Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.)

Norma Paniego, Mercedes Echaide, Marianne Muñoz, Luis Fernández, Susana Torales, Paula Faccio, Irma Fuxan, Mónica Carrera, Rubén Zandomeni, Enrique Y. Suárez, and H. Esteban Hopp

Abstract: Development of microsatellite markers for sunflower (*Helianthus annuus* L.) was performed to estimate their frequency, nature (structure), levels of polymorphism, usefulness for genotype identification, and calculation of genetic relationships between inbred lines representing the species diversity. Isolation was performed from a small-insert genomic library followed by hybridization screening using oligonucleotide probes containing different nucleotide arrays. In this work, 503 unique microsatellite clones were sequenced and 271 PCR primer sequences bordering the microsatellite repeat were designed. For polymorphism assessment, 16 *H. annuus* germplasm accessions were checked and 170 of the primers tested were shown to be polymorphic for the selected lines. The polymorphic microsatellites produced an average of 3.5 alleles/locus and an average polymorphism information content (PIC) of 0.55. The most frequently found motifs within polymorphic simple-sequence repeats (SSRs) were: $(GA)_n$, $(GT)_n$, $(AT)_n$, followed by trinucleotides $(ATT)_n$, $(TGG)_n$, and $(ATC)_n$, and the tetranucleotide $(CATA)_n$. Most of the 170 SSRs obtained showed important differences in the 16 reference inbred lines used for their characterization. In this work, 20 of the most informative SSRs destined to sunflower genotyping and legal fingerprinting purposes are fully described.

Key words: sunflower, molecular markers, microsatellites, simple-sequence repeats.

Résumé : Le développement et caractérisation de marqueurs microsatellites chez le tournesol (*Helianthus annuus*) ont été accomplis pour estimer leur fréquence, naturesse (structure), niveau du polymorphisme, utilité pour l'identification et calcule des relations génétiques entre lignées endocries qui représentent la diversité génétique du tournesol. L'isolement a été accompli en utilisant une banque génomique de petits inserts de tournesol. Dans ce but, le criblage basé sur l'hybridation a été effectué en utilisant des sondes oligonucléotidiques composées de différents arrangements de nucléotides. 503 clones microsatellites positifs ont été séquences et 271 séquences d'amorces adjacentes à ces microsatellites ont été synthétisées. Pour l'évaluation du polymorphisme, 16 sources de accessions du *H. annuus* ont été vérifiées et 170 des amorces testées ont montré un polymorphisme pour les lignées étudiées. Les microsatellites polymorphiques ont produit une moyenne de 3.5 allèles par gène et une moyenne de 0.5 PIC (contenu en information de polymorphisme). Les motifs les plus fréquemment trouvés dans les SSR polymorphes étaient: $(GA)_n$, $(GT)_n$, $(AT)_n$, suivis des trinucléotides $(ATT)_n$, $(TGG)_n$ et $(ATC)_n$ et du tetranucléotide $(CATA)_n$. La majeure partie des 170 séquence simple répétitive (SSR) obtenues ont montré des différences importantes dans les 16 lignées pures de référence utilisées pour sa caractérisation, 20 des SSR les plus instructives en termes d'information destinée à la détermination de génotypes de tournesol, et d'empreintes digitales dans un but légal, sont amplement décrits dans ce travail.

Mots clés : tournesol, marqueur moléculaire, microsatellites, séquence simple répétitive.

[Traduit par la Rédaction]

Received March 6, 2001. Accepted September 14, 2001. Published on the NRC Research Press Web site at http://genome.nrc.ca on December 5, 2001.

Corresponding Editor: J.B. Bell.

N. Paniego,[1] M. Echaide,[2] M. Muñoz, L. Fernández, S. Torales, P. Faccio, I. Fuxan, M. Carrera, R. Zandomeni,[3] E.Y. Suárez,[3] and H.E. Hopp.[3,4] Institutos de Biotecnología, de Recursos Biológicos y Laboratorio de Alta Complejidad Instituto de Microbiología y Zoología Agrícola , CNIA–INTA Castelar, CC 25, (1712) Castelar, Pcia. de Buenos Aires, Argentina.

[1]Corresponding author (e-mail: npaniego@cnia.inta.gov.ar).
[2]Present address: Laboratorio de Marcadores Moleculares, Instituto Nacional de Semillas, Av. Paseo Colon 922 4 piso, (1063) Ciudad de Buenos Aires, Argentina.
[3]Authors who directed different aspects of the work.
[4]Also belongs to Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, (1428) Ciudad de Buenos Aires, Argentina.

 DOI: 10.1139/G01-120

Introduction

Sunflower (*Helianthus annuus*) is the second most important oil crop worldwide, after soybean. Despite its economical importance, sunflower molecular genetics and genomics have remained unexplored until the last five years, during which four restriction fragment length polymorphism (RFLP) linkage maps of cultivated sunflower were published (Berry et al. 1995; Gentzbittel et al. 1995, 1999; Jan et al. 1998; Gedil et al. 2001). At the same time, genetic-diversity and coancestry analyses were carried out using random amplified polymorphic DNA (RAPD) (Arias and Rieseberd 1995), RFLP (Berry et al. 1995; Gentzbittel et al. 1995), and amplified fragment length polymorphism (AFLP) (Hongtrakul et al. 1997, Cheres and Knapp 1998) techniques. Considering that the progress in the genomic analysis of sunflower relies on the availability of molecular markers having a high polymorphic content, INTA and a consortium of seed companies have joined efforts for the isolation and characterization of microsatellites.

Microsatellites, or simple-sequence repeats (SSRs), are short sequence elements arranged in simple internal repeat structure (Tautz 1989) that are densely and randomly distributed throughout eukaryotic genomes. The number of microsatellites has been shown to be highly variable within and between species and subspecies. Because of their high mutation rate, they constitute the molecular markers with the highest polymorphic information content (PIC). This characteristic has promoted the application of microsatellites as molecular markers in fingerprinting (Weising et al. 1995; Diwan and Cregan 1997; Ashikawa et al. 1999), genome mapping (Marino et al. 1995; Broun and Tanksley 1996; McCouch et al. 1997; Röder et al. 1998; Winter et al. 1999; Scotti et al. 2000), phylogenetic and genetic relationship studies (Goldstein and Pollock 1997), marker assisted breeding, and population genetics (Goldstein et al. 1999). In addition to being highly polymorphic and well represented within eukaryotic genomes, the advantages of microsatellites as molecular markers include their codominant inheritance, easy scoring of the alleles, reproducibility, and accessibility to laboratories lacking highly sophisticated analysis equipment. These characteristics have strongly encouraged the microsatellite development in different plant species such as soybean (*Glycine max*; Akkaya et al. 1992; Morgante and Olivieri 1993; Cregan et al. 1994; Rongwen et al. 1995; Maughan et al. 1995), rice (*Oryza sativa*; Wu and Tanksley 1993; Panaud et al. 1995; McCouch et al. 1997), maize (*Zea mays*; Senior and Heun 1993), grape (*Vitis vinifera* L.; Thomas and Scott 1993), *Arabidopsis thaliana* (Bell and Ecker 1994), wheat (*Triticum aestiviun* L.; Röder et al. 1995), rapeseed (*Brassica napus*; Kresovich et al. 1995), kiwifruit (*Actinidia chinensis*; Weising et al. 1996), common bean (*Phaseolus vuglaris*; Yu et al. 1999), pine (*Pinus strobus*; Echt et al. 1996), chickpea (*Cicer arietinum* L.; Hüttel et al. 1999), lettuce (*Lactuca sativa* L.; van de Wiel et al. 1999), and others. The results of these surveys confirmed that microsatellites are also abundant and well distributed in plant genomes.

This work describes the survey of microsatellite markers in sunflower to estimate their frequency, nature (structure), ease of clean, locus-specific marker development, levels of polymorphism, usefulness for genotype identification, and calculation of genetic relationships between inbred lines representing the species diversity.

Materials and methods

Plant material

DNA from a commercial hybrid was used for the development of genomic libraries. Sixteen elite inbred sunflowers, described in Table 1, were used to screen for microsatellite polymorphisms. Ing. Agr. Carlos Sala (Nidera S.A., Venado) and Ing. Agr. Guillermo Pozzi Jáuregui (Tuerto, Argentina) kindly provided details on the pedigree information of these inbred lines. Part of this information is available at http://www.ag.ndsu.nodak.edu/aginfo/seedstock/varieties/VH-SUNF.htm.

Construction of small insert genomic library

Plants were grown in the greenhouse and total DNA was extracted from leaf tissue belonging to a single plant by cetyltrimethylammonium bromide (CTAB) protocol (Saghai-Maroof et al. 1984) and quantified by gel-electrophoresis staining comparison. Single- and double-digestion assays were conducted using different enzymes such as *Alu*I, *Dra*I; *Hae*III, *Hinf*I, *Hha*I, *Mse*I, *Ssp*I, and *Tsp*509I. Finally, 10 μg of sunflower genomic DNA were digested with *Tsp*509I (New England Biolabs, Beverly, Mass.) and fragments from 300 to 700 bp in size were isolated from agarose gels using the QiaEx purification kit following manufacturer recomendations (Qiagen, Hilden, Germany). The selected fragments were ligated into the *Eco*RI restriction endonuclease site of the λ-Zap II vector (Stratagene, La Jolla, Calif.). The library was amplified in *Escherichia coli* DH5αF′ according to manufacturer instructions. The efficiency of the library was 5×10^5 and 4.4×10^8 plaque forming units (PFU)/mL after amplification.

Isolation of clones containing microsatellite motifs

The library was plated and plaque lifts were made using Hybond N+ filters (Amersham-Pharmacia, Little Chalfont, Buckinghamshire, U.K.). The phage filters were probed with a set of ^{32}P-labeled synthetic poly di-, tri-, and tetranucleotides. The oligonucleotides were labeled using either [α-^{32}P]dCTP (specific activity 3000 Ci/μg, New England Nuclear, Boston, Mass.) and a random primer labeling kit (Prime-a-gene, Promega, Madison, Wis.) to label guanosine-containing oligonucleotides, or [α-^{32}P]dATP (specific activity 3000 Ci/μg, New England Nuclear) to label timidine-containing oligonucleotides. Filters probed with (GA)- or (GT)-rich motifs were prehybridized and hybridized in PAES-based hybridization mixture (0.1% polyanetholesulfonic acid (Sigma, St. Louis, Mo.); 0.01% sodium pyrophosphate, 1% SDS; and 2.5% 20× SSPE buffer (Sigma)) for 16 h at 55°C and then washed 5 min at room temperature with a stringency of 6× SSC and 0.1% SDS, followed by 4 min at 50°C with the same stringency. Filters probed with AT-rich motifs were hybridized at 40°C and then washed twice at room temperature with a stringency of 6× SSC and 0.1% SDS in a procedure similar to that described by Sambrook et al. (1989). Screening of the libraries was performed by making two rounds of plaque hybridizations. Positive clones were identified by autoradiography.

Table 1. Description of the inbred sunflower lines used for polymorphism screening.

Inbred lines	Pedigree	Location of breeding or reselection	Commercial use
H 52	Putatively Romanian germplasm[a]	South Africa	Oilseed maintainer
HA 61	'953-88-3'/'Armavirski 3497'	U.S.A.	Oilseed maintainer
HA 89	'Vniimk 8931'	U.S.A.	Oilseed maintainer
HA 292	'Commander'*3/'Mennonite RR'[b]	U.S.A.	Non-oilseed maintainer
HA 303	'Voshod'	U.S.A.	Oilseed maintainer
HA 369	'Teguá INTA' (Arg. 8018)	Argentine	Oilseed maintainer
HA 370	'RK-74-198'	South Africa	Oilseed maintainer
HA 821	'HA 300' (derived from 'Peredovik 301')	U.S.A.	Oilseed maintainer
HA-R2	'Impira INTA' Selection 5	Argentine	Oilseed maintainer
HA-R3	'Charata INTA'[c] selection	Argentine	Oilseed maintainer
HA-R5	'Guayacán INTA'[d] selection	Argentine	Oilseed maintainer
KLM 280	'KLM'[e] selection	Argentine	Oilseed maintainer
RHA 274	('cmsPI343765'/'Ha119'/'Ha64-4-5')/T66006-2[f]	U.S.A.	Oilseed restorer
RHA 293	'HA155'/'HIR34'/2/'RHA282'	U.S.A.	Non-oilseed restorer
RHA 374	'Arg-R43'	U.S.A.	Oilseed restorer
V 94[g]	'Mp543'* h./*H. argophyllus*	Argentine	Oilseed maintainer

[a]'H52' is an accession putatively originating from Romanian germplasm bred in Potchestfrom, Transvaal, South Africa.
[b]Third generation backcross of 'Mennonite RR' to 'Commander'.
[c]'Charata INTA' was obtained by interspecific crossings with wild germplasm belonging to species *H. annuus* subsp. *annuus* and *H. petiolaris*.
[d]'Guayacán INTA' derived from a cross between the Argentine variety Klein and 'CM953-102' and backcrossed once again with 'Klein'.
[e]'KLM' is a multiple cross between cultivars Klein × Local (a pool of local varieties of INTA Pergamino breeding program like 'Guayacán INTA', 'Charata INTA', etc.) × 'Manfredi' (a pool of varieties from INTA Manfredi breeding program like 'Impira INTA', 'Cordobés INTA', 'Manfredi INTA', etc.).
[f]T66006-2 comes from Peredovik*2/953-102-1-1-41.
[g]'V94' is another Argentine selection of a cross between cultivated sunflower ('MP543') and wild species (*H. argophyllus*), 'MP543' derives from 'MPRR' (mezcla precoz resistente a roya; pool of early material resistant to sunflower rust), which also derives from wide crossings with *Helianthus* wild species.

Nucleotide sequencing and primer design

Positive clones were isolated and excised from the λ-Zap II vectors using the Exasist/SOLR system (Stratagene) and plasmids were sequenced using the Sanger protocol and a DNA ABI 373 automatic sequencing device (Perkin Elmer, Foster City, Calif.). Sequences were edited manually to remove vector and ambiguous sequences at the ends. The trimmed sequences were stored as a text file to be compared by pairwise alignment against a data base containing all the sequences accumulatively generated by this project. In this process, those regions containing microsatellite repeats were filtered with the purpose of avoiding spurious similarities. Those pairs of sequences showing similarity greater than 85% were considered redundant. Finally, primers for unique sequences flanking the microsatellite motif were designed using a combination of computer programs, such as Primer 3 (Rozen and Skaletsky 1998), and visual selection.

PCR amplification and evaluation of polymorphism

To estimate the degree of polymorphism, the previously mentioned 16 inbred lines were evaluated for each microsatellite after gel electrophoresis of the amplified sequences using standard denaturing polyacrylamide sequencing gels. DNA was prepared as described above, starting from 1 g of leaf tissue belonging to a pool of plants from the same inbred line. Thus, in those few cases where 2 microsatellite bands (alleles)/locus were detected, we cannot determine if it was due to heterogeneity of the inbred line or heterozygosity of any of the individual plants. PCR reactions were performed in a total volume of 12 μL using a PT-100 DNA thermocycler (MJ Research Inc., Watertown, Mass.). The reaction mixture contained 30 ng of genomic DNA, 1.5 mM $MgCl_2$, 0.2 mM of each deoxyribonucleotide, 0.25 μM each of 3′- and 5′-end primers, 1× PCR buffer (Promega), and 1 U Taq DNA polymerase (Promega). The amplification reaction was performed by enrichment of template by a touch down of 70–55°C or 68–50°C according to the annealing temperatures of the pair of primers tested. The enrichment step was followed by 25 cycles, each consisting of a 94°C denaturing step (60 s), the lower annealing temperature of the pair of primers (60 s), and a 72°C elongation step (60 s). After the PCR reaction, the amplification products were resolved by electrophoresis on denaturing polyacrylamide gels (6% acrylamide/bisacrylamide, 20:1, 8 M urea in TBE (pH 8.3)). Bands were revealed using a sequencing gel silver-staining kit (Promega).

Data collection and analysis

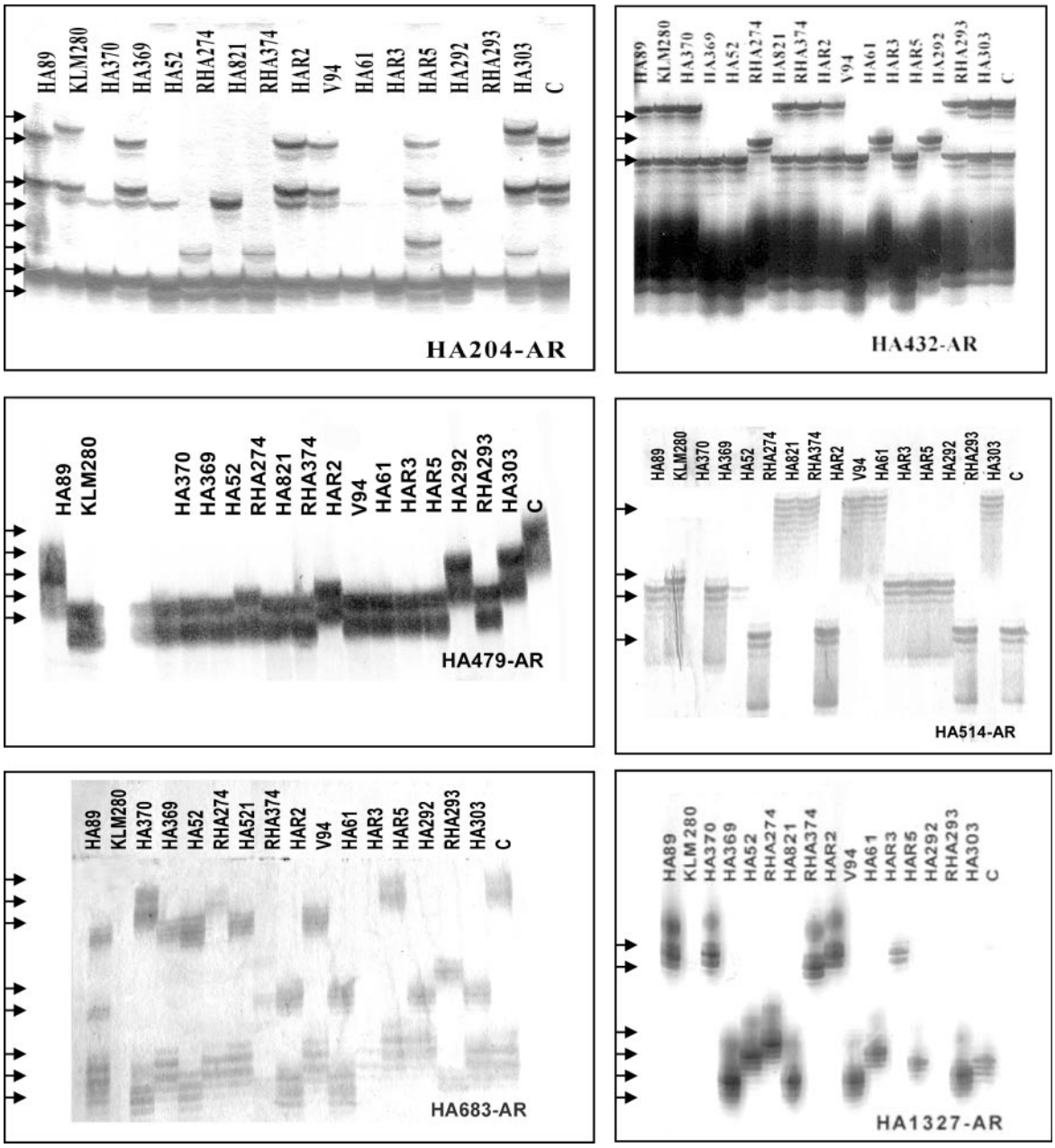
Genetic diversity was measured by evaluating the differences in allele number per locus, polymorphic index content, and similarity values. The amplification profile for each microsatellite was scored visually and independently twice. Ambiguous data were re-examined in a new electrophoresis run.

The degree of polymorphism was quantified using the polymorphic information content (PIC, Anderson et al. 1993), also referred to as "expected heterozygosity" by Powell et al. (1996) for soybean or "diversity index" by Milbourne et al. (1997) for potato

$$\mathrm{PIC} = 1 - \sum_{i}^{n} p_g^2$$

where p_g is the frequency of an individual genotype. Bands with the same mobility were considered identical, receiving equal values. SSR markers were usually considered to reveal a single locus per primer combination. The presence of only

Fig. 1. Characterization of different microsatellite-containing repeats by silver-stained denaturing PAGE of PCR amplification products.



one allele of a given microsatellite was considered a homozygous state of the allele, assuming the absence of null alleles.

Presence or absence of each band was coded as 1 or 0, respectively, in a binary data matrix (BDM). In the case of primers revealing more than one band (owing to the presence of heterozygous plants or heterogeneity between plants belonging to the same inbred line; Fig. 1, *Ha432-ar*), both alleles were scored. No attempts to calculate and include relative frequencies of the different microsatellite alleles within each inbred line were carried out. In the case of primers revealing more than one locus at the same time, only one of the two loci was taken into account (Fig. 1, *Ha204-ar* and *Ha683-ar*). This was done to avoid scoring two different variables that are probably not independent (theoretically these loci are frequently closely linked from the genetic point of view, thus describing the same genomic region; see Discussion).

Similarity analyses were conducted using a complete BDM comprising data from the analysis of 170 microsatellites and estimated with the NTSYS 1.8 program (Exeter Software, Setauket, N.Y.; Rohlf 1992) using the Jaccard association coefficient of similarity (Sneath and Sokal 1973). The resulting pairwise similarities were expressed as triangular similarity matrices. Cluster analysis were conducted by applying the SAHN option on similarity estimates using the unweighted pair-group method arithmetic average (UPGMA) and the resulting clusters were expressed as dendrograms using the option TREE PLOT. To estimate the magnitude of distortion of the dendrogram, the

Table 2. Characterization of microsatellite sequences.

Motifs	Polymorphic microsatellites						
	Perfect			Imperfect and compound			
	Average TRN	Average PIC	Subtotal	Average TRN	Average PIC	Subtotal	Total
GA	15±4	0.5±0.2	34				
GT	9±3	0.4±0.2	28	23±13	0.4±0.2	39	112
TGG	6±1	0.3±0.2	11				
ATT	26±10	0.6±0.1	9				
ATC	8±5	0.4±0.2	8	25±12	0.5±0.1	36	56
CATA	10±8	0.5±0.4	3				

cophenetic matrix was computed and correlated with the similarity matrix using the MATRIX PLOT function.

A second BDM was constructed including data from a selected group of 20 highly polymorphic microsatellites to evaluate the fingerprinting potential of these markers.

Results

Library construction and microsatellite screening

Sunflower genomic DNA was cut with several restriction-endonuclease combinations. Among them, *Tsp*509I was selected based on the size of the fragments obtained and the compatibility of the digested DNA sticky ends with the *Eco*RI-produced sticky ends in the vector DNA. Thus, *Tsp*509I-digested *H. annuus* DNA was cloned in the *Eco*RI site of λ-Zap II vector with an efficiency of about 5×10^5 PFU/mL. Assuming an average insert size of 250 bp, the library covers about 1.25×10^8 bp, representing 4% of *H. annuus* genome. This library was later amplified to reach a total of 4.4×10^8 PFU/mL. Approximately $5.3 \times 10^5 \pm 0.5 \times 10^5$ PFU were screened using ^{32}P-labeled oligonucleotide probes containing $(AG)_n$, $(GT)_n$, and $(TGG)_n$, and $3.2 \times 10^5 \pm 0.7 \times 10^5$ PFU were screened with probes containing $(ATC)_n$, $(ATT)_n$, $(TCT)_n$, $(GAG)_n$, $(CAG)_n$, $(CATA)_n$, $(GATA)_n$, and $(GGAT)_n$. A total number of 1856 putative positive clones were identified. After a second round of plaque hybridization, about 35% of these clones were not confirmed. The remaining positive clones were isolated and sequenced. To identify redundancy, the raw sequence data were trimmed by removing 5′ and 3′ vector sequences and comparing them with the aid of a computer algorithm against a cumulative sequence database, masking out those regions containing microsatellite repeats. About 50% of the remaining sequenced clones were discarded by this procedure, returning 608 true, non-redundant, potentially useful sequences. The frequencies observed per motif in this library were $3.7 \times 10^{-4} \pm 0.5 \times 10^{-4}$ for $(GA)_n$, $2.5 \times 10^{-4} \pm 0.3 \times 10^{-4}$ for $(GT)_n$, $8.9 \times 10^{-5} \pm 1.1 \times 10^{-5}$ for $(TGG)_n$, $1.5 \times 10^{-4} \pm 0.4 \times 10^{-4}$ for $(ATC)_n$, $1.4 \times 10^{-4} \pm 0.4 \times 10^{-4}$ for $(ATT)_n$, and $8.0 \times 10^{-5} \pm 2.3 \times 10^{-5}$ for $(CATA)_n$. The isolation of clones containing microsatellite motifs like $(CGG)_n$, $(TTG)_n$, $(TCT)_n$, $(GAG)_n$, $(GATA)_n$, and $(GGAT)_n$ was due to their circumstantial association with any of the other motifs described above.

Primer design and evaluation

Primer pairs flanking the microsatellite motifs could be designed for 54% of the remaining 503 useful sequences. In the rest of the clones, too little flanking sequence was left for primer design either on one or both sides of the microsatellite motif. Forty-two percent of these primer sets were derived from clones containing a GA motif, whereas 31% of the primers were derived from the GT motif, 9% of the primers were derived from the TGG motif, 9% of the primers were derived from the ATT motif, 5% of the primers were derived from the ATC motif, and 3% of the clones were derived from the CATA motif.

Primer pairs were tested using the inbred line set, under identical PCR conditions and two thermal regimes according to the annealing temperature of each set. Of the primer sets tested, 216 (79% of the total) produced PCR products of the expected size, whereas the rest either failed to amplify or produced a complex pattern of bands that was difficult to evaluate. Of these 216 microsatellites, 170 produced clear and reproducible banding patterns that showed polymorphisms among sunflower lines on silver-stained polyacrylamide electrophoresis gels (Fig. 1).

Polymorphic microsatellites description

The most frequent motifs found among the 170 polymorphic SSRs characterized in this work were: $(GA)_n$ (38%) and $(GT)_n$ (37%), followed by trinucleotide motifs $(ATT)_n$ (7.1%), $(TGG)_n$ (6.5%), and $(ATC)_n$ (4%), and tetranucleotide motifs $(CATA)_n$ (4%). These microsatellite sequences could be classified as perfect, imperfect, and compound according to Weber (1990). Fifty-seven percent of the polymorphic microsatellites characterized in this work, comprising the six analyzed microsatellite types, were classified as perfect microsatellites (Table 2).

The imperfect and compound microsatellite types represent 43% of the polymorphic microsatellites studied. Compound microsatellites were the most common of the two types of complex repeats (Table 2). The association among dinucleotide repeats was commonly found in the microsatellite structure of the clones analyzed in this work. The most frequent association was $(GA)_n$–$(GT)_n$, followed by $(GT)_n$–$(AT)_n$, and less frequently the three dinucleotide repeats were found together. For both $(GA)_n$ and $(GT)_n$ microsatellites, other associated repeats were trinucleotide or tetranucleotide motifs. Perfect and compound $(GT)_n$ microsatellites were often flanked by long stretches of poly(C). Imperfect microsatellites, defined as those containing two or more strings of uninterrupted repeats broken up by a stretch of nonrepeated bases, were more frequent in the $(GA)_n$ motif. Usually, imperfect microsatellite motifs are interrupted by repeat-like sequences mainly composed of the

Table 3. PCR primer pairs amplifying microsatellites from cultivated sunflower. The last column refers of alleles found in 16 sunflower accessions.

Locus	GenBank accession No.	Motif	Primer sequences (5′→3′) Forward	Reverse	Expected size range (bp)	Allele No.
Ha95-ar	G67405	$(GA)_{24}$	GAACATGGCCATAACTCATAGACG	CCTTCGACCCAACATC	166–162	5
Ha140-ar	G67517	$(GA)_{13}$	GTCTCCTTCTCTTTCTCGGC	CTAGCAACCAACCTCATTG	149–129	3
Ha196-ar	G67518	$(GA)_{13}\ G_{16}$	GGTGTTAGAGTAAAGTATGCCG	AGGGGTTTCCTTTCAG	182–176	4
Ha293-ar	G67519	$(GA)_7\ N_4\ (GA)_3\ TA\ (GA)_4\ N_5\ (GA)_{12}$	GGGGACATCTCCCGTCCACC	CCTCATCCATCTCCATCCAATC	111–107	3
Ha357-ar	G67522	$(GA)_{15}$	GTGGGTGTGGAAGGAAGAATC	CAGACACATGCTAGTCGTCGTG	132–92	3
Ha360-ar	G67406	$(GA)_{15}$	CAACAAGGAACCGATAACTGCT	CACCCTTCATCTCCTTC	199–191	4
Ha432-ar	G67407	$(GT)_{10}$	CTTTATCCCCCACCCCCTCC	GGGTTTAGTGGCCAGTAGTTGTC	178–168	3
Ha494-ar	G67408	$(GA)_{17}\ A\ (GA)_2\ N_{12}\ T_{15}$	GCGTTGGTTAAGGCCTGAGGTC	GAGCAGCAAACAGAGGGTACACC	209–193	4
Ha514-ar	G67409	$(GA)_{13}$	GGTCAACGGATTTAGAGTC	GTATTGATTCCAACATCCAG	200–164	4
Ha806-ar	G67410	$(GT)_8\ N_{27}\ (GA)_6$	GATGTTCCTTCCTGCAC	GGTTGGATAAATGGGGCAGC	199–189	4
Ha991-ar	G67520	$(GA)_4\ T\ (GA)_{12}$	GCCCCCTTGATGCCCTTTTC	GAATCGCCATTTGAATCGCCAG	145–137	3
Ha1167-ar	G67411	$(GT)_9\ N_2\ (GT)_4$	CGATGTCGCGATTCGGACTGGAG	CCCCATCTACACTTCAATACTG	165–155	3
Ha1209-ar	G67412	$(ATT)_{26}$	CATTGGCCTCACAAACACTTG	GATGTGAAACAGCCTCCATACTC	160–142	5
Ha1287-ar	G67413	$(GA)_{26}$	GATATGAGCCCATCACTCATC	GAAGATATGTCAGGTCACACCC	171–151	8
Ha1327-ar	G67414	$(ATT)_{30}$	CCGTTAGGTATGTTACTTGCGACC	GGTGGGGGGAATATTCTGAGGTG	225–201	6
Ha1354-ar	G67415	$(AT)_{16}\ AAT\ N_7\ (GA)_4\ T_2\ (GA)_{11}$	CATATGTGGCACATTCATGG	GGTTGATTGGGGTTAAGGAG	212–182	8
Ha1442-ar	G67521	$(ATT)_{31}$	GCTTATGTGCTTACGTGTTCCTG	CTAAACAGTTCGGCGAGTGTAGG	244–202	8
Ha1608-ar	G67416	$(ATT)_{25}$	GATCTTAGGTCCGCCAC	GATGGCATTTGGCTAGAC	331–169	7
Ha1626-ar	G67417	$(ATT)_{35}$	GATGTTACACGTTAGCAACG	GAACTCAGCCTAAAAGTC	170–146	6
Ha1796-ar	G67418	$(ATT)_{33}$	CGAAGGAAGGAACCTGCCTC	CCATACGCGTTTACTTCTCAGG	230–152	7

same bases that constitute the basic motif of the microsatellite.

A wide range of repeat variation was observed among members of these groups; the average tandem repeat number (TRN) and the average PIC found per motif are presented in Table 2.

Microsatellite polymorphism among sunflower inbred lines

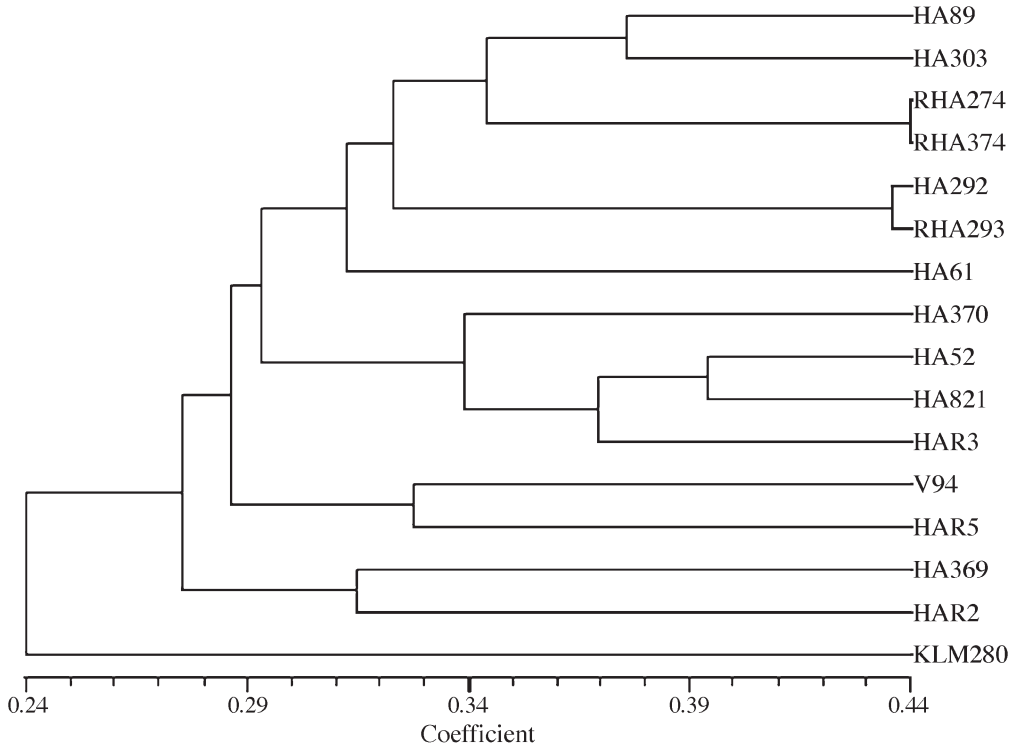
The 170 polymorphic microsatellites produced an average of 3.5 alleles/locus and an average PIC of 0.55 within the reference group of 16 inbred lines. Analyzing the average PIC of the different motifs (Table 2), it can be inferred that there is a subtle association between the number of repeat units and the polymorphism degree, and that those characteristics can also be correlated with the presence of the AT-rich motifs in the microsatellite structure. Furthermore, it was observed that microsatellites containing trinucleotide motifs, particularly $(ATT)_n$, exhibited the largest differences in allele size.

In relation to the quality exhibited by microsatellite amplification profiles, no association between stuttering (multiple bands in a microsatellite allele) and specific dinucleotide motifs was found. However, it was observed that amplification products of $(ATT)_n$, $(TGG)_n$, and $(GA)_n$ motifs usually produced few stutter bands, whereas $(GT)_n$, $(CATA)_n$, and $(ATC)_n$ motifs exhibited a lower quality profile, but still produced scorable bands.

Finally, some primer combinations evidenced the amplification of multiple band patterns (see Fig. 1, *Ha204-ar*, *Ha432-ar*, and *Ha683-ar*). This kind of amplification profile can be explained by at least two main reasons: (*i*) heterozygosity or heterogeneity of the inbred lines, indicating that they are derived from heterozygous cultivar selections; and (*ii*) presence of two (or more) duplicated microsatellite loci (including their bordering sequences) in the sunflower genome. The number of microsatellites evidencing heterozygosity and (or) heterogeneity was 18 (out of 170), suggesting a high degree of inbreeding in the analyzed lines. Six microsatellites that contain either imperfect $(GA)_n$ or compound $(GA)_n(GT)_n$ or $(GT)_n(TA)_n$ motifs, look like they belong to duplicated genomic loci (Fig. 1, *Ha204-ar* and *Ha683-ar*). This kind of amplification pattern was described before by Akagi et al. (1998) in different species of the genus *Oryza*.

It is generally assumed that commercial sunflower genetic diversity is limited because of its extensive breeding as well as the fact that the crop was introduced in most of the important producing areas world wide (Eastern and Western Europe, Argentina, South Africa, India; sunflower center of origin is located in North America). Limited genetic variability in sunflower sometimes prevents the unambiguous identification (fingerprinting) of new cultivars or hybrids with the traditional morphological descriptors used for variety protection. To explore the use of the characterized microsatellites as a tool for the elaboration of a unique genotype identification or fingerprint for sunflower hybrids, inbred lines, and varieties, a selected group of 20 microsatellites was analyzed. The potential number of unique genotypes generated using just these 20 specific markers is 565 740 according to the formula provided by Brown et al. (1996). This formula considers the number of

Fig. 2. Dendrogram based on microsatellites. Sixteen sunflower accessions were analyzed with 170 primer pairs using NTSYS. Similarity matrix was calculated to cluster the data using the unweighted pair-group method with arithmetical average (UPGMA) algorithm.



alleles as well as their relative frequency in the studied population (PIC) pointing out the maximum number of unique combinations of microsatellite alleles that be can potentially found in a given genotype. Therefore, with a rather limited number of loci (20) it should be tentatively possible to unequivocally fingerprint most of the existing commercial germplasm. The sequences of the primers and amplified microsatellite motifs of these selected markers are listed in Table 3.

Another point of interest was to evaluate the use of these microsatellite markers for the study of genetic relationships within *H. annuus* sp. (the subspecies of cultivated sunflower). A similarity matrix was calculated to cluster the data using the unweighted pair-group method with arithmetical average (UPGMA) algorithm. The average similarity coefficient among cultivars was 0.36, ranging from 0.44 for related pairs of lines (RHA 274 and RHA374) to 0.24 for KLM280 vs. HA89 pair of lines. Regression between cophenetic and similarity matrices was high and significant ($r = 0.80$, $P < 0.05$), reflecting a good fit. Genetically related inbred lines clustered together at values of 0.44 or lower (Fig. 2), suggesting that the chosen inbred lines are quite different between them, supporting the criterion by which they were selected for polymorphism screening. The most diverse genotypes are those of Argentine origin (KLM280, HAR2, HA369, HAR5, V94, HAR3), which separate from the rest at similarity values of 0.29. This can be easily explained by the intensive utilization of wide interspecific crossing with *Helianthus petiolaris*, *Helianthus argophyllus*, and *H. annuus* subsp. *annus* in their breeding process, as well as the use of striped-hull achene Russian sunflower germplasm (instead of sunflowers with black-hull achenes used in American, South African, or Western European breeding programs). The exception is Argentine inbred line 'HAR3' derived from Russian variety 'Peredovik 301', which is also an ancestor of inbred lines 'HA821' and 'H52'. Thus, it is not surprising that this line clusters together with American and South African inbred lines and not with the other Argentine lines analyzed here (Fig. 2). The two representatives of the nonoilseed-type sunflower group ('HA292' and 'HA293') cluster together with a similarity index of 0.43, coherent with their breeding history (i.e., separated from the oilseed-type sunflowers). 'RHA274' and 'RHA374' are the closest related lines analyzed here (although the similarity coefficient is as low as 0.44). Both of them are restorer lines for oilseed-type sunflowers. This cluster relates to the one above because they share the cytoplasmic male-sterility donor 'cmsHA155'. At a similarity coefficient of 0.39, 'HA89' and 'HA303' group together. Both lines are derived from traditional Russian varieties of the black-hull achene type (associated with larger oil content) that also participate in the genealogy of some of the inbred lines described above, which explains why they group with the previous mentioned genotypes, as well as with 'HA61', at a similarity coefficient of 0.32.

Discussion

Primary screening by hybridization of a non-enriched, amplified sunflower *Tsp*509I genomic library with different di-, tri-, and tetra-nuceotide repeats resulted in the isolation of 503 unique microsatellite-containing clones from 1200 hybridization-positive clones (42%). Primer pairs were designed for 54% of these unique clones; about 78% of

these primers allowed the amplification of clear PCR product bands, whereas 79% of the amplified PCR products showed polymorphisms among 16 sunflower elite inbred lines. The usefulness described for this library was comparable to those reported for other nonenriched, as well as for enriched, microsatellite libraries. For instance, Chen et al. (1997), working with rice *Tsp*509I genomic library, identified 587 putative positive clones. About 24% of these clones yielded unique microsatellite sequences. For 64% of them primer sets could be designed, 73% allowed the amplification of clear PCR products, and 94% of the amplified PCR bands were polymorphic among the parental genotypes of a mapping population. In contrast, Yu et al. (2000) sequenced 980 clones from a sunflower genomic DNA library enriched for several dinucleotide and tetranucleotide microsatellite motifs. Of these, 35.5% were unique sequences. Primer sets were designed for 87% of the single clones, 74.5% of primer sets amplified clean SSRs for genomic DNA, and 71% of the amplified products evidenced polymorphisms among 16 elite inbred lines.

Inferences of microsatellite abundance in sunflower genome based on this data could not be carried out owing to the high level of biased redundancy introduced by enrichment of the library. Nevertheless, the frequency per motif obtained for the *Tsp*509I library, gave an approximation of microsatellite representation in sunflower genome, and guided the current screening of non-amplified libraries in our lab. Besides, the predominance of GA motifs over GT motifs inferred here agrees with previously described results for sunflower by Dehemer and Friedt (1998) on the basis of nonradioactive probe hybridizations. This profile for dinucleotide motifs was widely described in plant genomes (Morgante and Olivieri 1993; Wang et al. 1994; Panaud et al. 1995; Echt et al. 1996; Smulders et al. 1997; Hüttel et al. 1999). Another aspect observed in this work was the abundance of AT-rich microsatellites found, preferably in association with $(GT)_n$ and $(GA)_n$ motifs. This kind of association was also observed in grasses (Röder et al. 1995; Taramino and Tingley 1996). Regarding trinucleotide motifs, $(ATT)_n$ was the most encountered, followed by ATC and $(TGG)_n$. This profile was different from that reported by Dehemer and Friedt (1998) and not described in sunflower or lettuce previously, which could be explained by the screening and (or) enrichment strategy followed in those works. Our results agree again with those reported by the same authors in relation to the predominance of $(CATA)_n$ among tetranucleotide microsatellite repeats. Taken together, the representation of microsatellite motifs in sunflower genome seem to follow the general pattern described within plant genomes, where GA-, GT-, and AT-rich motifs compose the most frequent type of microsatellites.

The amplification rate of the 271 microsatellites isolated in this work was high (78%), and comparable to other rates obtained for primers derived from restriction-enzyme-digested DNA library clones (Chen et al. 1997; Panaud et al. 1996), as well as for primers derived from enriched DNA library clones (Yu et al. 2000; Van del Wiel et al. 1999).

The amplification of dinucleotide microsatellites did not produce stutter bands as has been frequently reported in other species (Cregan et al. 1994). However, differences were observed in the amplification quality of some $(GT)_n$-containing microsatellites. This fact was also described for maize and wheat (Taramino and Tingley 1996; Ma et al. 1996), where it was hypothesized that the poor quality observed during $(GT)_n$-containing microsatellite amplifications could be a consequence of the high complexity exhibited by this type of microsatellites.

Even with the rather selective group of breeding lines used for polymorphism screening, and considering the narrow genetic background of cultivated sunflower, 79% of the microsatellites analyzed evidenced polymorphism within the sixteen sunflower inbred lines used for characterization. The average PIC value of 0.50 and the average of 3.5 alleles/locus are comparable with the values found for lettuce microsatellites in 18 genotypes, even though 7 of the last belong to a wider germplasm representation that include accessions not normally used in commercial breeding (van de Wiel et al. 1999).

As happens in other living species, the most frequent polymorphic microsatellites belong to the dinucleotide class of repeats (see Table 2). Regarding microsatellites containing trinucleotide repeats, those containing AT-rich motifs exhibited the maximum level of polymorphism as well as the widest allele size range observed among the studied motifs. These characteristics, together with the fact that they mostly amplify low copy number loci and produce PCR products almost free of band artifacts, reinforce the interest in trinucleotide microsatellites as molecular markers for mapping and fingerprinting purposes.

A subgroup of polymorphic microsatellites produced the amplification of multiple-band patterns. This happens, mostly, only with some of the accessions and can be easily attributed to heterozygosity or heterogeneity of the inbred lines used for screening (see Fig. 1, *Ha432-ar*, for one example). These double bands could correspond to two microsatellite alleles belonging to the same locus, reflecting that the segregation of the loci present in those inbred lines descend from ancestors that are not 100% homozygous. In other cases, a double-band pattern was observed in all the genotypes (Fig. 1, *Ha204-ar* and *Ha683-ar*). This type of amplification patterns has been explained in other species by the occurrence of a duplication process within the genome and evolution of families of repetitive sequences (Akagi et al. 1998). If this were the case, the usefulness of multilocus microsatellites for molecular marker-assisted selection, for fingerprinting, or for calculation of genetic relationships would be the same as those microsatellites revealing just a single locus, because the two (or more) loci are genetically linked, thus representing the same genomic region.

Microsatellite patterns have been shown to represent genetic relationships quite accurately in many species and taxa. High gene diversity scores shown by microsatellite markers make them useful in distinguishing closely related genotypes (Giancola 1998). Indeed, the Plant Variety Protection Office of the USDA Agriculture Marketing Service now accepts microsatellite allelic profiles as supporting evidence for the uniqueness of a new cultivar (Diwan 1997). Here, 20 microsatellite markers were enough to establish a unique and specific fingerprint for the 16 inbred lines, as well as for many different commercial hybrids tested in our lab (data not shown), confirming the power of these microsatellite markers to reveal genetic diversity in sunflower germplasm.

One of the controversial points in the application of molecular markers to variety registration is the predictable reduction of the minimal genetic distances that will distinguish two varieties and establish the limits of the so called "essentially derived varieties." To be defined as such, varieties should ideally differ in a considerable proportion of their genomes, not just in a few loci. The careful selection of molecular markers widely located in the genetic map should take care of this point. This work shows that, for the case of breeding sunflower germplasm the calculated maximum genetic similarity is about 0.44, suggesting the feasibility of this approach (Fig. 2, RHA274 vs. RHA374). The best system to differentiate and identify varieties must be practical, precise, reliable, and robust. It must offer high discrimination power, lack of interaction with the environment, capacity to generate equivalent results among laboratories, consistency of calculated genetic relationships with pedigree data, good genomic coverage, public availability, susceptibility to automation, and adequate cost–profit relationship. According to our results, the use of SSRs shows the closest possibility to generate such a system to differentiate and identify commercial varieties of sunflower. This work demonstrates that SSRs provide a powerful way of discriminating between lines and that they appear to be useful for identification, assessment of distinctness, and diversity.

In conclusion, this work achieved the isolation and characterization of a significant number of sunflower microsatellites. This allowed us to get an insight of their relative abundance, organization, and polymorphism. The primer sets designed in this work for sunflower showed to be useful markers for breeding program purposes and identification of sunflower inbred lines, cultivars, and accessions that was the aim of this work.

Acknowledgements

We are grateful to Dr. Susan McCouch and Dr. Xiuli Chen for their excellent scientific and technical assistance during the first part of this work. We thank C.C. Julio Tillería for the development of the computationals algorithms needed to curate the sunflower microsatellite database and Nancy Lopez for her technical assistance. This research was supported in part by a consortium of eight seed companies (Advanta Semillas SAIC, Buck, Cargill, Dekalb, Mycoyen, Nidera, Novartis, Sursem) to which the authors are very grateful. This project was also partially supported by PID/98:024 from FONCyT (BID 802 OC-AR, ANPCyT-SETCyP, Argentina). Dr. N. Paniego and Dr. R. Zandomeni are career members of the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) and Dr. H.E. Hopp is a career member of the Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) and Professor at the Facultad de Ciencias Exactas y Naturales, University of Buenos Aires (UBA).

References

Akagi, H., Yokzeki, Y., Inagaki, A., and Fujimura, T. 1998. Origin and evolution of twin microsatellites in the genus *Oryza*. Heredity, **81**: 187–197.

Akkaya, S.M., 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., Autrique, J.E., Tanksley, S.D., and Sorrells, M.E. 1993. Optimizing parental selection for genetic linkage maps. Genome, **36**: 181–186.

Arias, D.M., and Rieseberg, L.H. 1995. Genetic relationship among domesticated and wild sunflower (*Hellianthus annuus*, Asteraceae). Econ. Bot. **49**: 239–248.

Ashikawa, I., Kurata, N., Saji, S., Umehara, Y., and Sasaki, T. 1999. Application of restriction fragment fingerprinting with a rice microsatellite sequence to assembling rice YAC clones. Genome, **42**: 330–7.

Bell, C.J., and Ecker, J.H. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics, **19**: 137–144.

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

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.

Brown, S.M., Hopkins, M.S., Mitchell, S.E., Senior, M.L., Wang, T.Y., Duncan, R.R., Gonzalez-Candelas, F., and Kresovich, S. 1996. Multiple methods for the identification of polymorphic simple sequence repeat, (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor. Appl. Genet. **93**: 190–198.

Chen, X., Temnykh, 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. 1998. Ancestral origins and genetic diversity of cultivated sunflower: coancestry analysis of public germplasm. Crop. Sci. **38**: 1476–1482.

Cregan, P.B., Bhagwat, A.A., Akkaya, M.S., and Rongwen, J. 1994. Microsatellite fingerprinting and mapping of soybean. Methods Mol. Cell. Biol. **5**: 49–61.

Dehemer, K.J., and Friedt, W. 1998. Evaluation of different microsatellite motifs for analyzing genetic relationships in cultivated sunflower (*Helianthus annuus* L.). Plant Breed. **1**: 39–42.

Diwan, N., and Cregan, P.B. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. **95**: 723–733.

Echt, C.S., May-Marquardt, P., Hseih, M., and Zahorchak, R. 1996. Characterization of microsatellite markers in eastern white pine. Genome, **39**: 1102–1108.

Gedil, M.A., Wye, C., Berry, S., Segers, B., Peleman, J., Jones, R., Leon, A., Slabaugh, M.B., and Knapp, S.J. 2001. 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., Berville, A., and Nicolas, P. 1995. Development of a consensus linkage RFLP map for cultivated sunflower. Theor. Appl. Genet. **90**: 1079–1086.

Gentzbittel, L., Mestries, E., Mouzeyrat, F., Badaoqui, S., Vear, F., Tourvieille de Labrouhe, D., and Nicolas, P. 1999. A composite map of expressed sequences and phenotypic traits of the sunflower (*Helliantus annuus* L.) genome. Theor. Appl. Genet. **99**: 218–234.

Giancola, S.M. 1998. Caracterización y diferenciación de germoplasma de soja (Glycyne max L. Merr.) mediante marcadores moleculares. PhD. Thesis. Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.

Goldstein, D.B., and Pollock, D.D. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. J. Hered. **88**: 335–42.

Goldstein, D.B., Roemer, G.W., Smith, D.A., Reich, D.E., Bergman, A., and Wayne, R.K. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics, **151**: 797–801.

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

Hüttel, B., Winter, P., Weising, K., Choumane, W., Weigand, F., and Kahl, G. 1999. Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Genome, **42**: 210–217.

Jan, C.C., Vick, B.A., Miller, J.K., Kahler, A.L., and Butler, E.T.I. 1998. Construction of a RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. **96**: 15–22.

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

Ma, Z.Q., Röder, M., and Sorrells, M.E. 1996. Frequencies and sequence characteristics of di-, tri-, and tetra-nucleotide microsatellites in wheat. Genome, **39**: 132–130.

Marino, M.A., Turni, L.A., Del Rio, S.A., Williams, P.E., and Cregan, P.B. 1995. The analysis of simple sequence repeat DNA in soybean by capillary gel electrophoresis. Appl. Theor. Electrophor. **5**: 1–5.

Maughan, P.J., Saghai Maroof, M.A., and Buss, R.G. 1995. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome, **38**: 715–723.

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.

Milbourne, D., Meyer, R., Bradshow, J.E., Baird, E., Bonar, N., Powel, J.W., Morgante, M., Andre, C., Hanafey, M., Vogel, M., Tingey, S., and Rafalski, A. 1997. The utility of RFLP, RAPID, AFLP and SSR (microsatellite) markers for germoplasm analysis. Mol. Breed. **3**: 225–238.

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

Panaud, O., Chen, X., and McCouch, S.R. 1995. Frequency of microsatellite sequences in rice (*Oryza sativa* L.). Genome, **38**: 1170–1176.

Panaud, O., Chen, X., and McCouch, S.R. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol. Gen. Genet. **252**: 597–607.

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

Röder, M.S., Plaschke, J., Konig, S.U., Borner, A., Sorrells, M.E., Tanksley, S.D., and Ganal, M.W. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. Mol. Gen. Genet. **246**: 327–33.

Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.H., Leroy, P., and Ganal, M.W. 1998. A microsatellite map of wheat. Genetics, **149**: 2007–2023.

Rohlf, F.J. 1992. NTSYS-pc. Numerical taxonomy and multivariate analysis system version 1.7. Owner's manual. Exeter Publishers Ltd., Setauket, N.Y.

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.

Rozen, S., and Skaletsky, H.J. 1998. Primer3. Available from http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., and Allard, R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U.S.A. **81**: 8014–8018.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular cloning: a laboratory manual. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Scotti, I., Magni, F., Fink, R., Powell, W., Binelli, G., and Hedley, P.E. 2000. Microsatellite repeats are not randomly distributed within Norway spruce (*Picea abies* K.) expressed sequences. Genome, **43**: 41–6.

Senior, M.L., and Heun, N. 1993. Mapping maize microsatellites and polymerase chain reaction confirmation of the target repeats using a CT primer. Genome, **36**: 884–889.

Smulders, M.J.M., Bredmeijer, G., Rus-Kortekass, W., Arens, P., and Vosman, B. 1997. Use of short microsatellites data-base sequences to generate polymorphisms among *Lycopersicun esculentum* cultivars and accessions of other *Lycopersicun* species. Theor. Appl. Genet. **97**: 264–272.

Sneath, P.H.A., and Sokal, R.R. 1973. Numerical taxonomy: the principles and practice of numerical classification. W.H. Freeman, San Fransisco, Calif.

Taramino, G., and Tingley, S. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. Genome, **39**: 277–287.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. **17**: 6463–6471.

Thomas, M.R., and Scott, N.S. 1993. Micorsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). Theor. Appl. Genet. **86**: 986–990.

van de Wiel, C., Arens, P., and Vosman, B. 1999. Microsatellite retrieval in lettuce (*Lactuca sativa* L.). Genome, **42**: 139–149.

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.

Weber, J.L. 1990. Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphism. Genomics, **7**: 524–530.

Weising, K., Atkinson, R.G., and Gardner, R.C. 1995. Genomic fingerprinting by microsatellite-primed PCR: a critical evaluation. PCR Methods Appl. **4**: 249–255.

Weising, K., Fung, R.W.M., Keeling, D.J., Atkinson, R.G., and Gardner, R.C. 1996. Characterization of microsatellites from *Acnidia chinensis*. Mol. Breed. **2**: 117–131.

Winter, P., Pfaff, T., Udupa, S.M., Huttel, B., Sharma, P.C., Sahi, S., Arreguin-Espinoza, R., Weigand, F., Muehlbauer, F.J., and Kahl, G. 1999. Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. Mol. Gen. Genet. **262**: 90–101.

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

Yu, K., Park, S.J., and Poysa, V. 1999. Abundance and variation of microsatellite DNA sequences in beans (*Phaseoulus* and *Vigna*). Genome, **42**: 27–34.

Yu, J.-K., Mangor, J., Edwards, E., Thompson, L., and Knapp, S.J. 2000. Simple sequence repeats in sunflower: lengh polymorphisms among elite inbred lines for several motifs. *In* Abstracts of Plant and Animal Genome Conference VIII. 9–12 January, 2000, San Diego, Calif. Available at http://www.intl-pag.org/pag/8/abstracts/pag8735.html.