., Kresovich, S., and Ziegle, J. (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95: 163-173.

Smith, J.S.C., Kresovich, S., Hopkins, M.S., Mitchell, S.E., Dean, R.E., Woodman, W.L., Lee, M., and Porter, K. (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequences repeats. Crop Sci. 40:226-232.

Smith, O.S., Smith, J.S.C., Bowen, S.L., Tenborg, R.A., and Wall, S.J. (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F1 heterosis, and RFLPs. Theor. Appl. Genet. 80: 833-840.

Sossey-Alaoui, K., Serieys, H., Tersc, M., Lambert, P., Schilling, E., Griveau, Y., and Berville, A. (1998) Evidence of several genomes in *Helianthus*. Theo. Appl. Genet. 97: 422-430.

Struss, D., and Plieske, J. (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor. Appl. Genet. 97: 308-315.

Summer, A.T. (1994) Functional aspects of the longitudinal differentiation of chromosomes. Eur. J. Histochem. 38: 91-109.

Tang, S., Yu, J.-K., Slabaugh, M.B., Shintani, D.K., and Knapp, S.J. (2001) Simple sequence repeat map for the sunflower. Genetics (submitted)

Tautz, D., and Renz, M. (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research 12: 4127-4137.

Tautz, D. (1989) Hypervariability of simple sequences as a general resource for polymorphic DNA markers. Nucleic Acids Research 17: 6463-6471.

Temnykh, S., Park, W.D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, T. G., Ishii, T., and McCouch, S.R. (2000) Mapping and genome organization of microsatellite sequences in rice *Oryza sativa* L. Theo. Appl. Genet. 100: 697-712.

Ungerer, M.C., Baird, J., Pan, J., and Rieseberg, L.H. (1998) Rapid hybrid speciation in wild sunflowers. Proc. Natl. Acad. Sci. USA 95: 11757-11762.

Vear, F., Gentzbittel, L., Philippon, J., Mouzeyar, S., Mestries, E., Roeckel-Drevet, P., Tourvielle De Labrouhe, D. and Nicolas, P. (1997) The genetics of resistance to five races of downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). Theor. Appl. Genet. 95: 584-589.

Wang, G., Mahalingam, R., and Knap, H.T. (1998) (C-A) and (G-A) anchored simple sequence repeats (ASSRs) generated polymorphism in soybean, *Glycine max* (L.) Merr. Theor. Appl. Genet. 96: 1086-1096.

Wang, Z., Weber, J.L., Zhong, G., and Tanksley, S.D. (1994) Survey of plant short tandem DNA repeats. Theor. Appl. Genet. 88: 1-6.

Webb, D.M., and Knapp, S.J. (1990) DNA extraction from a previously recalcitrant plant genus. Mol. Biol. Rep. 8: 180-185.

Weber, J.L., and May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388-396.

Whitton, J., Rieseberg, L.H., and Ungerer, M.C. (1997) Microsatellite loci are not conserved across the *Asteraceae*. Mol. Biol. Evol. 14. 204-209.

Wu, K.S., and Tanksley, S.D. (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gen. Genet. 241: 225-235.

Yu, J.-K, Tang, S., Slabaugh, M.B., and Knapp, S.J. (2001) Allelic diversity of simple sequence repeat markers among elite inbred lines in cultivated sunflower. Genome (submitted)

Zamir, D., and Tadmor, T. (1986) Unequal segregation of nuclear genes in plants. Bot. Gaz. 147: 355-358.

Zhang, Y.X., Gentzbittel, L., and Nicolas, P. (1995) Assessment of inter- and intra-inbred line variability in sunflower (*Helianthus annuus*) by RFLPs. Genome 38: 1040-1048.