Theor Appl Genet (2009) 120:85–91
DOI 10.1007/s00122-009-1161-8

ORIGINAL PAPER

Development, polymorphism, and cross-taxon utility of EST–SSR markers from safflower (*Carthamus tinctorius* L.)

Mark A. Chapman · John Hvala · Jason Strever · Marta Matvienko · Alexander Kozik · Richard W. Michelmore · Shunxue Tang · Steven J. Knapp · John M. Burke

Received: 10 February 2009 / Accepted: 11 September 2009 / Published online: 10 October 2009

Abstract Due to their highly polymorphic and codominant nature, simple-sequence repeat (SSR) markers are a common choice for assaying genetic diversity and genetic mapping. In this paper, we describe the generation of an expressed-sequence tag (EST) collection for the oilseed crop safflower and the subsequent development of EST–SSR markers for the genetic analysis of safflower and related species. We assembled 40,874 reads into 19,395 unigenes, of which 4,416 (22.8%) contained at least one SSR. Primer pairs were developed and tested for 384 of these loci, resulting in a collection of 104 polymorphic markers that amplify reliably across 27 accessions (3 species) of the genus *Carthamus*. These markers exhibited a high level of polymorphism, with an average of 6.0 ± 0.4 alleles per locus and an average gene diversity of 0.54 ± 0.03 across *Carthamus* species. In terms of cross-taxon transferability, 50% of these primer pairs produced an amplicon in at least one other species in the Asteraceae, and 28% produced an amplicon in at least one species outside the safflower subfamily (i.e., lettuce, sunflower, and/or *Gerbera*). These markers represent a valuable resource for the genetic analysis of safflower and related species, and also have the potential to facilitate comparative map-based analyses across a broader array of taxa within the Asteraceae.

Introduction

Simple-sequence repeats (SSRs) are found throughout eukaryotic genomes, occurring in both protein-coding and

Communicated by A. Bervillé.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1161-8) contains supplementary material, which is available to authorized users.

M. A. Chapman · J. Hvala · J. Strever · J. M. Burke (✉)
Department of Plant Biology, University of Georgia,
Miller Plant Sciences Building, Athens, GA 30602, USA
e-mail: jmburke@uga.edu

M. A. Chapman
e-mail: mchapman@plantbio.uga.edu

J. Hvala
e-mail: jhvala@plantbio.uga.edu

J. Strever
e-mail: jstrever@uga.edu

M. Matvienko · A. Kozik · R. W. Michelmore
Department of Plant Sciences, The Genome Center,
University of California, Davis, CA 935616, USA
e-mail: matvienko@gmail.com

A. Kozik
e-mail: akozik@atgc.org

R. W. Michelmore
e-mail: rwmichelmore@ucdavis.edu

S. Tang · S. J. Knapp
Center for Applied Genomic Technologies,
University of Georgia, 111 Riverbend Road,
Athens, GA 30602, USA
e-mail: shunxue.tang@gmail.com

S. J. Knapp
e-mail: sjknapp@uga.edu

non-coding regions (Levinson and Gutman 1987). Variation in SSR length occurs primarily due to slipped-strand mispairing during replication (Levinson and Gutman 1987) and mutations of this sort occur at a relatively high frequency. As such, SSRs are typically codominant and highly polymorphic, and have been a common source of markers for genetic mapping, molecular breeding, and population genetic analyses in a wide variety of species (e.g., Chase et al. 1996; Xiong et al. 1999; Matsuoka et al. 2002; Tang et al. 2003; Neeraja et al. 2007).

Although the utility of SSR markers has been well-established, their *de novo* development can be a costly and time-consuming endeavor on a locus-by-locus basis (Squirrell et al. 2003). Consequently, expressed-sequence tag (EST) databases have become an increasingly valuable resource for SSR marker development (e.g., Pinto et al. 2004; Feingold et al. 2005; Pashley et al. 2006; Yi et al. 2006; Becher 2007; Hanai et al. 2007; Laurent et al. 2007; Heesacker et al. 2008). EST databases are now available for more than 140 biologically and economically important plant species (Heesacker et al. 2008), and SSRs are typically abundant within these sequence collections (Ellis and Burke 2007). A key advantage of these EST-SSRs is that they are often more transferable across species as compared to so-called 'anonymous' SSRs from non-coding sequences (e.g., Cordeiro et al. 2001; Yu et al. 2004; Pashley et al. 2006), thereby facilitating comparative genetic analyses. This higher level of cross-taxon marker portability facilitates comparative genetic analyses.

The primary goal of the present study was to develop a set of polymorphic EST-SSR markers for use in safflower (*Carthamus tinctorius* L., Asteraceae) and related species. Safflower is an herbaceous crop that is native to the hot, dry climates of the Fertile Crescent. While it was traditionally grown on a local scale in the Old World as a source of dye (carthamine), food coloring/flavoring, and medicinal extracts (Weiss 1971), safflower cultivation expanded to the New World near the end of the 19th century. By the 1950s, safflower was being bred for commercial oil production (Knowles 1958) and, over the past decade, there has also been growing interest in using it as a platform for the large-scale production of plant-made pharmaceuticals (Lacey et al. 1998).

Safflower is a member of the Asteraceae, which is one of the largest and most diverse flowering plant families, comprising more than 23,000 species (Stevens 2006).The most recent molecular phylogenetic analysis of the family recognized 12 subfamilies (Panero and Funk 2008), though three of these (the Asteroideae, Cichorioideae, and Carduoideae) account for nearly 95% of the species. In recent years, molecular marker development efforts within the Asteraceae have focused primarily on the two most economically important species within the family, sunflower (*Helianthus annuus*; Asteroideae) and lettuce (*Lactuca sativa*; Cichorioideae) with little attention having been paid to members of the Carduoideae.

To date, genetic analyses of safflower have been performed using a variety of randomly amplified nuclear markers (i.e., amplified fragment length polymorphisms [AFLPs], inter-simple sequence repeats [ISSRs], and randomly amplified polymorphic DNAs [RAPDs]; Sehgal and Raina 2005; Johnson et al. 2007). Unfortunately, such marker systems are typically dominant, and it can also be difficult to establish the homology of such markers across populations/species, thereby limiting their utility in comparative studies. In order to remedy this situation, we mined the publicly available safflower EST database for SSRs, designed primers to amplify a subset of these loci, screened them for polymorphism, and tested their transferability across the genus *Carthamus*, as well as more broadly across the family.

Methods

Generation of EST library, EST sequencing and contig assembly

Library construction began with RNA extraction from a single safflower accession (AC Sunset; PI 592391). Tissues (roots from 14 day old seedlings, young leaves, and leaf stalk from 8-leaf stage seedlings held 2 days in the dark, pre-anthesis florets and developing achenes) were harvested into liquid nitrogen, ground to a powder and RNA extracted using TRIzol following the manufacturer's protocol (Invitrogen, Carlsbad, CA, USA).

The separate RNA samples were combined following extraction and cDNA was synthesized using SMART cloning technology (Takara Clontech, Palo Alto, CA). cDNA libraries were subsequently normalized using the Trimmer-Direct kit (Evrogen, Moscow, Russia) and size-fractionated into small, medium and large cDNA pools. Each pool was directionally cloned into the pBRcDNASfiAB vector (see http://compgenomics.ucdavis.edu) and transformed into *E. coli*.

Clones were sequenced at the Joint Genome Institute using ABI3730 automated DNA sequencers (Applied Biosystems, Foster City, CA, USA) and then deposited at the Arizona Genomic Institute (http://www.genome.arizona.edu).Sequence processing, annotation and assembly were carried out using the CGPDB pipeline (http://cgpdb.ucdavis.edu/cgpdb2/; Kozik et al. 2002) as described by Heesacker et al. (2008).

SSR discovery, primer design and amplification

Following assembly, the unigene set was searched for SSRs using SSR-IT (Temnykh et al. 2001) with a minimum size

of n (number of motif repeats) ≥ 5 for dinucleotide SSRs and $n \geq 4$ for tri- and tetranucleotide SSRs. PCR primers were then designed to flank SSRs using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Primer pairs were screened for successful amplification on a screening panel consisting of DNA samples from six diverse safflower lines plus two individuals of *C. oxyacanthus*. DNA was extracted using a modified CTAB method (Doyle and Doyle 1990) from fresh tissue grown from seed.

PCR amplification utilized a modification of the three-primer amplification protocol described by Schuelke (2000), as detailed in Wills et al. (2005). Each reaction contained 10 ng of template DNA, 30 mM tricine pH 8.4-KOH, 50 mM KCl, 2 mM $MgCl_2$, 100 μM each dNTP, 0.02 μM forward primer (with the M13(-29) sequence appended to the 5′ end), 0.1 μM reverse primer, 0.1 μM fluorescently labeled M13(-29) primer, and one unit of *Taq* DNA polymerase. Cycling conditions followed a 'touch-down' protocol, as follows: initial denaturation at 95°C for 3 min; 10 cycles of 30 s at 94°C, 30 s at 65°C (annealing temperature was reduced by 1° per cycle), and 45 s at 72°C; 30 cycles of 30 s at 94°C, 30 s at 55°C, and 45 s at 72°C; and a final extension time of 20 min at 72° C.

PCR products were resolved on agarose gels to check for amplification. In cases where no amplification was evident, or if fewer than six of the eight individuals screened produced amplicons, the primer pair was discarded. Likewise, primer pairs that produced apparently multilocus amplification profiles were not considered further. For loci that appeared to produce a single-locus product, the amplicons were diluted 1:50 and visualized on an ABI 3730xl DNA sequencer (Applied Biosystems) with MapMarker 1000 ROX size standards (BioVentures Inc., Murfreesboro, TN, USA) included in each lane to allow for accurate fragment size determination. Alleles were called using the software package GeneMarker (SoftGenetics LLC, State College, PA, USA). Markers that were monomorphic in the eight individuals were not considered further, as they are likely to represent loci with no (or very low) SSR variation.

Once a suitable panel of polymorphic markers was identified, they were used to genotype a set of 24 safflower accessions (Table 1), 12 individuals of *C. palaestinus* (PI 235663), which is the apparent progenitor of safflower, as well as twelve individuals of *C. oxyacanthus* (accessions PI 426488 and PI 426428), which is a more distantly related species in the same section. In addition, the primer pairs were used to test for cross-species amplification outside the genus *Carthamus*. The cross-taxon screening included five members of the safflower subfamily (the Carduoideae) plus representatives of three other subfamilies (lettuce, Cichorioideae; sunflower, Asteroideae and *Gerbera,* Mutisioideae; Table 2). Seed of each species were obtained from the USDA National Plant Germplasm System (http://www.ars-grin.gov/npgs/) with the exception of lettuce (*L. sativa* cv. Salinas), which was obtained from the Michelmore lab, and *Gerbera,* which came from a plant in the greenhouse at the University of Georgia.

Table 1 The 24 safflower accessions used in the study of marker diversity

Accession PI	Accession name	Country of origin
PI 167390	BJ-2645	Turkey
PI 193473	BJ-673	Ethiopia
PI 195895	BJ-681	Morocco
PI 209287	BJ-774	Romania
PI 209296	BJ-782	Kenya
PI 209300	BJ-786	Kenya
PI 220283	BJ-795	Afghanistan
PI 239042	BJ-820	Morocco
PI 250081	BJ-986	Egypt
PI 250533	BJ-1048	Egypt
PI 251285	Col. No. K1111	Jordan
PI 251291	BJ-1131	Jordan
PI 253527	BJ-2485	France
PI 253895	BJ-1222	Syria
PI 257582	BJ-1252	Ethiopia
PI 262433	BJ-2716	Ethiopia
PI 271070	TOZI SPINY	Sudan
PI 279051	U. Cal 61-20	India
PI 279345	Yamagata4	Japan
PI 301053	N-86	Turkey
PI 401470	BJ-2023	Bangladesh
PI 405984	BJ-2081	Iran
PI 544041	Honghua	China
PI 576992	CART 72/86	North Korea

Table 2 The eight species outside *Carthamus* used to determine marker portability

Accession PI	Species
W6 7111	*Centaurea cyanus* L.
W6 30011	*Saussurea nuda* Ledeb.
W6 30135	*Cirsium quercetorum* (A. Gray) Jeps.
PI 503532	*Cynara cardunculus* L.
PI 639183	*Echinops ritro* L.
PI 599773	*Helianthus annuus* L.
–[a]	*Lactuca sativa* L.
–	*Gerbera* spp.

[a] Not from the USDA (see text)

Data analysis

The level of polymorphism per locus (number of alleles, nA, and expected heterozygosity [i.e., gene diversity], H_e) was calculated using the program GDA (Lewis and Zaykin 2001). We also performed a principal coordinates analysis (PCO) of the 48 individuals based on genotypes at all 104 loci using GenAlEx (Peakall and Smouse 2002). Polymorphism information content (PIC) values were calculated using PIC calculator (http://www.liv.ac.uk/~kempsj/pic.html).

Results

Prevalence of SSRs

The 40,874 safflower (*C. tinctorius*) ESTs (Genbank numbers EL372565-EL412381 and EL511108-EL511145) assembled into 19,395 unigenes (7,154 contigs and 12,241 singletons). Approximately 79% of the safflower unigenes showed sequence similarity with an *Arabidopsis* gene (see http://compgenomics.ucdavis.edu). 6,180 SSRs were detected in 4,416 unigenes (with one to eight per unigene). Excluding SSRs which were too close (within 50 bp) to the ends of the sequence (precluding primer design) the database contained 4,835 SSRs in 3,572 unigenes. The majority of these (71.2%) were trinucleotide repeats, with di- and tetranucleotide-repeats making up 24.5 and 4.4%, respectively. The longest dinucleotide-repeat spanned 56 bp (i.e., 28 repeat units) whereas the longest tri- and tetranucleotide-repeats were 42 and 40 bp, respectively (corresponding to 14 and 10 ten repeat units, respectively).

Amplification in safflower and wild *Carthamus* species

Primer pairs were designed for 384 randomly chosen EST-SSR loci. Of these, 216 (56.3%) produced an apparently single-locus amplification product in at least six of the eight individuals on the screening panel, with 148 being polymorphic. One hundred and four primer pairs, presented in Table S1, were selected from this larger set for further analysis. No tetranucleotide SSRs were represented in this set of 104 markers, presumably due to their low prevalence in the entire database.

On average, these markers amplified extremely well across the diverse panel of 24 safflower lines, with the average number of successful amplifications per primer pair being 22.5 of the 24 safflower individuals. For *C. palaestinus* and *C. oxyacanthus* the corresponding average was somewhat lower (~9.5 of 12 individuals), as might be expected given that the primer pairs were designed based on safflower sequences.

These loci harbored between two and fifteen alleles across the full panel of *Carthamus* individuals (6.0 ± 0.4; mean ± SE) and gene diversity (expected heterozygosity) ranged from 0.05 to 0.91 (0.54 ± 0.03). Across the full panel, dinucleotide SSRs were significantly more polymorphic than trinucleotide SSRs (nA and H_e both $P = 0.001$; 2-sample t test). Sixty-five of the markers (62.5%) were polymorphic in all three species. The potential future utility of these markers in safflower is evident from the high levels of polymorphism within the crop; 93 (89.4%) of the markers were polymorphic, with an average gene diversity (per polymorphic locus) of 0.40 ± 0.03 (3.9 ± 0.2 alleles per locus). PIC values ranged from 0.00 to 0.85 (mean 0.32 ± 0.02). The number of polymorphic loci was somewhat lower in *C. palaestinus* (74.0% of loci), however, gene diversity for these polymorphic loci was higher (0.42 ± 0.02). These values are based on multiple individuals drawn from a single accession, and therefore may appear quite high. It is important to note, however, that the origin of this accessions (e.g. how many plants were collected) is not given by the USDA NPGS. The two accessions of *C. oxyacanthus* exhibited a high level of polymorphism, with 89.4% polymorphic loci and gene diversity per polymorphic locus of 0.55 ± 0.02.

Relationships between *Carthamus* species and between safflower accessions

The PCO based on these genotypic data clearly shows that the three species are genetically distinct (Fig. 1). Axis one separates *C. oxyacanthus* from the other two species and accounts for 22.7% of the variation. Axis two separates safflower and *C. palaestinus* and accounts for 9.3%, consistent with the view that these two species are more closely-related to each other than to *C. oxyacanthus* (Chapman and Burke 2007).

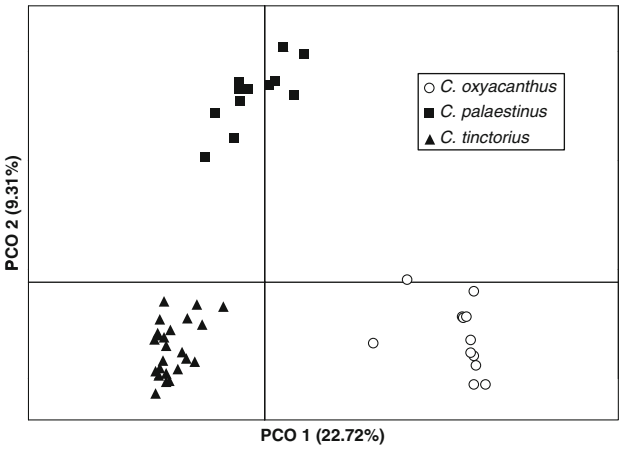

Fig. 1 Principal coordinates analysis (PCO) of the 48 individuals based on genotypic information from 104 loci. Each species is indicated by a different symbol

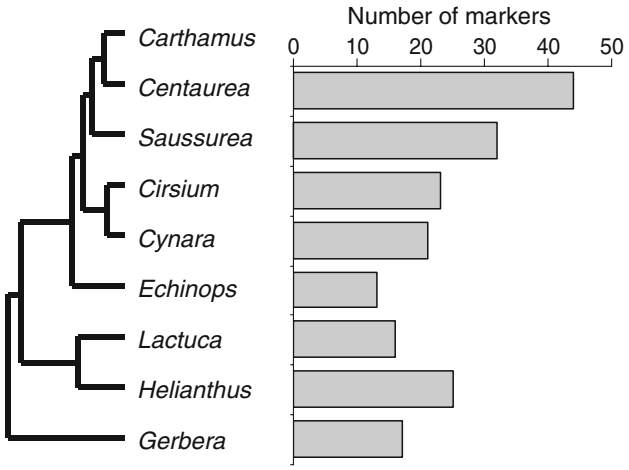

Fig. 2 Cross-species marker portability. The phylogenetic tree to the left is based on the relationships inferred by Susanna et al. (2006). The number of successful PCR amplifications (of 104 attempted) is given for five members of the Carduoideae plus three species outside this subfamily (*Helianthus*, *Lactuca* and *Gerbera*)

Amplification of the safflower SSRs in related species

The 104 SSRs were tested for amplification across a panel of five members of the Carduoideae subfamily that varied in their relatedness to the genus *Carthamus* (based on the tribal classification of Susanna et al. 2006) as well as sunflower, lettuce and *Gerbera*, which are members of the other major subfamilies within the Compositae. The number of successful amplifications in members of the Carduoideae generally followed a predictable pattern, decreasing with increased evolutionary distance (Fig. 2). Forty-four (42.3%) of the primer pairs amplified from *Centaurea*, a member of the same subtribe (the Centaurinae) as *Carthamus*, whereas only 13 (12.5%) amplified from *Echinops* (in the more distantly related subtribe Echinopinae). Slightly more primer pairs amplified from sunflower, lettuce and *Gerbera* (with an average of 19.3 primer pairs) than for the most distantly related member of the Carduoideae (*Echinops*).

Discussion

The markers described in this paper represent a valuable resource for the genetic analysis of safflower and related species. To our knowledge, this is the first set of SSR markers that has been published for safflower and our results further indicate that the safflower EST collection will be a valuable source for the development of additional SSR markers for the genetic analysis of safflower and related species. The EST-SSR markers amplified well across safflower accessions as well as other species in the section and exhibited high levels of polymorphism. As such, they should prove useful for both genetic map-based analyses as well as population genetic studies. The PIC values reported herein can serve as a guide to selecting the loci that are most likely to be informative in safflower, and possibly in other *Carthamus* species. Our PCO analysis serves as a proof-of-concept for the utility of these markers for population genetic studies within *Carthamus* and provides further support for the close relationship between safflower and *C. palaestinus* (Chapman and Burke 2007). Relationships within species are difficult to discern due to the divergent nature of the *C. oxyacanthus* individuals included in the plot.

Overall, the safflower EST collection surveyed herein contained an exceptionally high frequency of repeats, with nearly 23% of unigenes containing at least one SSR. While the apparent frequency of SSR-bearing ESTs is highly dependent on the search parameters, this value was considerably higher than expected based on data from other taxa. For example, Kantety et al. (2002) estimated that 2–5% of all plant-derived ESTs contain an SSR. More recently, Ellis and Burke (2007) searched EST collections from 33 plant species using search parameters similar to those employed here and found that an average of 9% of ESTs contain at least one SSR (range = 2.5–21.1%). Comparing to other species in the Asteraceae, safflower has approximately twice the frequency of SSR-bearing ESTs as sunflower (Heesacker et al. 2008) and approximately five times the frequency found in lettuce (Simko 2008), though the latter study employed much more stringent search criteria.

In terms of marker portability across greater evolutionary distances, several previous studies have reported reasonably high levels of transferability across taxa, with the success rate declining with relatedness (e.g., Gupta et al. 1994; Thiel et al. 2003; Pashley et al. 2006; Saha et al. 2006). Consistent with these earlier findings, many of the primer pairs described in this paper produced amplicons in other species of the Asteraceae, though success rates generally decreased with increasing evolutionary distance. For example, 44 of the 104 SSR markers (42.3%) amplified in *Centaurea*, a genus in the same subtribe as *Carthamus* which contains several important weedy species (e.g., yellow starthistle [*C. solstitialis*] and spotted knapweed [*C. maculosa*]). Moving outside the Carduoideae, ~28% of these 104 markers produced amplicons in sunflower, lettuce, and/or *Gerbera*.

In a similar study in sunflower, Heesacker et al. (2008) found somewhat lower transferability of EST-SSR markers from sunflower to safflower (14.8%) as compared to the safflower-to-sunflower success rate documented here (24.0%). While the cause of the apparently higher transfer rate in the present study is not immediately evident, it is important to note that we only selected primer pairs for

analysis that amplified across the three species of *Carthamus*. As such, the EST-SSRs analyzed in the present study may have been enriched for loci whose primer sequences are more conserved.

Given that numerous genetic maps for sunflower and lettuce have now been published (e.g., Landry et al. 1987; Kesseli et al. 1994; Gentzbittel et al. 1999; Johnson et al. 2000; Burke et al. 2002; Tang et al. 2002; Yu et al. 2003; Truco et al. 2007; Wills and Burke 2007; McHale et al. 2009), the EST-SSR resources described in this paper, along with other recently developed marker resources within the family (e.g., Chapman et al. 2007; Heesacker et al. 2008), have the potential to not only facilitate genetic map-based analyses of safflower and its close relatives, but to also enable syntenic analyses across the three major subfamilies of the Asteraceae. Such research promises to provide insights into genome evolution across this diverse family and to also aid in gene discovery across species (e.g., Gale and Devos 1998; Paterson et al. 2000).

Acknowledgments We would like to thank the greenhouse staff at UGA for maintenance of the safflower plants, the JGI for sequencing, data processing and delivery of the EST data and Huaqin Xu for maintenance of the CGPDB website. This work was supported by grants to J.M.B. from the National Science Foundation Plant Genome Research Program (DBI-0332411) and the Plant Genome Program of the USDA Cooperative State Research, Education, and Extension Service-National Research Initiative (03-35300-13104), and to the Compositae Genome Project from the USDA Initiative for Future Agriculture and Food Systems (IFAFS 2000-04292) and the NSF Plant Genome Research Program (DBI-0421630).

References

Becher R (2007) EST-derived microsatellites as a rich source of molecular markers for oats. Plant Breed 126:274–278

Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. Genetics 161:1257–1267

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

Chapman MA, Chang J-C, Weisman D, Kesseli RV, Burke JM (2007) Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). Theor Appl Genet 115:747–755

Chase MR, Moller C, Kesseli R, Bawa KS (1996) Distant gene flow in tropical trees. Nature 383:398–399

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160:1115–1123

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

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125–132

Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). Theor Appl Genet 111:456–466

Gale MD, Devos KM (1998) Plant comparative genetics after ten years. Science 282:656–658

Gentzbittel L, Mestries E, Mouzeyar S, Mazeryat F, Badaoui S, Vear F, de la Tourvieille Brouhe D, Nicolas P (1999) A composite map of expressed sequences and phenotypic traits of the sunflower (*Helianthus annuus* L.) genome. Theor Appl Genet 99:218–234

Gupta M, Chyi Y-S, Romero-Stevenson J, Owen JL (1994) Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. Theor Appl Genet 89:998–1006

Hanai LR, de Campos T, Camargo LEA, Benchimol LL, de Souza AP, Melotto M, Carbonell SAM, Chioratto AF, Consoli L, Formighieri EF, Siqueira M, Tsai SM, Vieira MLC (2007) Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. Genome 50:266–277

Heesacker A, Kishore VK, Gao WX, Tang SX, Kolkman JM, Gingle A, Matvienko M, Kozik A, Michelmore RM, Lai Z, Rieseberg LH, Knapp SJ (2008) SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. Theor Appl Genet 117:1021–1029

Johnson WC, Jackson LE, Ochoa O, van Wijk R, Peleman J, St. Clarir DA, Michelmore RW (2000) Lettuce, a shallow-rooted crop, and *Lactuca serriola*, its wild progenitor, differ at QTL determining root architecture and deep soil water exploitation. Theor Appl Genet 101:1066–1073

Johnson RC, Kisha TJ, Evans MA (2007) Characterizing safflower germplasm with AFLP molecular markers. Crop Sci 47:1728–1736

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Kesseli RV, Paran I, Michelmore RW (1994) Analysis of a detailed genetic linkage map of *Lactuca sativa* (Lettuce) constructed from RFLP and RAPD markers. Genetics 136:1435–1446

Knowles PF (1958) Safflower. Adv Agron 10:289–323

Kozik A, Michelmore RW, Knapp SJ, Matvienko MS, Rieseberg L, Lin H, van Damme M, Lavelle D, Chevalier P, Ziegle J, Ellison P, Kolkman J, Slaubaugh MB, Livingstone K, Zhou LZ, Church S, Edberg S, Jackson L, Bradford KJ (2002) Sunflower and lettuce ESTs developed by the Compositae Genome Project (http://compgenomics.ucdavis.edu/)

Lacey DJ, Wellner N, Beaudoin F, Napier JA, Shewry PR (1998) Secondary structure of oleosins in oil bodies isolated from seeds of safflower (*Carthamus tinctorius* L.) and sunflower (*Helianthus annuus* L.). Biochem J 334:469–477

Landry BS, Kesseli RV, Farrara B, Michelmore RW (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331–337

Laurent V, Devaux P, Thiel T, Viard F, Mielordt S, Touzet P, Quillet MC (2007) Comparative effectiveness of sugar beet microsatellite markers isolated from genomic libraries and GenBank ESTs to map the sugar beet genome. Theor Appl Genet 115:793–805

Levinson G, Gutman GA (1987) Slipped-strand mispairing—a major mechanism for DNA-sequence evolution. Mol Biol Evol 4:203–221

Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data. http://lewis.eeb.uconn.edu/lewis-home/software.html

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002) A single domestication for maize is shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

McHale LK, Truco MJ, Kozik A, Lavelle DO, Ochoa OE, Wroblewski T, Knapp SJ, Michelmore RW (2009) Genomic architecture of disease resistance in lettuce. Theor Appl Genet 118:565–580

Neeraja CN, Maghirang-Rodriguez R, Pamplona A, Heuer S, Collard BCY, Septiningsih EM, Vergara G, Sanchez D, Xu K, Ismail AM, Mackill DJ (2007) A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. Theor Appl Genet 115:767–776

Panero JL, Funk VA (2008) The value of sampling anomalous taxa in phylogenetic studies: major clades of the Asteraceae revealed. Mol Phylogenet Evol 47:757–782

Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST databases as a source for molecular markers: lessons from *Helianthus*. J Hered 97:381–388

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang CX, Katsar CS, Lan TH, Lin YR, Ming RG, Wright RJ (2000) Comparative genomics of plant chromosomes. Plant Cell 12:1523–1539

Peakall R, Smouse PE (2002) GenAlEx V5.04: genetic analysis in excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia. http://www.anu.edu.au/BoZo/GenAlEx/

Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, de Souza AP (2004) Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47:795–804

Saha MC, Cooper JD, Mian MAR, Chekhovskiy K, May GD (2006) Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor Appl Genet 113:1449–1458

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

Sehgal D, Raina SN (2005) Genotyping safflower (*Carthamus tinctorius*) cultivars by DNA fingerprints. Euphytica 146:67–76

Simko I (2008) Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). J Hered. doi:10.1093/jhered/esn072

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339–1348

Stevens PF (2006) Angiosperm phylogeny Website. Version 7. http://www.mobot.org/MOBOT/research/APweb

Susanna A, Garcia-Jacas N, Hidalgo O, Vilatersana R, Garnatje T (2006) The Cardueae (Compositae) revisited: Insights from ITS, trnL-trnF, and matK nuclear and chloroplast DNA analysis. Ann Mo Bot Gard 93:150–171

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1136

Tang S, Kishore VK, Knapp SJ (2003) PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. Theor Appl Genet 107:6–19

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1141–1452

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Truco MJ, Antonise R, Lavelle D, Ochoa O, Kozik A, Witsenboer H, Fort S, Jeuken MJW, Kesseli RV, Lindhout P, Michelmore RW, Peleman J (2007) A high-density, integrated genetic linkage map of lettuce (Lactuca spp.). Theor Appl Genet 115:735–746

Weiss EA (1971) Castor, sesame and safflower. Barnes and Noble, Inc, New York, NY

Wills DM, Burke JM (2007) QTL analysis of the early domestication of sunflower. Genetics 176:2589–2599

Wills DM, Hester ML, Liu A, Burke JM (2005) Chloroplast SSR polymorphisms in the Compositae and the mode of organellar inheritance in *Helianthus annuus*. Theor Appl Genet 110:941–947

Xiong LX, Liu KD, Dai XK, Xu CG, Zhang QF (1999) Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O.* rufipogon. Theor Appl Genet 98:243–251

Yi GB, Lee JM, Lee S, Choi D, Kim BD (2006) Exploitation of pepper EST-SSRs and an SSR-based linkage map. Theor Appl Genet 114:113–130

Yu JK, Tang S, Slabaugh MB, Heesacker A, Cole G, Herring M, Soper J, Han F, Chu WC, Webb DM, Thompson L, Edwards KJ, Berry S, Leon AJ, Grondona M, Olungu C, Maes N, Knapp SJ (2003) Towards a saturated molecular genetic linkage map for cultivated sunflower. Crop Sci 43:367–387

Yu JK, Dake TM, Singh S, Benscher D, Li WL, Gill B, Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805–818