Electronic Journal of Biotechnology ISSN: 0717-3458
http://www.ejbiotechnology.info
DOI: 10.2225/vol13-issue6-fulltext-14

Genetic mapping of EST-SSRs, SSR and InDels to improve saturation of genomic regions in a previously developed sunflower map

Paola Talia[1] · Verónica Nishinakamasu[1] · Horacio Esteban Hopp[1,2] · Ruth Amelia Heinz[1,2] · Norma Paniego[1] ✉

1 Instituto de Biotecnología, CICVyA, CNIA, INTA Castelar, Buenos Aires, Argentina

2 Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

✉ Corresponding author: npaniego@cnia.inta.gov.ar
Received June 14, 2010 / Accepted September 22, 2010
Published online: November 15, 2010

Abstract In order to saturate a sunflower genetic map and facilitate marker-assisted selection (MAS) breeding for stress response, it is necessary to enhance map saturation with molecular markers localized in linkage groups associated to genomic regions involved in these traits. This work describes the identification and characterization of 1,134 simple sequence repeat (SSR) containing expressed sequence tags (ESTs) from unigenes available databases. Twelve of these functional markers as well as 41 public SSR markers were successfully localized in linkage groups, thus contributing to the saturation of specific regions on a reference genetic-linkage-map derived from recombinant inbred lines (RIL) mapping population from the cross between PAC2 x RHA266 lines. The enriched map includes 547 markers (231 SSR, 9 EST-SSR, 3 insertions/deletions (InDels) and 304 amplified fragment length polymorphisms (AFLPs) distributed in 17 linkage groups (LG), spanning genetic size to 1,942.3 cM and improving its mean density to 3.6 cM per locus. As consequence, no gaps longer than 13.2 cM remain uncovered throughout the entire map, which increases the feasibility of detecting genes or traits of agronomic importance in sunflower.

Keywords: EST-SSR, InDels, linkage map, SSR, sunflower

INTRODUCTION

Microsatellite markers have been widely used in genetic analysis of crop plants based on their ubiquitous presence in genomes and their genetic characteristics of being co-dominant, frequently multi-allelic and chromosome-specific. Although non-genic simple sequence repeats (SSRs) are more polymorphic than genic SSRs (Cho et al. 2000; Lee et al. 2004), the estimated frequency of genic SSRs is high (Morgante et al. 2002) and their location within putative candidate genes make them particularly interesting as functional markers. They can be adapted to high-throughput genotyping and thus, become suitable for the construction of high-density linkage maps, gene mapping and marker-assisted selection. However, their development is expensive, labor intensive and time consuming, especially if they derived from genomic libraries (Paniego et al. 2002; Varshney et al. 2002; Tang et al. 2003). Alternatively, SSRs can be identified by mining expressed sequence tags (EST) databases and used for SSR marker development (Morgante and Olivieri, 1993). These markers are commonly referred as EST-SSR and they have been applied to genetic studies in different plant species as *Eragrostis* (Cervigni et al. 2008) sugarcane (Cordeiro et al. 2001), wheat (Eujayl et al. 2002; Leigh et al. 2003; Yu et al. 2004; Zhang et al. 2005), barley (Thiel et al. 2003; Chabane et al. 2005), rye (Hackauf and Wehling, 2002), melon (Fernandez-Silva et al. 2008), quercus (Ueno et al. 2008), oil palm (Singh et al. 2008), tea (Sharma et al. 2009) and eucalyptus (Acuña el al. 2010) including sulflower (Kumpatla and Mukhopadhyay, 2005; Pashley et al. 2006; Heesacker et al. 2008).

Over recent years there has been an increase in the availability of ESTs for a wide range of plant species, including sunflower (Gentzbittel et al. 1999; Fernández et al. 2003; Heesacker et al. 2008)

enabling identification and genetic characterization of functional markers in sunflower (Kolkman et al. 2004; Kumpatla and Mukhopadhyay, 2005; Lai et al. 2005; Liu and Burke, 2006; Pashley et al. 2006; Kolkman et al. 2007; Fusari et al. 2008; Heesacker et al. 2008).

The usefulness of EST-SSR markers arises from their close linkage to potentially important genes, helping to identify candidate genes for quantitative trait loci (QTL). Moreover, these markers could also be of great assistance for comparative studies in related species due to higher heterologous conservation.

Over the past decades, several genetic linkage maps that differ in length and density were developed for cultivated sunflower, based on different molecular markers such as restriction fragment length polymorphism (RFLP) and/or random amplification of polymorphic DNA (RAPD) markers for the first reported maps (Berry et al. 1995; Gentzbittel et al. 1995; Berry et al. 1997; Jan et al. 1998; Rieseberg, 1998; Gentzbittel et al. 1999; Berry et al. 2003). Later, the addition of amplified fragment length polymorphisms (AFLPs) markers allowed further saturation of genetic linkage maps (Peerbolte and Peleman, 1996; Gedil et al. 2001; Al-Chaarani et al. 2004). More recently, the concomitant development of a large number of SSR markers and the automatization of the mapping procedures (Paniego et al. 2002; Tang et al. 2002; Tang et al. 2003) lead to the generation of SSR-anchorage linkage maps for different populations including those relevant for sunflower (Paniego et al. 2002; Tang et al. 2002; Tang et al. 2003; Yu et al. 2003; Kiani et al. 2007). Most of the SSR markers used in sunflower mapping are neutral (usually located in intergenic genomic regions), as they were developed from genomic libraries using microsatellite motives as hybridization probes (Paniego et al. 2002; Tang et al. 2002; Tang et al. 2003). In recent years, due to the rapid increase of sequence information, the generation of EST-SSRs and single nucleotide polymorphisms (SNPs) markers has become an attractive alternative to complement existing SSR marker collections (Pashley et al. 2006, Heesacker et al. 2008) allowing the inclusion of functional markers in genetic maps (Lai et al. 2005).

The purposes of this work were to mine the unigene sunflower database (*Helianthus annus* Gene Index, HAGI) developed for TIGR (now DFCI Gene Index Database) for the identification and characterization of novel EST-derived SSR in terms of frequency, type, and motifs of repetition, and their inclusion along with new genomic SSR and insertion/deletions (InDels) on a previously developed sunflower reference map (Kiani et al. 2007), exploring their colocalization with biotic and abiotic stress tolerance QTL.

MATERIALS AND METHODS

Plant material

Sunflower genomic DNAs extracted from PAC2, RHA266, HA89 and RHA801 inbred lines were used for amplification of EST-SSR and polymorphism detection considering that these lines constitute the parentals of two mapping populations used for characterization of stress response traits. A segregant population of 94 recombinant inbred lines (RILs), derived from the crossing of sunflower PAC2 and RHA266 (Roath et al. 1981; Gentzbittel et al. 1995) was used for linkage mapping.

EST-SSRs survey in the TIGR sunflower unigene database

In order to identify EST-SSR, the database of sunflower TIGR unigenes (HAGI), currently maintained and updated at the Dana-Farber Cancer Institute (http://compbio.dfci.harvard.edu/tgi/) was screened using the CUGI-SSR Server (Clemson University Genome Institute, USA) for the presence of repeat tandem sequences with di, tri, tetra and penta-nucleotide motives. The database used for this analysis included 36,743 unigenes (Release 4.0).

The conditions set to identify EST-SSR were the following: repetition length≥18 bases for the di -and trinucleotides and ≥20 bases for tetra and pentanucleotides. The output file was manually evaluated and those primers amplifying more than one simple sequence motif (compound microsatellites) or amplifying imperfect repetitions were ruled out of further analysis. A subset of 127 identified EST-SSR were selected and evaluated for polymorphism detection in four sunflower genotypes.

Table 1. BLASTX sequence similarities and GO classification of sunflower EST-SSRs included in the genetic map.

Sequence name	Putative function	Hit description	Hit ACC	E-value	GO/EC code	GO Terms	LG
TC18146	leucine zipper	gi|38503523|gb|AAR20445.2| putative leucine zipper protein [*Gossypium hirsutum*]	AAR20445	0	GO:0048529 GO:0009706 GO:0009535 GO:0015979 GO:0015995 GO:0055114 GO:0046914 EC:1.14.13.81	magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase activity; chloroplast inner membrane; chloroplast thylakoid membrane; photosynthesis; chlorophyll biosynthetic process; oxidation reduction; transition metal ion binding	10
TC18194	protein kinase	gi|224117232|ref|XP_002331754.1| predicted protein [*Populus trichocarpa*] >gi|222874451|gb|EEF11582.1| predicted protein [*Populus trichocarpa*]	XP_002331754	2,66e^{-036}	GO:0004672	protein kinase activity	3
TC21689	glycine-rich protein	gi|195648068|gb|ACG43502.1| glycine-rich protein 2b [*Zea mays]*	ACG43502	7,55e^{-006}			4
TC23602	CDPK-related protein kinase	gi|195616300|gb|ACG29980.1| CDPK-related protein kinase [*Zea ays*]	ACG29980	1,61e^{-023}	GO:0006468 GO:0005509 GO:0005524 GO:0004674 EC:2.7.11.0	protein amino acid phosphorylation; calcium ion binding; ATP binding; protein serine/threonine kinase activity	8
TC23762	zinc ring-type	gi|225449738|ref|XP_002270953.1| predicted: hypothetical protein [*Vitis vinifera*] >gi|157354027|emb|CAO46568.1| unnamed protein product [Vitis vinifera]	XP_002270953	2,37e^{-037}	GO:0005515 GO:0008270	protein binding; zinc ion binding	14
TC24992	hypothetical protein	gi|189242004|ref|XP_001807057.1| predicted: hypothetical protein, partial [*Tribolium castaneum*]	XP_001807057	5,72e^{-006}			8
TC25054	ferredoxin precursor	gi|37779195|gb|AAO42615.1| ferredoxin [*Helianthus annuus*]	AAO42615	4,02e^{-036}	GO:0005515 GO:0009055 GO:0051536 GO:0022900 GO:0006118	protein binding; electron carrier activity; iron-sulfur cluster binding; electron transport chain; electron transport	16
TC26323	photosystem II	gi|255543290|ref|XP_002512708.1| Photosystem II 11 kDa protein recursor, putative [*Ricinus communis]* gi|223548669|gb|EEF50160.1| Photosystem II 11 kDa protein precursor, putative [*Ricinus communis*]	XP_002512708	1,27e^{-034}	GO:0030095 GO:0009543	chloroplast photosystem II; chloroplast thylakoid lumen	17
TC26869	DNAJ heat shock n-terminal domain-containing protein	gi|211953609|gb|ACJ13966.1| DNAJ heat shock N-terminal domain-containing protein [*Helianthus annuus*]	ACJ13966	4,05e^{-035}			

Genomic SSR and identification of InDels

Two sets of SSR, the set of 50 SSR named 'ORS' (Tang et al. 2002; Tang et al. 2003) selected by its position in a composite map (Yu et al. 2003) and a set of 117 SSR named 'HA', previously developed in the lab and without genomic localization (Paniego et al. 2002) were evaluated for polymorphism in PAC2 and RHA266 parental lines.

DNA isolation and PCR amplification

Genomic DNA was isolated from lyophilized young leaves of the four parental inbred lines and 94 RILs using the CTAB method (Saghai-Maroof et al. 1984).

PCR was carried out in 12 µl (final volume) using a Mastercycler ep-gradient (Eppendorf, Germany). Each reaction was accomplished using 30 ng of genomic DNA in 1 x PCR buffer including 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.25 µM of each primer, 0.75 U of *Taq* DNA polymerase (Invitrogen, USA). All fragments were amplified using the following touch-down PCR profile: an initial denaturing step of 4 min at 95ºC; 13 cycles of 30 sec at 94ºC, 30 sec at 63ºC, and 1 min at 72ºC, annealing temperature decreasing to 55ºC by 1ºC per cycle; followed by 30 cycles of 30 sec at 94ºC, 30 sec at 50ºC, 1 min at 72ºC, followed by 10 min at 72ºC.

Amplified SSR and EST-SSRs fragments of parental lines were evaluated by conventional and automatic methods. Conventional method consisted on 6% denaturing polyacrylamide gels (AA:BIS = 19:1) and 7 M urea in 0.5 x TBE buffer. PCR fragment were mixed with equal volumes of loading buffer (formamide containing 0.8 mM EDTA and 5 x bromophenol blue and xylene cyanol), denatured at 95ºC for 5 min and cooled on ice. Samples were then loaded on pre-heated Gibco BRL Sequencing System (Life Technologies, USA), and run at 2,000 V for 3 up to 4 hrs, depending on the fragment length. After the run, the fragments were visualized by silver staining. For this aim the gel was fixed for 15 min in 10% acetic acid, rinsed in deionized water, stained for 30 min in 0.3% (w/v) silver nitrate, rinsed again in deionized cold water, and developed for approximately 15 min until the bands became visible. The gel was then fixed for 15 min in 10% acetic acid. Scoring was done by visual inspection.

Amplified polymorphic fragments from genomic SSR and EST-SSR were alternatively analyzed using fluorescent labelled PCR primers according to the procedures described by Tang et al. (2003) and Kiani et al. (2007). Polymerase chain reaction (PCR) fragments were resolved using electrophoresis through an ABI 3130xl DNA analyzer (Applied Biosystems, USA). Fragment sizing was done using the ROX 500 internal-lane standard (Applied Biosystems; ROX, 6-carboxy-x-rhodamine). GeneMapper 3.0 software (Applied Biosystems, USA) was used to score SSR alleles.

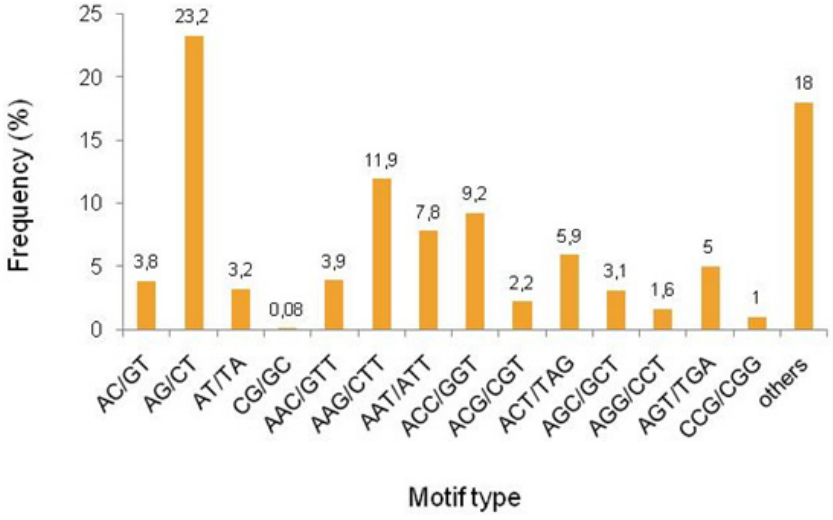


Fig. 1 Frequency distribution of EST-SSR based on motif size.

After evaluating allele profiles between the parental lines RHA266 and PAC2, useful polymorphisms were genotyped in the 94 RILs.

Three polymorphic InDels described by Fusari et al. (2008) between PAC2 and RHA266 (Fusari et al. 2008) were genotyped in the present work as well, using fluorescent labeled primers.

Gene Ontology (GO) annotations

Gene ontology (GO), (Ashburner et al. 2000) and Enzyme Commission (EC) annotations for the accessions shown in Table 1 were obtained using Blast2GO (Conesa and Götz, 2008). The BLASTX (Altschul et al. 1990) analysis against NCBI's NR Database was performed with an e-value threshold set at $1e^{-5}$. For the annotation step the e-value hit filter was $1e^{-6}$ and the annotation cut-off was 55 (Table 1).

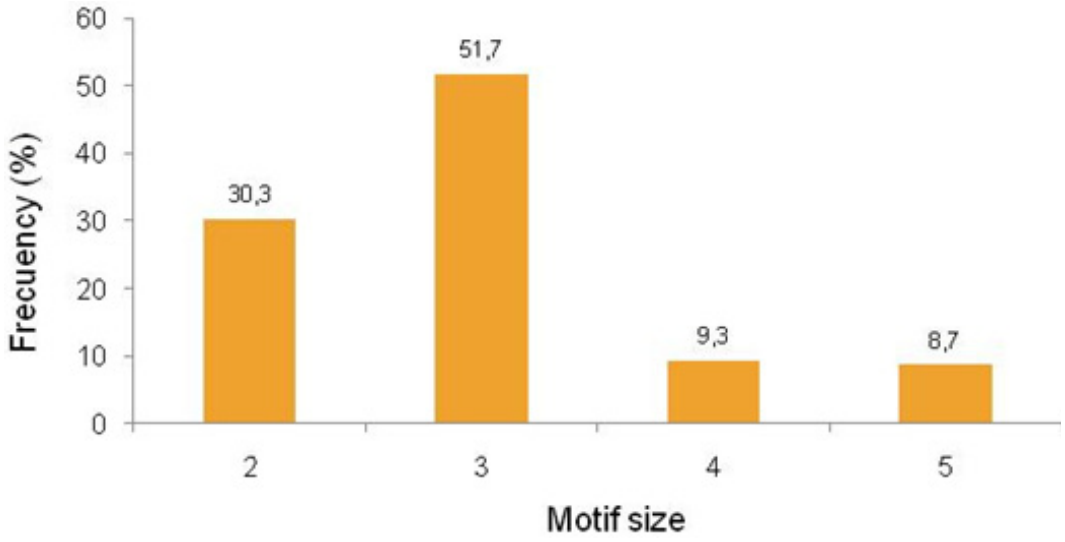


Fig. 2 Frequency distribution of EST-SSR based on motif sequence type.

Linkage analysis and map construction

Genotyping data from 94 RILs were included in a previously developed data matrix composed by 343 AFLP and 191 SSRs markers (Kiani et al. 2007). Segregation pattern of each marker allele in the progeny was assessed using the GQMOL program (Schuster and Cruz, 2004, available at http://www.ufv.br/dbg/gqmol/gqmol.htm) by FDR (False Discovery Rate) test (Benjamin and Hocheberg, 1995).

The map was constructed using Carthagene 0.999 (Schiex and Gaspin, 1997) and Mapmaker 3.0 (Lander et al. 1987). Loci were assembled into groups using likelihood odds (LOD) ratios with a LOD threshold of 4.0 and a maximum recombination frequency threshold of 0.35 (Kiani et al. 2007). The likelihoods of different locus-order possibilities were compared and the one having the highest likelihood was selected for each linkage group. Kosambi mapping function (Kosambi, 1994) was used to calculate map distances (cM) from recombination frequency. Locus order and map distance were also tested using MapMaker 3.0. Finally, the genetic map was drawn using the Mapchart 2.1 program (Voorrips, 2002).

Table 2. Information of SSR, EST-SSR and EST-InDel markers localized in the genetic linkage map. Marker name, the accession number in Genbank and/or TIGR, the repeat type and motive, the allele size in bp, the sequence of forward and reverse primers and the linkage group in the genetic map is provided.

SSR name	GenBank access number	Motive type	Repeat type	Product (bp)	Forward primer (5´- 3´)	Reverse primer (5´- 3´)	LG
HA77	BV727857	dinucleotide	GA	249	TGTAATCTGTATCACTTCCACC	GTTGTTCTGTTAGGTCGTTCCG	1
HA95	G67405	dinucleotide	GA	162-166	ACGCTTGATAGACAAATGCT	TAGGCAGCAGGTTTTACTCT	5
HA3621	BV728258	trinucleotide	ATC	158	TTAAATCTGGTGGCAACTTT	TACTCGCTCCACCTCTACAT	8
HA503	BV727935	dinucleotide	GA-TA	100	GTAGATCTTTCCCCTGCACCC	GATGCGTTAGAGATAACCTG	8
HA541	BV727938	dinucleotide	GT	220	CAACACACGCTTTGTTGGGTG	CGATAAGGAGGGAGGGGAAGATAG	3
HA911	BV727888	dinucleotide	GA	179	CAAAGTTCACCTCGTTTTTC	AAGTGGGAAGGTCTACGAGT	8
HA920	BV727955	trinucleotide	ATC	171	CCCCATGTAATCGCTTTAG	GTGGGGATATAGTGTTCGGC	10
HA1155	BV727896	dinucleotide	GA-GT	110	CAATCAAAACATGTGTATGTG	GGATTCTGCTGAAAGGAGACAAC	12
HA1626	G67417	trinucleotide	ATT	170	GATGTTACACGTTAGCAACG	GAACTCAGCCTAAAAGTC	13
HA1770	BV727917	dinucleotide	GA	186	GCCGGTCACCCATTCCCCCCTTAC	GGCGGAAGCCAGGCGTTGTAGCG	3
HA2190	BV728130	dinucleotide	GA	189	CAAACCCTAATCGCCCAATTGC	GGAAGCTTGTGAATCTTGGAGGG	3
HA2989	BV728090	hexa-trinucleotide	ATCCTC-ATC	145	GCCTGTGTCCTCCTCAGAGTC	GTTATGTGGGCTGGCTGCACC	14
HA3204	BV728181	trinucleotide	ATC	121	GCCCTTCAATCCTACCATTAA C	GGATATTGAGTTGGTTGTTGGG	10
HA3272	BV728357	dinucleotide	TA GA	123	TGTTCTCACACTTTTCACCA	TAGGCCCTTAAATTCAGATG	7
HA3312	BV728075	dinucleotide	GA	115	TAACTGCAACAGACATGGAA	ATAACGATTGCACAACACAA	6
HA3325	BV728106	dinucleotide	GA	222	GTCATTGCTTGAGTGTGAGA	CGACGCTGAACATATACTTG	2
HA3417	BV728230	dinucleotide	GA	122	TAATTGATTGGGGGTAAATG	TATGATTTGGTGTGCTCAGA	13
HA3640	BV728259	dinucleotide	GA	184	GCCATATCAATAACCAATCCC	CAAAGCATGTGAGGACCAGAATG	11
HA3703	-	dinucleotide	GA	217	CAAATGCTGATTCCACACTA	ATGGTTTCCTGTTTGAATTG	4
HA3878	BV728314	trinucleotide	ATC ATT	262	TTTGTTTAGCATCATCATCATC	GAGACCCTAACCATAACATGA	7
HA3950	BV728296	dinucleotide	C GA	145	CATTCCTCACCAACATAACA	CCCTTGAGTATGCTTCAAAC	14
HA3971	BV728318	dinucleotide	GA	179	GGATCTCTCCTGGTCCTAGT	GATCAATATCTTCCATGTTAT	17
HA4011	BV728360	dinucleotide	GA	126	ACTTCTACCCTCCCCTTCTT	CTGTACACGTGCTGCTTTAG	13
HA4023	BV728189	dinucleotide	GA	254	CGTCGTCTATCTCCAGTAACC	GAATCTTGGGACCTGAGTCACC	9

HA4057	BV728333	dinucleotide	GT	172	AAACCCTTCCGACTTATCTC	TAAAGAGAGAGCCCAACAAG	3
HA4058	BV728334	dinucleotide	AT	156	GATTATGATTCCATGAGCCAAG	CATCCTGTAATCCTGTCAATGTC	14
HA4105	BV728354	dinucleotide	GA	118	GGGACTAGTTTGTAATATCTCTC	GTTTGAGAGTGGTGATAGGTTATG	3
HA4112	BV728340	dinucleotide	GA	183	TTGACTCTCCTTCTCGTCTC	GCTCTCAAGAAATCGGTTAG	2
HA4149	BV728355	dinucleotide	GA	111	TCAATTCATCGTGCATATCG	CCAAAGTCCACCAAATCTTCC	17
HA4264	BV728346	dinucleotide	GT	169	TCCCAAACCTACACCTCTAA	AAACAAACTATGGATTAACAAGG	14
ORS420	BV005977	dinucleotide	GT	138	TCATGGTGTTTGGTTTGTGTC	TGCCAAATTCCTCTTCTTTCT	15
ORS460	BV005998	dinucleotide	GA	302	ACTCGGCTACCACCTCACAC	GCCCTTTGACCCTAACCAAG	7
ORS483	BV006013	dinucleotide	GT	271	CCGAACAACAATCTCCACAA	GGTTTAGGTGTCGCATCACA	6
HA4149	BV728355	dinucleotide	GA	111	TCAATTCATCGTGCATATCG	CCAAAGTCCACCAAATCTTCC	17
HA4264	BV728346	dinucleotide	GT	169	TCCCAAACCTACACCTCTAA	AAACAAACTATGGATTAACAAGG	14
ORS420	BV005977	dinucleotide	GT	138	TCATGGTGTTTGGTTTGTGTC	TGCCAAATTCCTCTTCTTTCT	15
ORS460	BV005998	dinucleotide	GA	302	ACTCGGCTACCACCTCACAC	GCCCTTTGACCCTAACCAAG	7
ORS483	BV006013	dinucleotide	GT	271	CCGAACAACAATCTCCACAA	GGTTTAGGTGTCGCATCACA	6
ORS662	BV006121	dinucleotide	AG	314	CGGGTTGGATATGGAGTCAA	CCTTTACAAACGAAGCACAATTC	1
ORS679	BV006133	dinucleotide	GA	240	CCCTCCTCCCTCTTCACTTT	CTCATCGGACAACCAGAACC	12
ORS691	BV006140	dinucleotide	CT	447	GCATCTGAGCAACTGCGTTA	ACCGTCCTTAGCTCTTGTGAG	10
ORS788	BV006202	dinucleotide	GA	263	CTGGATAAAGATGGGATAAAGAGAG	GGACCCACCAAGATTTGTTTT	16
ORS934	BV006334	dinucleotide	AG	252	CGTTCCGATCTACAAGGTAAAGG	AACATACACAAATCCACGCAGAG	6
ORS993	BV006383	dinucleotide	AG	327	GTGTTACAATCCTTGCGTCGATA	CAAACCACAGGGACCAAAATG	16
ORS1065	BV006445	dinucleotide	CT	295	ACCGCTGTCAACACCTTAAACTC	GGCTGGGAATCAACTGCTACTAC	2
ORS1265	BV006617	dinucleotide	CT	222	GGGTTTAGCAAATAATAGGCACA	ACCCTTGGAGTTTAGGGATCA	9

Unigene Name	TIGR access number	Motive type	Repeat type	Product (bp)	Forward primer (5´- 3´)	Reverse primer (5´- 3´)	LG
TC15366	TC15366	InDel	GGTTA	282	ATGAATTGAAGCATGCAGTAG	ATGAGTTCTTTATCAGCCAGTAT	10
TC18146	TC18146	InDel	CCAAATCT	447	ATGAATTTTGGCTTGAAGAAGG	TGCCGGGTATGTATATGGGAAT	10
TC18194	TC18194	dinucleotide	TG	305	AGAGGGAAGACACAAACTGAC	AATCAACTCATCATCAATCCA	3
TC20625	TC20625	trinucleotide	GAG	209	GGTGGTTCACAAACATAAGG	CTTCTTCTCCATTTCCAATCT	4
TC21689	TC21689	trinucleotide	GTG	357	GTTACGGAGGTGGTGGTC	CCAATAAACACAGAATCCAAA	4
TC23602	TC23602	tetranucleotide	GTTT	200	TTCCAACACAAGGCATCA	AAACAACAGTCACAATACATCAA	8
TC23762	TC23762	InDel	GTAGAAATTCTTTCTGAGGTA	779	AAACTTCACGCAATGAGACT	CAAGCTTCCCTACAAAGAAA	14
TC24343	TC24343	tetranucleotide	TTGT	220	TAAGGCAAAGAACCAGGGAC	ATGAAGCACAACTAAAGGCA	16
TC24992	TC24992	trinucleotide	AAT	180	ACATTCACTTCTTCTCCATCA	AGCCATACTGTCCTCCATACT	8
TC25054	TC25054	pentanucleotide	TTTAC	318	TACCTTGACGATGACCAGA	ACCTTACATTCCACCAAACA	16
TC26323	TC26323	dinucleotide	AT	303	ATCACAAACCTCCTGACAAC	CAACTCTTGGGTGGCTAA	17
TC26869	TC26869	trinucleotide	GAT	288	AGAGGAGGAGTTTGTTGATGT	CCAATAAACACAGAATCCAAA	16

RESULTS

Search for ESTs containing SSR motifs

HAGI database was used to identify EST-SSRs containing di, tri- tetra and penta-nucleotide with a minimum length of 18 (di- and tri-) to 20 (tetra-penta) bases. These parameters were chosen based on previous studies performed on different species that indicate that the polymorphism level decreases with repetitions shorter than 18 bases (Cho et al. 2000). CUGI-SSR Server identified 1,134 repetition motifs with at least one repeat from a total of 36,741 analyzed sequences (3.08%).

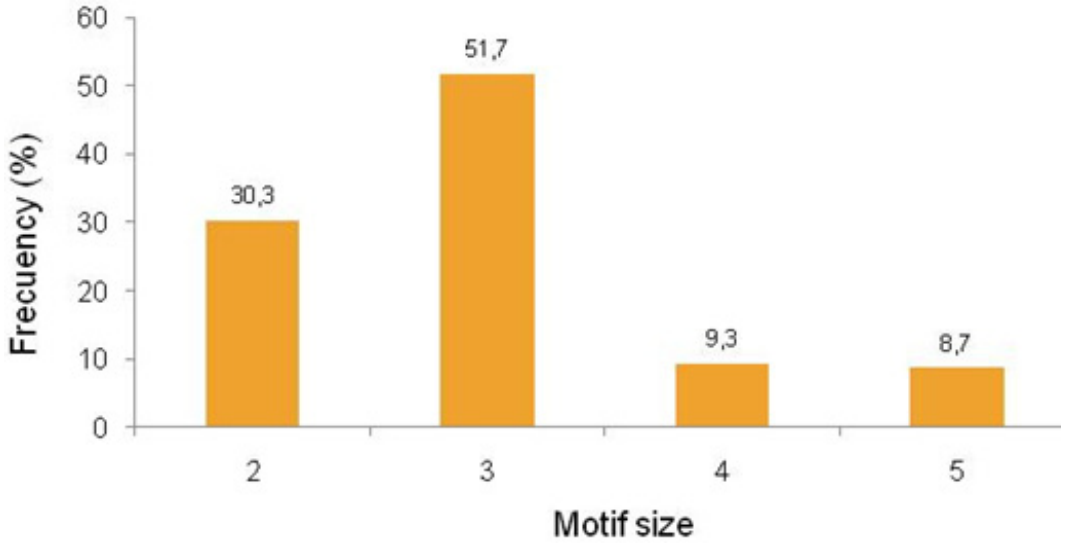


Fig. 2 Frequency distribution of EST-SSR based on motif sequence type.

The relative abundance of di-, tri-, tetra-, and penta-nucleotide was 30.28%, 51.65%, 9.32% and 8.74% respectively (Figure 1).

Mononucleotides were not taken into account in this study. One hundred and forty six unigenes (12.9%) out of 1,134 SSR exhibited more than one SSR motif.

The GA/CT motif (23.2%) was the most abundant dinucleotide SSR, while the AC/GT (3.8%), AT/TA (3.2%) and CG/GC (0.08%) were detected at a lower frequency. All possible combinations of trinucleotide repeat motifs were detected in the sunflower unigenes, showing AAG/CTT (11.9%), ACC/GGT (9.2%), and AAT/ATT (7.8%) higher frequencies (Figure 2).

BLASTX analysis and GO mapping were performed for functional annotation of the genotyped EST-SSR markers. Fifteen out of seventeen sequences had at least one significant BLAST hit and eleven accessions were mapped to one or more GO terms each (Table 1).

Amplification and polymorphism degree of EST-SSR and genomic SSR

Among the 127 EST-SSR primer pairs tested in four parental lines, 70 showed clear, easy scoring banding patterns, 50 failed to amplify, 3 amplified null alleles (0/1 or 1/0) and the remaining ones generated unspecific amplification or complex amplification patterns. Twenty three EST-SSR markers were polymorphic within the parents of one or both mapping populations.

The analysis of 167 genomic SSRs performed within the parents of the PAC2 x RHA266 mapping population showed 37.7% polymorphic markers, 40.6% of monomorphic, 15% of null alleles and 6.7% failed the amplification reaction for both alleles.

Table 3. Marker distribution among the linkage groups.

Linkage group	AFLP	SSR	EST-SSR	InDels	Total marker number	Length (cM)	Density (cM/*locus*)
1	13	8	-	-	21	83.3	3.9
2	14	15	-	-	29	112.5	3.8
3	16	13	1	-	30	92.9	3.0
4	20	7	2	-	29	95.9	3.3
5	16	11	-	-	27	103.5	3.8
6	13	8	-	-	21	83.6	3.9
7	15	8	-	-	23	71.9	3.1
8	15	17	2	-	34	112.0	3.2
9	26	16	-	-	42	141.7	3.3
10	45	23	-	2	70	203.0	2.9
11	14	7	-	-	21	77.5	3.6
12	16	15	-	-	31	109.1	3.5
13	10	10	-	-	20	89.3	4.4
14	18	20	-	1	39	153.4	3.9
15	11	13	-	-	24	89.2	3.7
16	20	28	3	-	51	178.7	3.5
17	22	12	1	-	35	173.4	4.9
Total	304	231	9	3	547	1,942.3	3.6

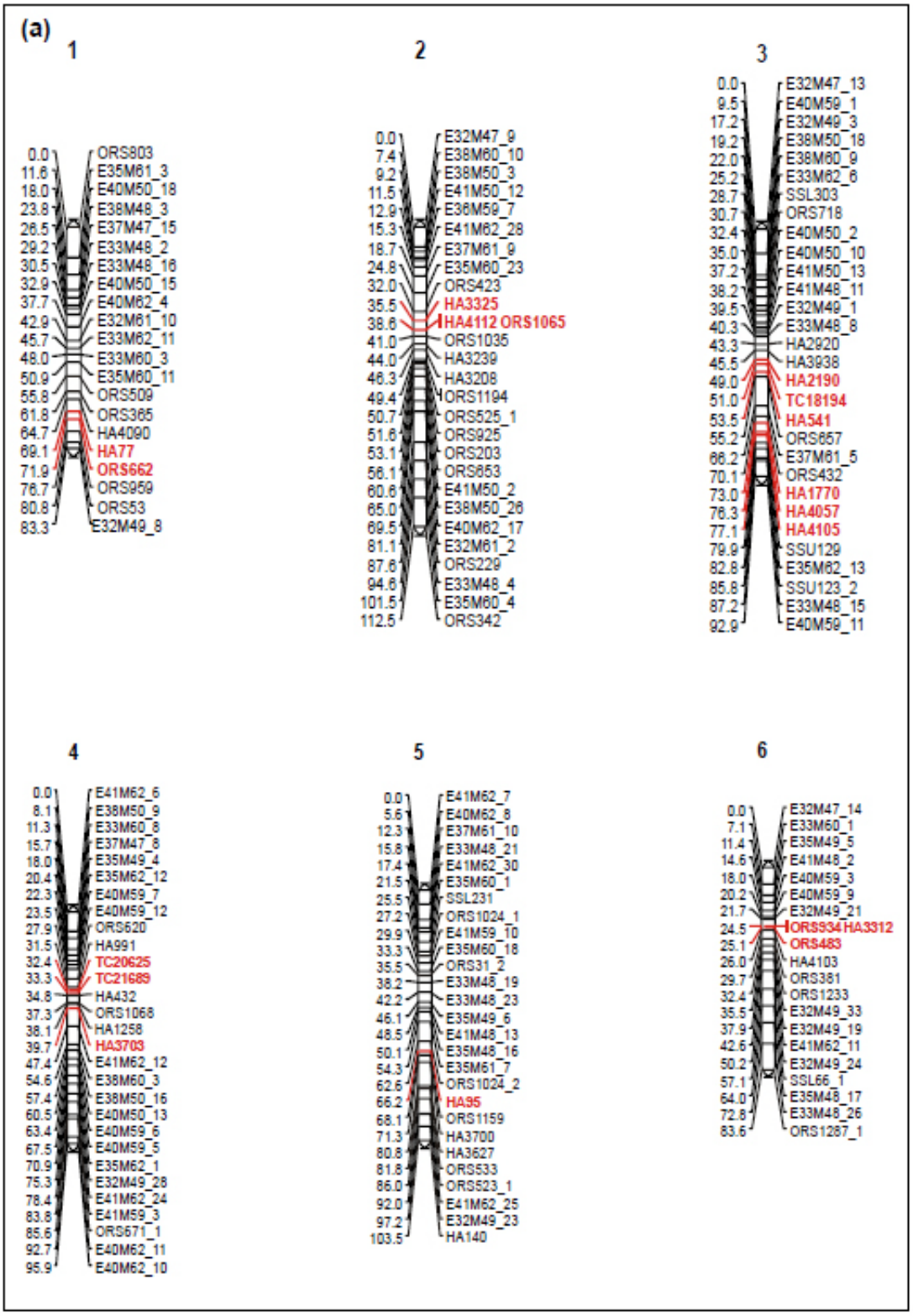
Genetic mapping

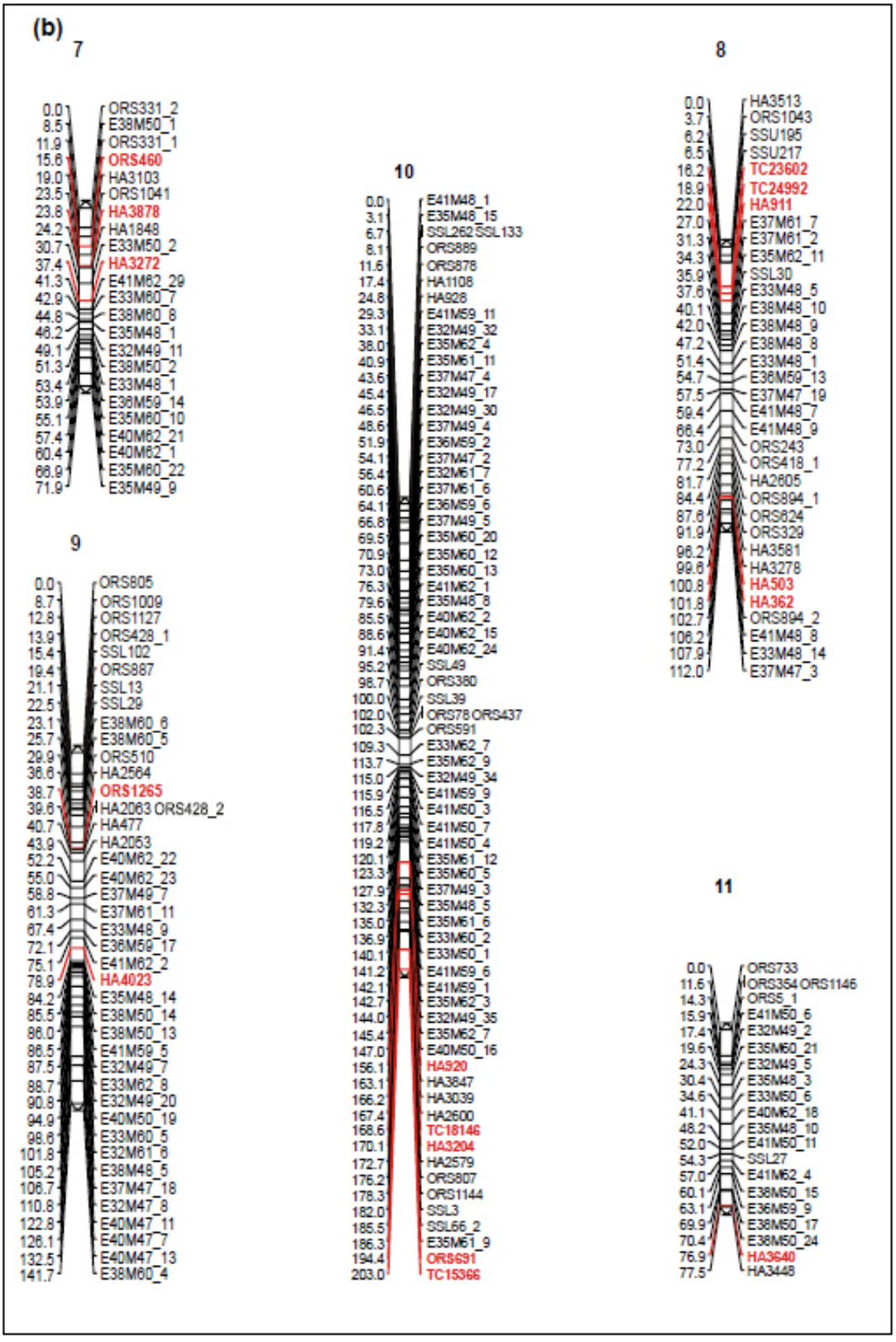
Sixty three SSR, 14 EST-SSR and the three previously described InDels displaying clear polymorphism between parental lines were assayed in 94 RILs. A total of 53 new markers could be positioned on the sunflower linkage map described by Kiani et al. (2007). Of these, 41 markers were genomic SSR (30 HAx and 11 ORSx) and 12 correspond to new functional markers (9 EST-SSR and 3 InDels) developed in this work (Table 2).

Molecular marker distribution among the linkage groups in the resulting map is presented in Table 3. The map now contains 547 markers (304 AFLP, 231 SSR, 9 EST-SSR and 3 InDels) placed in 17 linkage groups (Figure 3).

This linkage map was constructed using a minimum LOD score of 4.0 and a maximum recombination value of 0.35. The total map length is 1,942.3 cM with a mean density of 3.6 cM per locus (Table 3). Linkage group 10 presented the highest density value (2.9 cM/locus), while LG 17 presented the lowest one (4.9 cM/locus). The groups ranged in length from 71.9 (LG 7) to 203 cM (LG 10) and carried between 20 (GL 13) to 70 (LG 10) markers. Twenty two SSR and five EST-SSR markers remained unlinked to any group.

Among the newly identified functional markers, some are predicted to derive from key genes as those coding for transcription factors, kinases that may play important roles in signal transduction, stress resistance and metabolism (Table 1).


(a)
1
0.0 ORS803
11.6 E35M61_3
18.0 E40M50_18
23.8 E38M48_3
26.5 E37M47_15
29.2 E33M48_2
30.5 E33M48_16
32.9 E40M50_15
37.7 E40M62_4
42.9 E32M61_10
45.7 E33M62_11
48.0 E33M60_3
50.9 E35M60_11
55.8 ORS509
61.8 ORS365
64.7 HA4090
69.1 HA77
71.9 ORS662
76.7 ORS959
80.8 ORS53
83.3 E32M49_8
2
0.0 E32M47_9
7.4 E38M60_10
9.2 E38M50_3
11.5 E41M50_12
12.9 E38M59_7
15.3 E41M62_28
18.7 E37M61_9
24.8 E35M60_23
32.0 ORS423
35.5 HA3325
38.6 HA4112 ORS1065
41.0 ORS1035
44.0 HA3239
46.3 HA3208
49.4 ORS1194
50.7 ORS525_1
51.6 ORS925
53.1 ORS203
56.1 ORS653
60.6 E41M50_2
65.0 E38M50_26
69.5 E40M62_17
81.1 E32M61_2
87.6 ORS229
94.6 E33M48_4
101.5 E35M60_4
112.5 ORS342
3
0.0 E32M47_13
9.5 E40M59_1
17.2 E32M49_3
19.2 E38M50_18
22.0 E38M60_9
25.2 E33M62_6
28.7 SSL303
30.7 ORS718
32.4 E40M50_2
35.0 E40M50_10
37.2 E41M50_13
38.2 E41M48_11
39.5 E32M49_1
40.3 E33M48_8
43.3 HA2920
45.5 HA3938
49.0 HA2190
51.0 TC18194
53.5 HA541
55.2 ORS657
66.2 E37M61_5
70.1 ORS432
73.0 HA1770
76.3 HA4057
77.1 HA4105
79.9 SSU129
82.8 E35M62_13
85.8 SSU123_2
87.2 E33M48_15
92.9 E40M59_11
4
0.0 E41M62_6
8.1 E38M50_9
11.3 E33M60_8
15.7 E37M47_8
18.0 E35M49_4
20.4 E35M62_12
22.3 E40M59_7
23.5 E40M59_12
27.9 ORS620
31.5 HA991
32.4 TC20625
33.3 TC21689
34.8 HA432
37.3 ORS1068
38.1 HA1258
39.7 HA3703
47.4 E41M62_12
54.6 E38M60_3
57.4 E38M50_16
60.5 E40M50_13
63.4 E40M59_6
67.5 E40M59_5
70.9 E35M62_1
75.3 E32M49_28
78.4 E41M62_24
83.8 E41M59_3
85.6 ORS671_1
92.7 E40M62_11
95.9 E40M62_10
5
0.0 E41M62_7
5.6 E40M62_8
12.3 E37M61_10
15.8 E33M48_21
17.4 E41M62_30
21.5 E35M60_1
25.5 SSL231
27.2 ORS1024_1
29.9 E41M59_10
33.3 E35M60_18
35.5 ORS31_2
38.2 E33M48_19
42.2 E33M48_23
46.1 E35M49_6
48.5 E41M48_13
50.1 E35M48_16
54.3 E35M61_7
62.6 ORS1024_2
66.2 HA95
68.1 ORS1159
71.3 HA3700
80.8 HA3627
81.8 ORS533
86.0 ORS523_1
92.0 E41M62_25
97.2 E32M49_23
103.5 HA140
6
0.0 E32M47_14
7.1 E33M60_1
11.4 E35M49_5
14.6 E41M48_2
18.0 E40M59_3
20.2 E40M59_9
21.7 E32M49_21
24.5 ORS934 HA3312
25.1 ORS483
26.0 HA4103
29.7 ORS381
32.4 ORS1233
35.5 E32M49_33
37.9 E32M49_19
42.6 E41M62_11
50.2 E32M49_24
57.1 SSL66_1
64.0 E35M48_17
72.8 E33M48_26
83.6 ORS1287_1


(b)
7
0.0 ORS331_2
8.5 E38M50_1
11.9 ORS331_1
15.6 ORS460
19.0 HA3103
23.5 ORS1041
23.8 HA3878
24.2 HA1848
30.7 E33M50_2
37.4 HA3272
41.3 E41M62_29
42.9 E33M60_7
44.8 E38M60_8
46.2 E35M48_1
49.1 E32M49_11
51.3 E38M50_2
53.4 E33M48_1
53.9 E38M59_14
55.1 E35M60_10
57.4 E40M62_21
60.4 E40M62_1
66.9 E35M60_22
71.9 E35M49_9
9
0.0 ORS805
8.7 ORS1009
12.8 ORS1127
13.9 ORS428_1
15.4 SSL102
19.4 ORS887
21.1 SSL13
22.5 SSL29
23.1 E38M60_6
25.7 E38M60_5
29.9 ORS510
36.6 HA2564
38.7 ORS1265
39.6 HA2063 ORS428_2
40.7 HA477
43.9 HA2053
52.2 E40M62_22
55.0 E40M62_23
58.8 E37M49_7
61.3 E37M61_11
67.4 E33M48_9
72.1 E36M59_17
75.1 E41M62_2
78.9 HA4023
84.2 E35M48_14
85.5 E38M50_14
86.0 E38M50_13
86.5 E41M59_5
87.5 E32M49_7
88.7 E33M62_8
90.8 E32M49_20
94.9 E40M50_19
98.6 E33M60_5
101.8 E32M61_6
105.2 E38M48_5
106.7 E37M47_18
110.8 E32M47_8
122.8 E40M47_11
126.1 E40M47_7
132.5 E40M47_13
141.7 E38M60_4
10
0.0 E41M48_1
3.1 E35M48_15
6.7 SSL262 SSL133
8.1 ORS889
11.6 ORS878
17.4 HA1108
24.8 HA928
29.3 E41M59_11
33.1 E32M49_32
38.0 E35M62_4
40.9 E35M61_11
43.6 E37M47_4
45.4 E32M49_17
46.5 E32M49_30
48.6 E37M49_4
51.9 E36M59_2
54.1 E37M47_2
56.4 E32M61_7
60.6 E37M61_6
64.1 E36M59_6
66.8 E37M49_5
69.5 E35M60_20
70.9 E35M60_12
73.0 E35M60_13
76.3 E41M62_1
79.6 E35M48_8
85.5 E40M62_2
88.6 E40M62_15
91.4 E40M62_24
95.2 SSL49
98.7 ORS380
100.0 SSL39
102.0 ORS78 ORS437
102.3 ORS591
109.3 E33M62_7
113.7 E35M62_9
115.0 E32M49_34
115.9 E41M59_9
116.5 E41M50_3
117.8 E41M50_7
119.2 E41M50_4
120.1 E35M61_12
123.3 E35M60_5
127.9 E37M49_3
132.3 E35M48_5
135.0 E35M61_6
136.9 E33M60_2
140.1 E33M50_1
141.2 E41M59_6
142.1 E41M59_1
142.7 E35M62_3
144.0 E32M49_35
145.4 E35M62_7
147.0 E40M50_16
156.1 HA920
163.1 HA3847
166.2 HA3039
167.4 HA2600
168.6 TC18146
170.1 HA3204
172.7 HA2579
176.2 ORS807
178.3 ORS1144
182.0 SSL3
185.5 SSL66_2
186.3 E35M61_9
194.4 ORS691
203.0 TC15366
8
0.0 HA3513
3.7 ORS1043
6.2 SSU195
6.5 SSU217
16.2 TC23602
18.9 TC24992
22.0 HA911
27.0 E37M61_7
31.3 E37M61_2
34.3 E35M62_11
35.9 SSL30
37.6 E33M48_5
40.1 E38M48_10
42.0 E38M48_9
47.2 E38M48_8
51.4 E33M48_1
54.7 E36M59_13
57.5 E37M47_19
59.4 E41M48_7
66.4 E41M48_9
73.0 ORS243
77.2 ORS418_1
81.7 HA2605
84.4 ORS894_1
87.6 ORS824
91.9 ORS329
96.2 HA3581
99.6 HA3278
100.8 HA503
101.8 HA362
102.7 ORS894_2
106.2 E41M48_8
107.9 E33M48_14
112.0 E37M47_3
11
0.0 ORS733
11.6 ORS354 ORS1146
14.3 ORS5_1
15.9 E41M50_6
17.4 E32M49_2
19.6 E35M60_21
24.3 E32M49_5
30.4 E35M48_3
34.6 E33M50_6
41.1 E40M62_18
48.2 E35M48_10
52.0 E41M50_11
54.3 SSL27
57.0 E41M62_4
60.1 E38M50_15
63.1 E36M59_9
69.9 E38M50_17
70.4 E38M50_24
76.9 HA3640
77.5 HA3448

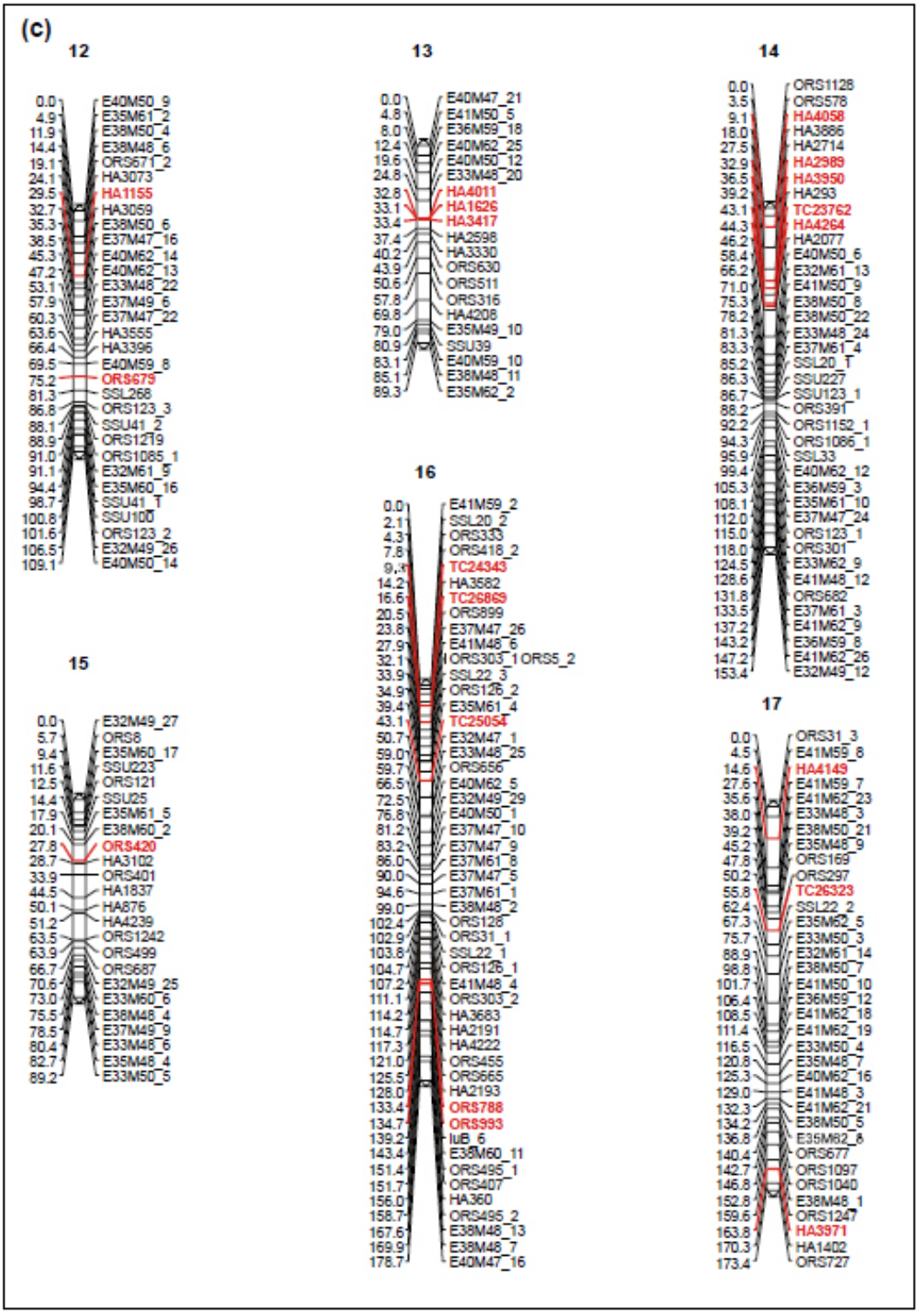


Fig. 3 Genetic linkage map of SSRs, EST-SSR, EST-InDels and AFLP on a population of RILs derived from crossing of PAC2 x RHA266. Linkage group described was numbered according to the nomenclature of SSR map described by Tang et al. (2003). Map distances (cM) between loci are indicated on the left side of each linkage group. Markers incorporated to this map version are highlighted in red.

DISCUSSION

A strategy for identification and characterization of EST-SSR from a sunflower unigene database was implemented to unravel new functional markers which were further mapped along with other genomic SSR and InDels, in a previously developed reference map (Kiani et al. 2007).

SSR derived from the unigenes database of *H. annuus* were detected in 3.08% of the analyzed sequences, under the criteria set of 18 bases for the di-and trinucleotide, and 20 bases for the tetra- and pentanucleotides. This result is consistent with that found by other authors that reported rates of EST-SSR ranging between 1.5% to 4.7% for species like barley, maize, rice, sorghum and wheat (Kantety et al. 2002; Gao et al. 2003). However, in other species such as wheat and pepper, frequencies with values above 11% have been reported (Nicot et al. 2004; Yi et al. 2006). Kumpatla and Mukhopadhyay (2005) analyzed EST-SSR in 49 dicotyledonous species, including sunflower and described that the frequencies may vary from 2.65% to 16.82%. These authors analyzed 60,007 sunflowers EST deposited in GenBank database (http://www.ncbi.nlm.nih.gov/dbEST/) and found a frequency of ESTs-SSRs of 6.18%. The criteria used for the EST-SSR identification was set in a minimum repetition value of 10 nucleotides. These authors conclude that mononucleotide SSR were the most frequent motives, followed by dinucleotide and trinucleotide SSRs. On the other hand, Pashley et al. (2006) also identified SSRs in sunflower ESTs sequences deposited in CGPDB database (http://cgpdb.ucdavis.edu), with a selection criterion of a minimum of 10 repeat units for dinucleotide and 12 for tri-and tetranucleotide motifs. They found that trinucleotides were the most abundant type of repetitions, followed by tetranucleotides and then dinucleotides. The most abundant trinucleotide motif was ATG/CAT. A selection of 188 sequences was used in their study to design primers, which were amplified in 12 species of wild *H. annuus* with the aim to analyze the transferability of these markers, and they identify a set of 48 functional EST-SSR. Recently, Heesacker et al. (2008) developed EST-SSR from a collection of 17,904 unigenes obtained from the assembly of 89,225 ESTs deposited in GenBank from *H. annuus*, *H. argophyllus* and *H. paradoxus*. Considering a repetition length ≥10 bp for dinucleotide, and a minimum number of repetitions ≥5 for the rest of the motives, they found that 10.9% of the 17,904 unigenes of *H. annuus* contained SSRs. Also, under these criteria, the dinucleotide motif presented the most frequent rate of recurrence, followed by trinucleotide, and finally tetranucleotide.

In general, differences in results between these studies may be attributed to the different data sets used to conduct the studies (ESTs and/or unigenes from different assembly processes), the tools and criteria used in the different analysis. In particular, the differences between the observations of this study and those of other authors can be explained based on the conditions set for the present analysis, regarding the inclusion of only perfectly matched microsatellites with a length greater than or equal to 20 bp and the exclusion of mononucleotides.

According to the procedures applied in this work, trinucleotide repeats were the most abundant type of motives in sunflower unigenes database (HAGI), represented by 51.7% of the sequences, followed by dinucleotide (30.3%), then tetranucleotides (9.3%) and finally pentanucleotides with 8.7%. This dominance of trimeric SSRs over di-, tetra-, and pentameric SSRs may be explained on the basis of the suppression of non-trimeric SSRs in coding regions due to the risk of frameshift mutations that may occur when those microsatellites alternate in length in one unit (Rungis et al. 2004). This may be because repeat numbers of trinucleotide microsatellites can change without altering the reading frame of the messenger RNA. It also could be due to the higher number of possible trinucleotide combinations compared to those for dinucleotide repeats. Morgante et al. (2002) suggested that mutation pressure and positive selection for specific single amino acid stretches could account for the doubled frequency of tri-nucleotide repeats relative to mono- and dinucleotide repeats in the genes of plant species. The most abundant dinucleotide motif found in the database analyzed was GA/CT, present in 23.2% of the sequences. The encoding triplets that can be constructed based on dinucleotide repetitions may constitute different codons, depending on the reading frame, and be translated into different amino acids. For example, GA/CT can represent codons GAG, AGA, and CUC, UCU in a population of mRNA and translated into amino acids Arg, Glu, Ala and Leu, respectively. Ala and Leu are present with high frequency in proteins, 8% and 10% respectively (Lewin, 1994). This could be one of the reasons why GA/CT is present at high frequency in the collections of ESTs. Previous research showed that the dinucleotide AG was the most common in plant ESTs (Morgante et al. 2002; Varshney et al. 2002; Kumpatla and Mukhopadhyay, 2005; Parida et al. 2006; Yi et al. 2006). The most common sunflower trinucleotide found in this work was the AAG/CTT (11.9%), followed by the ACC/GGT (9.2%) and the AAT/ATT (7.8%). According to Kumpatla and Mukhopadhyay (2005),

AAG/CTT is the most abundant motive in dicot species. However, these authors reported that the most frequent trinucleotide in sunflower was GGT/ACC.

The analysis of genomic markers performed within the parents of the mapping population (RHA266 x PAC2) showed 37.7% polymorphic markers. These results agree with those reported by Tang et al. (2002), who found a polymorphism level of 41.2% between genotypes RHA280 x RHA801 and with the previous study (Kiani et al. 2007). The percentage of null alleles in the genomic analysis of the SSR was 15%, coinciding this with a study reported by Tang et al. (2002), which showed a rate of 13.7% in the loci analyzed. Validation of genic SSRs in four genotypes of sunflower (RHA266, PAC2, HA89 and RHA801) resulted in amplification of 74 sequences from a total of 127 analyzed. Out of them, 13% represented polymorphic loci, 45% monomorphic, 5% null alleles and the remaining 37% showed either no amplification product, nonspecific amplification or complex or difficult to resolve banding patterns. The percentage of polymorphism observed coincides with that reported by Heesacker et al. (2008), which conclude that less than 10% of the transcribed loci in sunflower can be genetically mapped using SSR, and in agreement with reports for other species (Eujayl et al. 2004; Fraser et al. 2004; Varshney et al. 2005). The frequency of EST-SSR marker polymorphism was 18%. In general, EST-SSR has demonstrated less polymorphism compared with genomic SSRs in crop plants due to higher DNA sequence conservation in transcribed regions (Cho et al. 2000; Scott et al. 2000; Rungis et al. 2004).

Here we report an improved high density genetic linkage map of sunflower constructed using a basic matrix of 584 markers (343 AFLPs, 258 SSR, 14 EST-SSR and 3 EST-InDels) after incorporation of 53 new markers (41 SSR, 9 EST-SSR and 3 EST-InDels). The genome length spanned by this new linkage map is longer (1,942.3 cM) than that constructed using a 94 RILs population from a cross between RHA280 x RHA801 based on SSR/InDel markers (Tang et al. 2002), or from that constructed using SSR and SNP markers as described by Lai et al. (2005) which are 1,368.3 and 1,349.3 cM, respectively. This may be the result of using a combination of different marker types, such as AFLPs and SSRs which exhibit a differential distribution pattern (Sebastian et al. 2000; Mei et al. 2004; Syed et al. 2006; Zhang et al. 2006). Moreover, additional effects such as population size, marker clusters, the effects of distorted segregation markers, or the process of framework map construction could contribute to these differences (Cervera et al. 2001; Zhang et al. 2006).

Although most of the newly added markers locations spread all along the genome, 3 of them mapped to the outermost positions of their respective linkage groups when compared to the previously published map (Kiani et al. 2007). These markers were: ORS691 and TC15366 (located in LG 10) and HA3640 which is located to 0.6 cM from telomeric marker HA3448 (LG 11). In addition to enlarging the current genetic map of sunflower in telomeric regions, some other new markers helped to close previous gaps. TC23602, TC24992 and HA911 closed a gap of 10.8 cM present in LG 8, reducing the gap to 5 cM. Incorporation of a single marker (ORS420) in LG 15 allowed closing two previous gaps of 11.6 and 13.6 cM. As a consequence, no gaps longer than 13.2 cM (which is present in LG 17) can be found throughout the entire map, which increases the feasibility of detecting genes or traits of agronomic importance in sunflower.

However, the most interesting perspective of application of the newly mapped markers will be their use to better define and characterize previously published QTLs in the added chromosome positions. These markers were mapped to different chromosome locations (Table 2) and some of them mapped close to QTLs involved in the genetic control of water status and osmotic adjustment (Kiani et al. 2007), chlorophyll fluorescence parameters (Kiani et al. 2008), resistance to *P. macdonaldi* (Alfadil et al. 2007; Darvishzadeh et al. 2007), fatty acid composition in water-stressed conditions in the greenhouse and the field (Ebrahimi et al. 2008). Among the above mentioned QTL overlapping markers, some correspond to functional markers (Table 1). For example, TC26869 (which corresponds to a putative heat shock related protein gene) mapped close to a water status/osmotic adjustment QTL and TC26323 (encoding for a putative photosynthesis related protein gene) that colocalizes with a chlorophyll fluorescence determining QTL (Kiani et al. 2008). Although the functional demonstration that these genes actually are determinants of these traits is beyond the objectives of the present work, both the colocalization and the putative functions are very likely suggesting hints for their hypothetical involvement.

The fact that EST-SSRs exhibited sequence similarity to genes with a wide range of functions (Table 1) suggests that there is potential to identify markers that may be directly involved in determining agronomical important characters. For example plant peroxidases are a well-studied group of heme-

containing enzymes for which many different functions have been proposed. In the majority of plant species investigated they occur as distinctive isoenzymes which can be constitutively expressed or induced in response to external factors such as wounding, stress and attack by pathogens (Veitch, 2004).

In conclusion, this map complements the previously published public map with a dense framework of AFLP, SSR, EST-SSR and EST-InDels markers loci in place for analysis of *H. annuus* (Kiani et al. 2007) suitable for the development of approaches that could be used to characterize traits which are difficult to manipulate through breeding programs.

Financial support: This work was supported by INTA-AEBIO1330 /241331, and granted from the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT/ PAV2004-137), Argentina.

REFERENCES

ACUÑA, C.V.; FERNANDEZ, P.; VILLALBA, P.; GARCIA MARTÍN N.; HOPP, H. E. and MARCUCCI POLTRI, S.N. (2010). Discovery, validation and *in silico* functional characterization of EST-SSR markers in *Eucalyptus globulus*. *Tree Genetics & Genomes*. In press, 2010.

AL-CHAARANI, G.R.; GENTZBITTEL, L.; HUANG, X.Q. and SARRAFI, A. (2004). Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). *TAG Theoretical and Applied Genetics,* vol. 109, no. 7, p. 1353-1360.**[CrossRef]**

ALFADIL, T.A.; KIANI, S.P.; DECHAMP-GUILLAUME, G.; GENTZBITTEL, L. and SARRAFI, A. (2007). QTL mapping of partial resistance to Phoma basal stem and root necrosis in sunflower (*Helianthus annuus* L.). *Plant Science,* vol. 172, no. 4, p. 815-823. **[CrossRef]**

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W. and LIPMAN, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, vol. 215, no. 3, p. 403-410. **[CrossRef]**

ASHBURNER, M.; BALL, C.A.; BLAKE, J.A.; BOTSTEIN, D.; BUTTLER, H.; CHERRY, J.M.; DAVIS, A.P.; DOLINSKY, K.; DWIGHT, S.S.; EPPIG, J.T.; HARRIS, M.A.; HILL, D.P.; ISSEL-TARVER, L.; KASARSKIS, A.; LEWIS, S.; MATESE, J.C.; RICHARDSON, J.E.; RINGWALD, M.; RUBIN, G.M. and SHERLOCK, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics,* vol. 25, no. 1, p. 25-29. **[CrossRef]**

BENJAMIN, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistics Society B,* vol. 57, no. 1, p. 289-300.

BERRY, S.T.; LEON, A.J.; HANFREY, C.C.; CHALLIS, P.; BURKHOLZ, A.; BARNES, S.R.; 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. *TAG Theoretical and Applied Genetics,* vol. 91, no. 2, p. 195-199. **[CrossRef]**

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: *Proceedings of the Sunflower Research Workshop* (19, 9th - 10th, January, 1997, Fargo, ND, USA). p.113-118.

BERRY, S.T.; LEON, A.J.; CHALLIS, P.; LIVIN, C.; JONES, R.; HANFREY, C.C.; GRIFFITHS, S. and ROBERTS, A. (2003). Construction of a high density, composite RFLP linkage map for cultivated sunflower *Helianthus annuus*. In: *Proceedings of the International Sunflower Conference*. (14, 12th - 20th June, 1996, Beijing, China). Paris, International Sunflower Association, vol. 2, p.1150-1160.

CERVERA, M.-T.; STORME, V.; IVENS, B.; GUSMÃO, J.; LIU, B.H.; HOSTYN, V.; VAN SLYCKEN, J.; VAN MONTAGU, M. and BOERJAN, W. (2001). Dense genetic linkage maps of three Populus species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics,* vol. 158, no. 2, p. 787-809.

CERVIGNI, G.D.L.; PANIEGO, N.; DIAZ, M.; SELVA, J.P.; ZAPPACOSTA, D.; ZANAZZI, D.; LANDERRECHE, I.; MARTELOTTO, L.; FELITTI, S.; PESSINO, S.; SPANGENBERG, G. and ECHENIQUE, V. (2008). Expressed sequence tag analysis and development of gene associated markers in a near-isogenic plant system of *Eragrostis curvula*. *Plant Molecular Biology,* vol. 67, no. 1-2, p. 1-10. **[CrossRef]**

CHABANE, K.; ABLETT, G.A.; CORDEIRO, G.M.; VALKOUN, J. and HENRY, R.J. (2005). EST *versus* genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genetic Resources and Crop Evolution,* vol. 52, no. 7, p. 903-909. **[CrossRef]**

CHO, Y.G.; ISHII, T.; TEMNYKH, S.; CHEN, X.; LIPOVICH, L.; MCCOUCH, S.R.; PARK, W.D.; AYRES, N. and CARTINHOUR, S. (2000). Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *TAG Theoretical and Applied Genetics,* vol. 100, p. 713-722. **[CrossRef]**

CONESA, A. and GÖTZ, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics,* vol. 2008, article ID 619832, p. 619-832. **[CrossRef]**

CORDEIRO, G.M.; CASU, R.; MCINTYRE, C.L.; MANNERS, J.M. and HENRY, R.J. (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Science,* vol. 160, no. 6, p. 1115-1123. **[CrossRef]**

DARVISHZADEH, R.; KIANI, S.P.; DECHAMP-GUILLAUME, G.; GENTZBITTEL, L. and SARRAFI, A. (2007). Quantitative trait loci associated with isolate specific and isolate nonspecific partial resistance to *Phoma macdonaldii* in sunflower. *Plant Pathology,* vol. 56, no. 5, p. 855-861. [CrossRef]

EBRAHIMI, A.; MAURY, P.; BERGER, M.; KIANI, S.P.; NABIPOUR, A.; SHARIATI, F.; GRIEU, P. and SARRAFI, A. (2008). QTL mapping of seed-quality traits in sunflower recombinant inbred lines under different water regimes. *Genome,* vol. 51, no. 8, p. 599-615. [CrossRef]

EUJAYL, I.; SORRELLS, M.; BAUM, M.; WOLTERS, P. and POWELL, W. (2001). Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica,* vol. 119, no. 1-2, p. 39-43. [CrossRef]

EUJAYL, I.; SLEDGE, M.K.; WANG, L.; MAY, G.D.; CHEKHOVSKIY, K.; ZWONITZER, J.C. and MIAN, M.A.R. (2004). *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *TAG Theoretical and Applied Genetics,* vol. 108, no. 3, p. 414-422. [CrossRef]

FERNÁNDEZ, P.; PANIEGO, N.; LEW, S.; HOPP, H.E. and HEINZ, R.A. (2003). Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. *BMC Genomics,* vol. 4, no. 40. [CrossRef]

FERNANDEZ-SILVA, I.; EDUARDO, I.; BLANCA, J.; ESTERAS, C.; PICÓ, B.; NUEZ, F.; ARÚS, P.; GARCIA-MAS, J. and MONFORTE, A.J. (2008). Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.). *TAG Theoretical and Applied Genetics,* vol. 118, no. 1, p. 139-150. [CrossRef]

FRASER, L.G.; HARVEY, C.F.; CROWHURST, R.N. and SILVA, H.N. (2004). EST-derived microsatellites from *Actinidia* species and their potential for mapping. *TAG Theoretical and Applied Genetics,* vol. 108, p. 1010-1016. [CrossRef]

FUSARI, C.M.; LIA, V.V.; HOPP, H.E.; HEINZ, R.A. and PANIEGO, N.B. (2008). Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. *BMC Plant Biology,* vol. 8, no. 7. [CrossRef]

GAO, L.; TANG, J.; LI, H. and JIA, J. (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding,* vol. 12, no. 3, p. 245-261. [CrossRef]

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,* vol. 44, no. 2, p. 213-221. [CrossRef]

GENTZBITTEL, L.; VEAR, F.; ZHANG, Y.-X.; BERVILLÉ, A. and NICOLAS, P. (1995). Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). *TAG Theoretical and Applied Genetics*, vol. 90, no. 7-8, p. 1079-1086. [CrossRef]

GENTZBITTEL, L.; MESTRIES, E.; MOUZEYAR, S.; MAZEYRAT, F.; BADAOUI, S.; VEAR, F.; LABROUHE, D.T. and NICOLAS, P. (1999). A composite map of expressed sequences and phenotypic traits of the sunflower *(Helianthus annuus* L.) genome. *TAG Theoretical and Applied Genetics,* vol. 99, no. 1-2, p. 218-234. [CrossRef]

HACKAUF, B. and WEHLING, P. (2002). Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breeding,* vol. 121, no. 1, p. 17-25. [CrossRef]

HEESACKER, A.; KISHORE, V.K.; GAO, W.; TANG, S.; KOLKMAN, J.M.; GINGLE, A.; MATVIENKO, M.; KOZIK, A.; MICHELMORE, R.M.; LAI, Z.; RIESEBERG, L.H. and KNAPP, S.J. (2008). SSRs and INDELs mined from the sunflower EST database: Abundance, polymorphisms, and cross-taxa utility. *TAG Theoretical and Applied Genetics,* vol. 117, no. 7, p. 1021-1029. [CrossRef]

JAN, C.C.; VICK, B.A.; MILLER, J.F.; KAHLER, A.L. and BUTLER, E.T. (1998). Construction of an RFLP linkage map for cultivated sunflower. *TAG Theoretical and Applied Genetics,* vol. 96, no. 1, p. 15-22. [CrossRef]

KANTETY, R.V.; LA ROTA, M.; MATTHEWS, D.E. and SORRELLS, M.E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology,* vol. 48, no. 5-6, p. 501-510. [CrossRef]

KIANI, S.P.; TALIA, P.; MAURY, P.; GRIEU, P.; HEINZ, R.; PERRAULT, A.; NISHINAKAMASU, V.; HOPP, E.; GENTZBITTEL, L.; PANIEGO, N. and SARRAFI, A. (2007). Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under water treatments. *Plant Science,* vol. 172, no. 4, p. 773-787. [CrossRef]

KIANI, S.P.; MAURY, P.; SARRAFI, A. and GRIEU, P. (2008). QTL analysis of chlorophyll fluorescence parameters in sunflower (*Helianthus annuus* L.) under well-watered and water-stressed conditions. *Plant Science,* vol. 175, no. 4, p. 565-573. [CrossRef]

KOLKMAN, J.M.; SLABAUGH, M.B.; BRUNIARD, J.M.; BERRY, S.; BUSHMAN, B.S.; OLUNGU, C.; MAES, N.; ABRATTI, G.; ZAMBELLI, A.; MILLER, J.F.; LEON, A. and KNAPP, S.J. (2004). Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. *TAG Theoretical and Applied Genetics,* vol. 109, no. 6, p. 1147-1155. [CrossRef]

KOLKMAN, J.M.; BERRY, S.T.; LEON, A.J.; SLABAUGH, M.B.; TANG, S.; GAO, W.; SHINTANI, D.K.; BURKE, J.M. and KNAPP, S.J. (2007). Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics,* vol. 177, no. 1, p. 457-468. [CrossRef]

KOSAMBI, D.D. (1994). The estimation of a map distance from recombination values. *Annals of Eugenics,* vol. 12, no. 3, p. 172-175.

KUMPATLA, S. and MUKHOPADHYAY, S. (2005). Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome,* vol. 48, no. 6, p. 985-998. [CrossRef]

LAI, Z.; LIVINGSTONE, K.; ZOU, Y.; CHURCH, S.A.; KNAPP, S.J.; ANDREWS, J. and RIESEBERG, L.H. (2005). Identification and mapping of SNPs from ESTs in sunflower. *TAG Theoretical and Applied Genetics*, vol. 111, no. 8, p. 1532-1544. [CrossRef]

LANDER, E.S.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALY, M.J.; LINCOLN, S.E. and NEWBURG, L. (1987). MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics,* vol. 1, no. 2, p. 174-181. [CrossRef]

LEE, J.M.; NAHM, S.H.; KIM, Y.M. and KIM, B.D. (2004). Characterization and molecular genetic mapping of microsatellite loci in pepper. *TAG Theoretical and Applied Genetics,* vol. 108, no. 4, p. 619-627. [CrossRef]

LEIGH, F.; LEA, V.; LAW, J.; WOLTERS, P.; POWELL, W. and DONINI, P. (2003). Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica*, vol. 133, no. 3, p. 359-366. [CrossRef]

LEWIN, B. (1994). *Genes V.* New York, Oxford University Press.1272 p. ISBN-10 0198542879.

LIU, A. and BURKE, J.M. (2006). Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics,* vol. 173, no. 1, p. 321-330. [CrossRef]

MEI, M.; SYED, N.H.; GAO, W.; THAXTON, P.M.; SMITH, C.W.; STELLY, D.M. and CHEN, Z.J. (2004). Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). *TAG Theoretical and Applied Genetics,* vol. 108, no. 2, p. 280-291. [CrossRef]

MORGANTE, M. and OLIVIERI, A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal,* vol. 3, no. 1, p. 175-182. [CrossRef]

MORGANTE, M.; HANAFEY, M. and POWELL, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics,* vol. 30, no. 2, p. 194-200. [CrossRef]

NICOT, N.; CHIQUET, V.; GANDON, B.; AMILHAT, L.; LEGEAI, F.; LEROY, P.; BERNARD, M. and SOURDILLE, P. (2004). Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *TAG Theoretical and Applied Genetics,* vol. 109, no. 4, p. 800-805. [CrossRef]

PANIEGO, N.; ECHAIDE, M.; MUNOZ, M.; FERNANDEZ, L.; TORALES, S.; FACCIO, P.; FUXAN, I.; CARRERA, M.; ZANDOMENI, R.; SUAREZ, E.Y. and HOPP, H.E. (2002). Microsatellite isolation and characterization in sunflower (*Helianthus annuu*s L.). *Genome,* vol. 45, no. 1, p. 34-43. [CrossRef]

PARIDA, S.K.; KUMAR, K.A.R.; DALAL, V.; SINGH, N.K. and MOHAPATRA, T. (2006). Unigene derived microsatellite markers for the cereal genomes. *TAG Theoretical and Applied Genetics,* vol. 112, no. 5, p. 808-817. [CrossRef]

PASHLEY, C.H.; ELLIS, J.R.; MCCAULEY, D.E. and BURKE, J.M. (2006). EST databases as a source for molecular markers: Lessons from *Helianthus*. *Journal of Heredity,* vol. 97, no. 4, p. 381-388. [CrossRef]

PEERBOLTE, R.P. and PELEMAN, J. (1996). The Cartisol sunflower RFLP map (146 loci) extended with 291 AFLP markers. In: *Proceedings of the 18th Sunflower Research Forum.* (18, 1th - 12th January, 1996, Fargo, ND, USA). p. 174-178.

RIESEBERG, L.H. (1998). Genetic mapping as a tool for studying speciation**.** In: SOLTIS D.E.; SOLTIS P.S. and DOYLE, J.J. eds*. Molecular systematics of plants II DNA sequencing.* New York, Kluwer Academic Publisher. p. 459-487.

ROATH, W.W.; MILLER, J.F. and GULYA, T.J. (1981). Registration of RHA 801 sunflower germplasm1 (Reg. No. GP_5). *Crop Science,* vol. 21, no. 3, p. 479. [CrossRef]

RUNGIS, D.; BÉRUBÉ, Y.; ZHANG, J.; RALPH, S.; RITLAND, C.E.; ELLIS, B.E.; DOUGLAS, C.; BOHLMANN, J. and RITLAND, K. (2004). Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *TAG Theoretical and Applied Genetics,* vol. 109, no. 6, p. 1283-1294. [CrossRef]

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. *Proceeding of the National Academy of Sciences of the United States of America,* vol. 81, no. 24, p. 8014-8018. [CrossRef]

SCHIEX, T. and GASPIN, C. (1997). CARTHAGENE: Constructing and joining maximum likelihood genetic maps. In: *Proceedings of the 5th International Conference on Intelligent Systems for Molecular Biology*. (5, 21th - 26th June, 1997, Halkidiki, Greece). vol. 5, p. 258-267.

SCHUSTER, I. and CRUZ, C.D. (2004). Estatística genômica aplicada a populações derivadas de cruzamentos controlados. Viçosa, Brasil, Ediciones Universidade Federal de Viçosa, 568 p. ISBN 85-7269-1919-X.

SCOTT, K.D.; EGGLER, P.; SEATON, G.; ROSSETTO, M.; ABLETT, E.M.; LEE, L.S. and HENRY, R.J. (2000). Analysis of SSRs derived from grape ESTs. *TAG Theoretical and Applied Genetics,* vol. 100, no. 5, p. 723-726. [CrossRef]

SEBASTIAN, R.L.; HOWELL, E.C.; KING, G.J.; MARSHALL, D.F. and KEARSEY, M.J. (2000). An integrated AFLP and RFLP *Brassica olearacea* linkage map from two morphologically distinct doubled-haploid mapping populations. *TAG Theoretical and Applied Genetics,* vol. 100, no. 1, p. 75-81. [CrossRef]

SHARMA, R.K.; BHARDWAJ, P.; NEGI, R.; MOHAPATRA, T. and AHUJA, P.S. (2009). Identification, characterization and utilization of unigene derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC Plant Biology,* vol. 9, no. 53. [CrossRef]

SINGH, R.; ZAKI, N.M.; TING, N.-C.; ROSLI, R.; TAN, S.-G.; LOW, E.-T.L.; ITHNIN, M. and CHEAH, S.-C. (2008). Exploiting an oil palm EST database for the development of gene-derived SSR markers and their exploitation for assessment of genetic diversity. *Biologia,* vol. 63, no. 2, p. 227-235. [CrossRef]

SYED, N.H.; SØRENSEN, A.P.; ANTONISE, R.; VAN DE WIEL, C.; VAN DER LINDEN, C.G.; VAN'T WESTENDE, W.; HOOFTMAN, D.A.P.; DEN NIJS, H.C.M. and FLAVELL, A.J. (2006). A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers. *TAG Theoretical and Applied Genetics*, vol. 112, no. 3, p. 517-527. [CrossRef]

TANG, S.; YU, J.-K.; SLABAUGH, M.B.; SHINTANI, D.K. and KNAPP, S.J. (2002). Simple sequence repeat map of the sunflower genome. *TAG Theoretical and Applied Genetics,* vol. 105, no. 8, p. 1124-1136. [CrossRef]

TANG, S.; KISHORE, V. K. and KNAPP, S.J. (2003). PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. *TAG Theoretical and Applied Genetics,* vol. 107, no. 1, p. 6-19. [CrossRef]

THIEL, T.; MICHALEK, W.; VARSHNEY, R. and GRANER, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *TAG Theoretical and Applied Genetics,* vol. 106, no. 3, p. 411-422. **[CrossRef]**

UENO, S.; TAGUCHI, Y. and TSUMURA, Y. (2008). Microsatellite markers derived from *Quercus mongolica* var. *crispula* (Fagaceae) inner bark expressed sequence tags. *Genes & Genetic Systems,* vol. 83, no. 2, p. 179-187. **[CrossRef]**

VARSHNEY, R.K.; THIEL, T.; STEIN, N.; LANGRIDGE, P. and GRANER, A. (2002). *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cellular and Molecular Biology Letters,* vol. 7, no. 2A, p. 537-46.

VARSHNEY, R.K.; GRANER, A. and SORRELLS, M.E. (2005). Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology,* vol. 23, no. 1, p. 48-55. **[CrossRef]**

VEITCH, N.C. (2004). Structural determinants of plant peroxidase function. *Phytochemistry Reviews,* vol. 3, no. 1-2, p. 3-18. **[CrossRef]**

VOORRIPS, R.E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*, vol. 93, no. 1, p. 77-78. **[CrossRef]**

YI, G.; LEE, J.; LEE, S.; CHOI, D. and KIM, B.-D. (2006). Exploitation of pepper EST-SSRs and an SSR-based linkage map. *TAG Theoretical and Applied Genetics,* vol. 114, no. 1, p. 113-130. **[CrossRef]**

YU, J.-K.; TANG, S.; SLABAUGH, M.B.; HEESACKER, A.; COLE, G.; HERRING, M.; SOPER, J.; HAN, F.; CHU, W.-C.; WEBB, D.M.; THOMPSON, L.; EDWARDS, K.J.; BERRY, S.; LEON, A.J.; GRONDONA, M.; OLUNGU, C.; MAES, N. and KNAPP, S.J. (2003). Towards a saturated molecular genetic linkage map for cultivated sunflower. *Crop Science,* vol. 43, no. 1, p. 367-387. **[CrossRef]**

YU, J.-K.; DAKE, T.M.; SINGH, S.; BENSCHER, D.; LI, W.; GILL, B. and SORRELLS, M.E. (2004). Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. *Genome,* vol. 47, no. 5, p. 805-818. **[CrossRef]**

ZHANG, F.; WAN, X.-Q. and PAN, G.-T. (2006). QTL mapping of *Fusarium moniliforme* ear rot resistance in maize: 1. Map construction with microsatellite and AFLP markers. *Journal of Applied Genetics,* vol. 47, no. 1, p. 9-15.

ZHANG, L.S.; LE CLERC, V.; LI, S. and ZHANG, D. (2005). Establishment of an effective set of simple sequence repeat markers for sunflower variety identification and diversity assessment. *Canadian Journal of Botany,* January 2005, vol. 83, no. 1, p. 66-72. **[CrossRef]**

How to cite this article:

TALIA, P.; NISHINAKAMASU, V.; HOPP, H.E.; HEINZ, R.A. and PANIEGO, N. (2010). Genetic mapping of EST-SSRs, SSR and InDels to improve saturation of genomic regions in a previously developed sunflower map. *Electronic Journal of Biotechnology*, vol. 13, no. 6. http://dx.doi.org/10.2225/vol13-issue6-fulltext-14