Theor Appl Genet (2009) 119:1027–1038
DOI 10.1007/s00122-009-1106-2

ORIGINAL PAPER

Genetic variability in sunflower (*Helianthus annuus* L.) and in the *Helianthus* genus as assessed by retrotransposon-based molecular markers

M. Vukich · A. H. Schulman · T. Giordani · L. Natali · R. Kalendar · A. Cavallini

Received: 7 March 2009 / Accepted: 30 June 2009 / Published online: 19 July 2009

Abstract The inter-retrotransposon amplified polymorphism (IRAP) protocol was applied for the first time within the genus *Helianthus* to assess intraspecific variability based on retrotransposon sequences among 36 wild accessions and 26 cultivars of *Helianthus annuus* L., and interspecific variability among 39 species of *Helianthus*. Two groups of LTRs, one belonging to a *Copia*-like retroelement and the other to a putative retrotransposon of unknown nature (*SURE*) have been isolated, sequenced and primers were designed to obtain IRAP fingerprints. The number of polymorphic bands in *H. annuus* wild accessions is as high as in *Helianthus* species. If we assume that a polymorphic band can be related to a retrotransposon insertion, this result suggests that retrotransposon activity continued after *Helianthus* speciation. Calculation of similarity indices from binary matrices (Shannon's and Jaccard's indices) show that variability is reduced among domesticated *H. annuus*. On the contrary, similarity indices among *Helianthus* species were as large as those observed among wild *H. annuus* accessions, probably related to their scattered geographic distribution. Principal component analysis of IRAP fingerprints allows the distinction between perennial and annual *Helianthus* species especially when the *SURE* element is concerned.

Communicated by A. Bervillé.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1106-2) contains supplementary material, which is available to authorized users.

M. Vukich · T. Giordani · L. Natali · A. Cavallini (✉)
Department of Crop Plant Biology, Genetics Section,
University of Pisa, Via del Borghetto 80, 56124 Pisa, Italy
e-mail: acavalli@agr.unipi.it

A. H. Schulman · R. Kalendar
MTT/BI Plant Genomics Laboratory,
Institute of Biotechnology, University of Helsinki,
Viikki Biocenter, P.O. Box 65, 00014 Helsinki, Finland

A. H. Schulman
MTT Agrifood Research Finland, Myllytie 1,
31600 Jokioinen, Finland

Introduction

Large eukaryotic genomes are comprised mainly of transposable elements (TEs), the bulk of which are the Class I elements or retrotransposons. Most of the abundant retrotransposons are dispersed throughout the genome (Suoniemi et al. 1996; Neumann et al. 2006). Retrotransposons propagate via a "copy and paste" mechanism that resembles the replication cycle of retroviruses (Wicker et al. 2007). Of the Class I elements, the LTR retrotransposons are bounded by two long terminal repeats (LTRs), which contain the promoter and RNA processing signals. Internal to the 5′ and 3′ LTRs, respectively, are the primer binding site (PBS) and polypurine tract (PPT), which provide the signals for reverse transcription of retrotransposon transcripts into the cDNA that is reintegrated. Autonomous retrotransposons contain, between the LTRs, one or more open reading frames (ORFs) encoding the enzymatic machinery for retrotransposition (Boeke and Corces 1989; Kumar and Bennetzen 1999). The Superfamilies of LTR retrotransposons, *Copia* and *Gypsy*, respectively, are defined by the order of the enzymes within the ORFs (Wicker et al. 2007). Both Superfamilies are ubiquitous throughout the eukaryotes and appear to have been present since the divergence of the plants, animals, and fungi. Non-autonomous retrotransposons do not contain ORFs, but do have the PBS, PPTs, and LTRs needed for transcription,

replication, and integration as cDNA (Sabot and Schulman 2006). They are mobilized by parasitizing the proteins of autonomous elements (Tanskanen et al. 2007). Two major groups of non-autonomous retrotransposons have been identified, the LArge Retrotransposon Derivatives (LARDs; Kalendar et al. 2004) and the Terminal-repeat Retrotransposons In Miniature (TRIMs; Witte et al. 2001), which include the recently described *Cassandra* family (Kalendar et al. 2008). The replicative activity of retrotransposons has made them a major force in genome diversification through insertion and recombinational loss (Kalendar et al. 2000; Neumann et al. 2006; Ammiraju et al. 2007; Hawkins et al. 2008; Morse et al. 2009).

The ubiquity, abundance, dispersion, and dynamism of LTR retrotransposons in plant genomes have made them excellent sources of molecular markers (Schulman et al. 2004; Kalendar and Schulman 2006). The methods generally rely on PCR amplification between a conserved retrotransposon feature, most often the LTR, and another abundant, dispersed and conserved feature in the genome. The second site is a restriction site adapter in sequence-specific amplified polymorphism (SSAP; Waugh et al. 1997), a microsatellite in retrotransposon-microsatellite amplified polymorphism (REMAP; Kalendar et al. 1999), or another retrotransposon in inter retrotransposon amplified polymorphism (IRAP; Kalendar et al. 1999). All of these produce dominant markers, meaning that alleles at a particular locus are represented by presence or absence of a DNA fragment and that the presence of a fragment does not allow to distinguish between homozygous and heterozygous loci. In a heterozygous state, the allele in which the locus contains the retrotransposon generates a scorable product, whereas the allele lacking the retrotransposon does not. The degree of current activity, epoch of past activity, and speed of clearing from the genome all affect the phylogenetic resolution obtained from retrotransposon-based markers. In general, one may choose a retrotransposon family yielding the desired resolution. For example, in *Pisum*, the genetic diversity trees generated by the nucleotide sequence variation of 39 genes overlapped well with those produced from retrotransposon insertional polymorphisms, showing the possibility of using retrotransposon-based markers for phylogenetic purposes (Jing et al. 2007). The IRAP protocol has been applied to many plant genomes, although generally more monocots than dicots (Kalendar and Schulman 2006).

Among dicots, *Asteraceae* is the largest family, counting 25,000 species, and includes species of great economic importance for human nutrition, bioenergy production, and flower-farming. Among the *Asteraceae*, the cultivated sunflower, *Helianthus annuus,* is an important oil crop. Analyses of chloroplast DNA (Schilling 1997) dates the origin of this genus between 4.75 and 22.7 million years ago. Species within the genus diverged between 1.7 and 8.2 million years ago (Schilling 1997), i.e. relatively recently. The genus *Helianthus* likely originated in Mexico, with subsequent migration through North America. A molecular genetics study shows that modern sunflower cultivars, collected primarily in the United States, are most closely related to wild sunflower populations in the midwestern United States, supporting the hypothesis that sunflower domestication process occurred in the eastern regions of North America (Harter et al. 2004). Although controversial, a recent study has shown an earlier presence of domesticated sunflower in Mexico, based on currently available biogeographic, archaeological, linguistic, ethnohistoric, and ethnological evidence, possibly suggesting an independent domestication event in this area (Lentz et al. 2008).

Concerning the evolution of the *Helianthus* genus, the first molecular study by Schilling (1997), based on chloroplast DNA restriction analysis, subdivided the *Helianthus* genus into four sections, one for the annual *H. agrestis*, another for the annual *H. porteri*, a third (sect. *Helianthus*) containing all other annuals including *H. annuus*, and a fourth for all perennials, though limited support for the nodes between sections were found. In a subsequent work, based on ribosomal internal transcribed spacer sequences, Schilling et al. (1998) found little differentiation among most *Helianthus* species. Sossey-Alaoui et al. (1998), using RAPD technology, obtained clear-cut separations of three main sections, *Helianthus* (annuals), *Atrorubentes* and *Ciliares* (both perennials), and separate positions of *H. agrestis* and *H. porteri*. It is to be noted that the separation between species is difficult to be established because of the recent species divergence and because many species are known of hybrid origin (Rieseberg et al. 1995; Ungerer et al. 2006).

Given the activity of retrotransposons in driving genome diversification, retrotransposon-based marker methods appear attractive to be used in sunflowers. Sequences belonging to both *Gypsy* and *Copia* Superfamilies have been identified in *H. annuus* (Santini et al. 2002; Natali et al. 2006). The aim of this work was to analyse, based on application of IRAP protocol, the extent of retrotransposon-related variability in the genomes of wild and cultivated genotypes of *H. annuus* and of *Helianthus* species, especially concerning the distinction between annuals and perennials.

Materials and methods

Plant materials and DNA isolation

The genotypes used in the reported experiments are listed in Supplementary materials S1. All genotypes analyzed but

two (from Department of Crop Plant Biology, DCPB, University of Pisa, Italy) are from USDA, ARS, National Genetic Resources Program (ARS-GRIN). Further data on analyzed genotypes can be found at National Germplasm Resources Laboratory homepage (http://www.ars-grin.gov/cgi-bin/npgs/acc/query.pl). For intraspecific and interspecific fingerprinting, genomic DNA was isolated from pools of five seedlings, an approach allowing to evaluate variability among species and among wild accessions or open pollinated varieties independently from variation in single individuals. However, in other experiments, variability at the intra-population level was also investigated, isolating DNA from individual seedlings of two wild accessions of *H. annuus* (from Texas and South Dakota), *H. petiolaris* ssp. *petiolaris* (a wild, diploid, and annual species) and *H. nuttalli* ssp. *nuttalli* (a wild, diploid, and perennial species). The DNA was isolated with Nucleospin Plant Isolation kit (Macherey-Nagel) using C1 lysis buffer, which is based on the CTAB procedure. DNA was purified by RNaseA treatment. The genomic DNA was dissolved with 1× TE (1 mM EDTA, 10 mM Tris–HCl, pH 8.0) solution at 55°C. DNA was quantified using spectrophotometric analyses and DNA quality was assessed by visualization after gel electrophoresis.

Isolation of *H. annuus* LTRs

In a first set of experiments, LTR fragments were isolated by using the iPBS method (Kalendar et al. 2008; Kalendar et al., submitted). The method involves PCR amplification between the (–) strand PBS, which are highly conserved in all retrotransposons and located adjacent to the 5′ LTR. The method was carried out as described previously (Kalendar et al. 2008), in which PBS-PBS amplification was used for the isolation of *Cassandra* elements (see Supplementary materials S2). Amplified fragments were cloned into the pGEM-5Zf (Promega) plasmid T-vector and sequenced using an ABI3700 (Applied Biosystems, USA) capillary sequencer.

Retrotransposon segments within the clones were identified by a combination of comparison to known LTRs and internal retrotransposon regions from other elements, by alignment of all sequenced PCR fragments (retrotransposons are repetitive, so many copies of each are expected), and by universal structural features shared by retrotransposon LTRs. The reverse transcriptase and integrase domains of retrotransposons are sufficiently conserved that BLAST searches on the nucleotide or translated protein levels generally produce phylogenetically diverse matches, aiding in identification. The shared universal features that aid in identification of retrotransposon domains include the expectation of two TIRs terminating the LTRs and the presence of a PBS domain within two nucleotides of a TIR. The structural features of LTR retrotransposons are described by Wicker et al. (2007).

In a second set of experiments a different approach was used to isolate full-length *Copia*-like retrotransposon LTRs. A two-step PCR protocol was developed and applied (see Supplementary materials S2). The first PCR was performed using a forward primer (5′-CGAGATGA GTGCGATGGGTGAAAT-3′) designed on the *RNAse* domain of the *Copia*-like retrotransposon sequence AJ532591 (Natali et al. 2006), coupled with a randomly annealing primer (Giordani, unpublished data), rich in GC at its 3′-end (5′-ACCATCGTCCTCAGGTTAGTCAGG-3′), whose sequence but the 3′ sticky-end was randomly designed. In this approach, we searched for the 3′ LTRs, which lie close to the RNase domain.

Sequences were amplified using 20 ng of genomic DNA as template. The PCR program consisted of 28 cycles of 94°C for 30 s, 60°C for 60 s and 72°C for 60 s, using *Taq* DNA polymerase (Solis BioDyne). The amplified fragments longer than 2,000 bp were cloned into pGEM-T Easy plasmid vector (Promega). The cloned fragments were sequenced as above.

LTR sequences are highly variable (except for the promoter region and the canonical 5′ TG and 3′ AC dinucleotides at the 5′ and 3′ LTR ends) both in sequence and in length (from hundreds up to thousand nucleotides), therefore, at this stage the presence of the complete 3′ LTR within the sequenced clones could not be addressed. Retroelement 3′ LTR is usually located 2 nucleotides downstream to the PPT (Suoniemi et al. 1997). Because both LTRs are identical at the time of retroelement integration (Ma and Bennetzen 2004), primers designed on 3′ LTR are reasonably expected to match also the 5′ LTR.

Consequently, in the second PCR reaction, a forward 3′ LTR primer was designed 30 bp after the putative polypurin tract (GGGGGAG) and coupled with a reverse PBS primer (5′-TAGGTCGGAACAGGCTCTGATACCA-3′, Kalendar et al. 2008) pointing towards the 5′ LTR.

Retrotransposon amplified fragments, expected to carry 5′ LTR, were sequenced. To define the full-length LTR, a comparison between the isolated 5′ and 3′ LTRs was performed by CLUSTAL multiple alignment between the sequences obtained in the two different PCR reactions. The defined element was named *Helicopia*.

PCR protocol for IRAP

Isolated *H. annuus* LTR sequences were aligned and clustered. For each putative LTR cluster, consensus regions were selected for primer design. Primers were designed using "FastPCR" software (Kalendar, http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm) and are reported in Table 1. PCR reactions for IRAP analyses were

Table 1 List of selected primers used to generate IRAP in *Helianthus* genotypes

Primer	Sequence (5′–3′)	Primer	Sequence (5′–3′)
U81 (UF)	TAACGGTGTTCTGTTTTGCAGG	U97	ACGTCGAACTGCTGTCCGTACG
U82 (UR1)	AGAGGGGAATGTGGGGGTTTCC	U98	CAGCAGTTCGACGTGGGATCTC
U83	TCTCTATTTATAGCCGGAGAGGTG	U99	TAATCGAGCCCGTGGGCCTACA
U84	GATCCGGTTTCACGGGACTTAC	U100	TTAGTGTAGGCCCACGGGCTCG
U85	CGAAGAACAAACCGAATCACC’	U101	CGGCGTGAGAATAAGCGATTGC
U86	AGCCTCTGAAAGACTCGTTCG	U102	CATCTCAGCTGACGTCACCAGG
U87	TGTTAGCCGTTCGAGCGATCC	U103	CCCACTAGCGAGTGCGGAATCC
U88	TGGGACAGCATGTGGACCGCT	U104	AGGTTGTTCTCGATCCTCCGAG
U89 (UR2)	TTAACCAGGCTCCGGCGTGAG	U105	GAGCTGGGTATATATACCCATGC
U90	CTCTTAACGAGTAACGGTGTTCTG	U106	AAAGTACAGACACAAGTGCACC
U91	AGGTTATGGGCTGATGGGCCT	U107	TGTTGGGATTGAACCCTACCAG
U92	AAGGCCCATCAGCCCATAACC	U108	TGAATCCATTGTTGTGATCCGG
U93	GGATCGAGTATGATCTCACTGAGG	U109	ATGAAAGCCAAAACCGGATCAC
U94	CCACATACTCAGTAAGCACTAGCT	U110	TGTCACTTTCCTGGTGACGTC
U95	CTCAGTGAGATCATACTCGATCCT	CF	GGTTTAGGTTCGTAATCCTCCGCG
U96	ACTGAGTATGTGGCGGAAACAC	CR	ACAGACACCAGTGGCACCAAC

CF, CR and UF, UR1 and UR2 are the *Helicopia*- and *SURE*-LTR specific primers selected for subsequent analyses

performed in a 20 µl reaction mixture containing: 20 ng genomic DNA, 1× PCR buffer (80 mM Tris–HCl, 20 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$, 0.02% w/v Tween-20), 200 nM each primer, 200 µM each dNTP, 1U Thermostable DNA polymerase, FIREPol (Solis BioDyne). After an initial denaturing step at 95°C for 3 min, thermocycling was performed at 95°C for 20 s, 55°C for 60 s and 72°C for 60 s, for 30 cycles, final extension at 72°C for 5 min.

Each primer was tested for efficiency in the yield of IRAP bands and for fingerprinting quality (i.e., for the possibility for the amplified loci to be distinguished and scored). Genomic DNAs from ten different wild accessions of *H. annuus* were used as templates. Primers were used singularly and coupled in PCR reactions. *Helicopia* primers were designed to match the LTR ends. Only those primers generating reproducible and robust patterns were retained. The PCR products were separated by electrophoresis at 60 V for 8 h in a 1.7% agarose gel (RESolute Wide Range, BIOzym). Gels were stained with EtBr, scanned using a FLA-5100 imaging system (Fuji Photo Film GmbH., Germany) and photographed with a Canon PSA700. Each electrophoresis was repeated three times and fingerprints were scored to prepare binary matrices (Kalendar and Schulman 2006).

Sequence analyses

Helianthus annuus LTRs were aligned with CLUSTAL W software (Thompson et al. 1994). Statistics of LTR polymorphism were performed using the DNAsp program version 3.51 (Rozas and Rozas 1999). Nucleotide diversity (π, i.e. the average number of nucleotide differences per site) and its sampling variance were calculated according to Nei (1987), Eqs. 8.4 and 8.12, replacing $2n$ by n.

IRAP analyses

IRAP bands were employed for analyses and interpreted as (1) for presence or (0) absence, assuming that each band represents a single locus. Each experiment was repeated three times; non-reproducible bands were very rare and were excluded from the analyses along with weak bands. Because of huge IRAP variability among species and large number of analysed genotypes, four independent matrices (among *Helianthus* species, among wild *H. annuus* accessions, among *H. annuus* cultivars, among *H. annuus* South Dakota, *H. annuus* Texas, *H. petiolaris* and *H. nuttalli* individuals) were prepared.

Jaccard’s (1908) genetic similarity index was used to calculate genetic similarity, employing the software NTSYS (Rohlf 2000). Given two genotypes, A and B, M_{11} represents the total number of bands where they both have a value of 1, M_{01} represents the total number of bands whose values are 0 in A and 1 in B, M_{10} represents the total number of bands whose values are 1 in A and 0 in B. The Jaccard’s similarity index, J, is given as

$$J = M_{11}/M_{01} + M_{10} + M_{11}.$$

The average Jaccard’s index was calculated keeping separate data obtained on *Helianthus* species, *H. annuus* wild accessions, and *H. annuus* cultivars.

Genetic similarity was also evaluated by means of Shannon's Index (H'_j, Shannon and Weaver 1949) that, for multilocus markers is defined as:

$$H'_j = -\sum p_i \log p_i$$

where p_i is the frequency of the ith fragment in the sample. It was calculated using PopGene software version 1.32 (Yeh et al. 1999). In order to compare levels of diversity detected by different primer combinations, the various parameters were calculated for each primer combination separately. One-way ANOVA and Tukey's test were performed on genetic similarity indices at VassarStats website (http://faculty.vassar.edu/lowry/VassarStats.html).

Principal component analysis (PCA), a technique able to analyze multivariate data sets (i.e. multiple loci and multiple samples) following the algorithm of Orloci (1978), was carried out using the GenAlex6 programme (Peakall and Smouse 2006). A pairwise genetic distance matrix was calculated and converted to a covariance matrix, according to Huff et al. (1993). The GenAlex6 programme calculates the distance matrix assigning a value of 0 for characters in the same state (0,0; 1,1) and a value of 1 for characters in different states (0,1; 1,0). By means of PCA the major axis of variation are located within a multidimensional data set.

Dot-blot hybridization and calculation of sequence copy number

Helicopia partial LTR sequences were amplified using inward-oriented IRAP primer pairs. The amplification products were dissolved in a NaOH solution at a final concentration of 0.4 M, in a dilution series representing LTR copies from 2.38×10^6 to 6.15×10^8. Similarly, using a genome size estimate of 2.74 pg for the South Dakota accession and 2.91 pg for Texas accession of *H. annuus*, 3.92 pg for *H. nuttalli* ssp. *nuttalli*, and 2.66 pg for *H. petiolaris* ssp. *petiolaris* (Cavallini, unpublished), genomic DNAs were spotted on the same membranes in a dilution series from 15 to 3,840 haploid genomes.

DNA samples were incubated at 37°C for 15 min and loaded onto positively charged nylon membranes (Roche) using a Bio-Dot apparatus (Biorad). Hybridization was performed with digoxigenin-11-dUTP labeled *Helicopia* LTR as probe; probe concentration was 10 ng/ml of hybridization solution. Filter washing, detection of digoxigenin in DNA–DNA hybrids, loading control, and estimation of the copy number of the sequence probed in the samples of genomic DNA were carried out as described in Santini et al. (2002).

Results

LTR isolation

Copia-like full-length LTRs were isolated from *H. annuus* DNA using a procedure (see Supplementary materials S2) consisting in the identification of putative PPT in retrotransposon fragments obtained using standard PCR methods (see "Materials and methods") and in the design of forward LTR primers 30 bp after putative PPT on 3′-LTR. Since both LTRs are expected to be similar, the forward primer was coupled with a reverse PBS primer, close to the 5′ LTR. The cloned *Copia*-like elements (successive Genbank accessions FM177911–FM177928), which we name *Helicopia* for *Heli*anthus *copia*, has not been assigned to a family; it is therefore labelled as "RLC" (class Retrotransposon, Order LTR, Superfamily *Copia*) according to the classification system of Wicker et al. (2007).

The more general PBS amplification method (Kalendar et al., submitted), which is not specific to *Copia* elements but targets nearby or nested retrotransposons, yielded 12 PCR fragments 400 to 1,500 bp in length that contained putative LTRs (see Supplementary materials S2).

Ten of these putative LTRs were non-redundant (successive Genbank accessions FJ791038–FJ791047) and were used as query strings in BLAST searches against EST, non-redundant finished sequences, genome survey sequences, and shotgun sequence databases. The searches were circumscribed by the limited *H. annuus* sequence data publicly available; these comprised only one BAC and 133,000 ESTs (release 022009). One clone, HA83 (FJ791040), contained a region that matched both the single available BAC, accession FJ269356, as well as 12 ESTs. The likelihood of a low copy sequence matching the single 107 kb *H. annuus* sequence is very small; this suggests that the LTR of HA83 is highly repetitive. The putative LTR matched FJ269356 in an un-annotated region from nt 23,189 to 24,104, defining an LTR of 916 bp. This putative LTR (see Supplementary materials S3) terminates with the canonical TG…CA, has perfect terminal inverted repeats (TIRs) of TGTTT…AAACA and is flanked by putative target site duplications (TSDs) of TTA…TTA. The TIR, which tends to be at least 6 bp in retrotransposons, is somewhat short, as is the TSD, which tends to be 4–6 bp. Both, however, may be imperfect, a result of mutation following long residence in the genome. This interpretation is supported by the putative LTR being solo; it is flanked by TSDs but there is no adjacent PBS as would be expected of an intact retrotransposon of either autonomous or non-autonomous type. We hereby name this solo LTR *SURE* for "sunflower unidentified retroelement" and

classify it as "RLX" (class retrotransposon, order LTR, superfamily unknown) according to the scheme of Wicker et al. (2007). In the absence of an ORF associated with the *SURE* LTR, it is not possible to assign it to a retrotransposon superfamily. Of the other nine clones, seven matched only EST sequences, from six to 18 hits each. Though not matching any coding sequence, BLAST (discontiguous megablast) matches suggests that these putative LTRs are transcribed and belong to active retroelements.

Analysis of polymorphism in the two groups of LTRs (*Helicopia* and *SURE*) is reported in Table 2. Nucleotide diversity is higher for the *SURE* LTR than for *Helicopia*: since the level of diversity depends on the accumulation of mutations and hence on the time of insertion of elements, these data should indicate a more recent activity of *Helicopia* compared to that of *SURE*. However, we would need to know more about the family sizes and transcriptional and transpositional activities across the families in order to form hypotheses about differences in diversity between these groups.

Application of IRAP protocol for intraspecific and interspecific analyses

The IRAP protocol can detect genomic loci bounded by retrotransposon LTRs if elements lie close enough to be amplified by a thermostable polymerase. In IRAP, PCR products may be produced from a single primer if two similar elements are oriented head-to-tail or from two primers designed to two different head-to-head oriented LTRs. Primers designed to putative LTRs (Table 1) produced a large number of bands (see Supplementary materials S4 for an example) indicating the repetitiveness of the sequenced clones. Primers U81 (UF) coupled with U82 (UR1) and U89 (UR2), which are derived from clone HA83, produced the largest number of easily scorable bands compared to other primer combinations and to single primers (see Supplementary materials S4). These two primer pairs were used in IRAP analyses based on the *SURE* LTRs. In addition, primer pair targeting the *Helicopia* LTR (CF and CR) were used.

IRAP analyses were performed in *H. annuus* wild accessions and cultivars, and among *Helianthus* species, producing highly polymorphic fingerprints with bands ranging from 100 bp up to 3,000 bp with both *Helicopia* and *SURE* selected primer pairs (Figs. 1, 2 and 3). Nearly identical patterns were obtained in three independent experiments. However, the rare non-reproducible bands were excluded from subsequent analyses.

Helianthus annuus intraspecific variability

IRAP was analyzed on 36 wild *H. annuus* accessions (Fig. 1). Both *Helicopia* and *SURE* retrotransposon families produced highly polymorphic fingerprints, evidencing large variability in their insertion sites. The large number of bands obtained using *SURE* LTR primers further indicates that this sequence, though its nature is unknown, is highly abundant. A total of 161 bands among *H. annuus* wild accessions were scored (Table 3), all showing

Table 2 Number of sequences, number of sites (total and excluding gaps and missing sites), number of segregating sites, nucleotide diversity (π, the average number of nucleotide differences per site) and its sampling variance, in LTR sequences of the retrotransposons analyzed in this study

Retroelement	No. of sequences	No. of sites	No. of sites (excluding gaps and or missing sites)	No. of segregating sites	π	π standard deviation
Helicopia	18	454	353	142	0.1325	0.0002
SURE	12	565	130	128	0.5263	0.0055

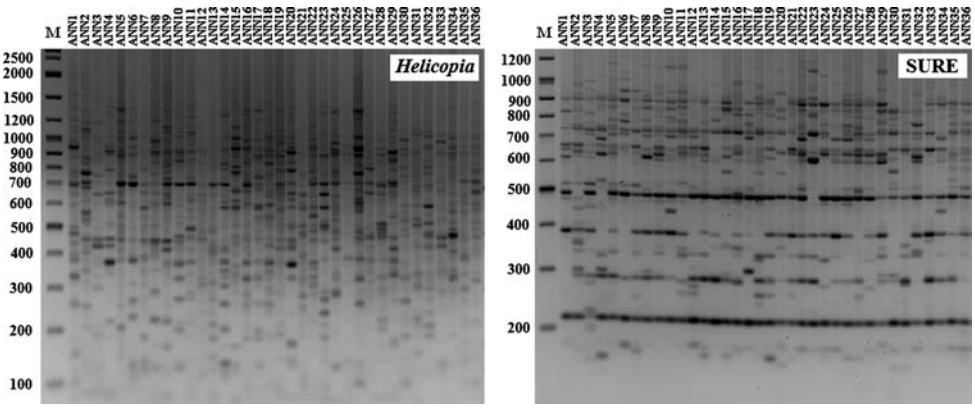

Fig. 1 IRAP fingerprints obtained with primers targeting *Helicopia* (CF-CR primers) and *SURE* LTRs (UF-UR2) in 36 wild accessions of *H. annuus*. Genotypes codes as listed in Supplementary materials S1. Molecular weight marker (M, Gene Ruler DNA Ladder Mix, Fermentas) was also loaded

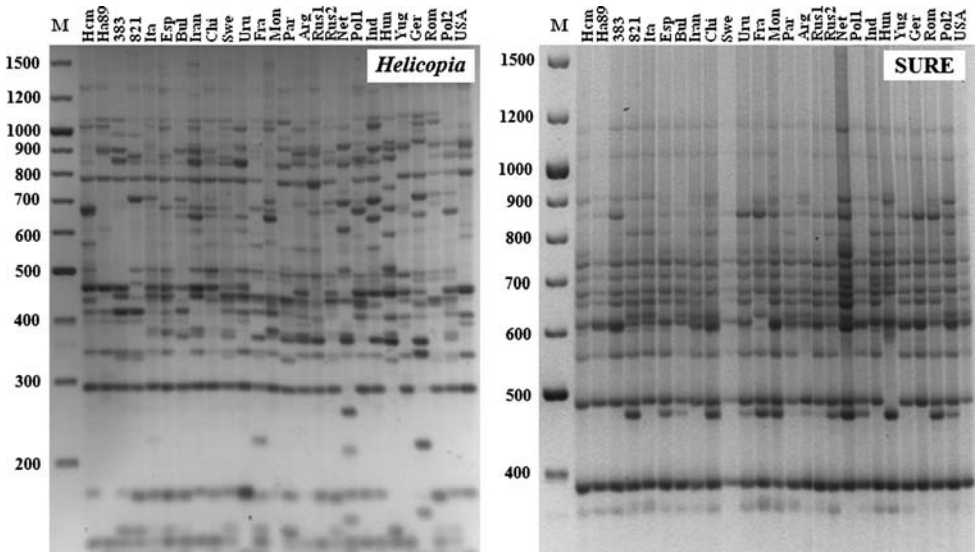

Fig. 2 IRAP fingerprints obtained with primers targeting *Helicopia* (CF-CR primers) and *SURE* LTRs (UF-UR2) in 26 *H. annuus* cultivars of different origin. Genotypes codes as listed in Supplementary materials S1. Molecular weight marker (M, Gene Ruler DNA Ladder Mix, Fermentas) was also loaded

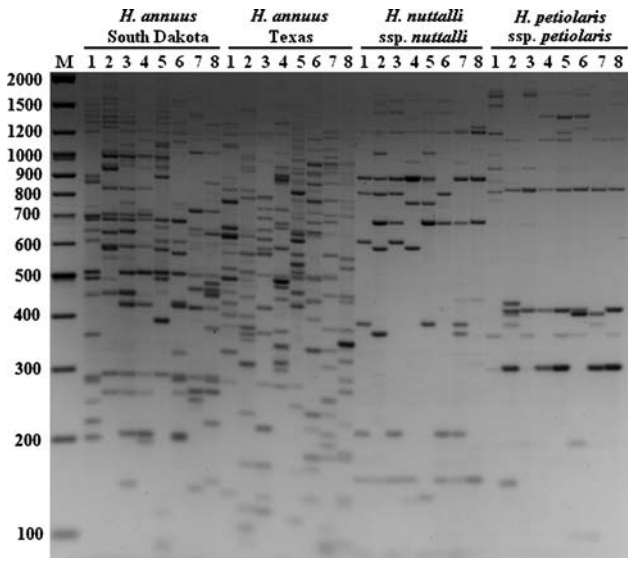

Fig. 3 IRAP fingerprints obtained with primers targeting *Helicopia* in eight individuals belonging to *H. annuus* South Dakota and Texas accessions, to *H. petiolaris* ssp. *petiolaris* and *H. nuttalli* ssp. *nuttalli*. Molecular weight marker (M, Gene Ruler DNA Ladder Mix, Fermentas) was also loaded

polymorphic. The number of polymorphic bands per primer combination ranged from 50 to 60, and was higher for the *Helicopia* primer pair, compared to both *SURE* primer pairs (Table 1). Shannon's Index for each primer combination is also reported in Table 3. This index is higher for the *Helicopia* than for the *SURE* family. The average Jaccard's Similarity Index for all primer combinations are also low (Table 3).

In other experiments, 26 cultivated genotypes of *H. annuus* were analyzed. These cultivars were randomly chosen from different countries in which sunflower is a major crop, representing a reliable sample of genetic diversity in the domesticated materials of this species. All cultivars are open pollinated; a few inbred lines were also included in the cultivar set. The IRAP protocol was applied using the same primer pairs as for wild accessions (Fig. 2). On the whole, genetic variability in domesticated genotypes was lower than in wild accessions; the total number of polymorphic bands are reduced compared to wild accessions (Table 3). Shannon's indices are reduced only for primer pairs designed to the *SURE* LTR, whereas for the *Helicopia* LTR these parameters are similar to those found for the wild genotypes (Table 3). The *Helicopia* primer pair produced twice the number of bands than each of the *SURE* primer pairs (Table 3), different from what was observed in wild *H. annuus* and in other *Helianthus* species (see below), in which the numbers of bands, percentages of polymorphic loci, and Shannon's indices were comparable. On the contrary, in *H. annuus* cultivars, the average Jaccard's Similarity Indices for each primer combination are by far the highest.

Interspecific versus intraspecific variability

The IRAP protocol was used to study genetic variability among different species of the genus *Helianthus* using the same primer pairs recognizing *Helicopia* and *SURE* LTRs as for the intraspecific analyses. The PCR reactions produced a high number of amplified fragments for all primer pairs (see Supplementary materials S5).

Fingerprints of the *Helianthus* species are characterized by more bands than those of *H. annuus* genotypes, for both *Helicopia* and *SURE* LTR primer pairs (Table 3). As among wild *H. annuus*, 100% of the bands were polymorphic. Shannon's indices are similar among the three groups of genotypes (*Helianthus* species, wild *H. annuus*, and *H. annuus* cultivars, Tables 3, 4). The average

Table 3 Number of bands, percentage of polymorphic loci, Shannon's Index, and average Jaccard's Similarity Index in 36 wild accessions and 26 cultivars of *H. annuus*, and in 32 species belonging to the *Helianthus* genus, measured for each primer pair used (CF-CR: *Helicopia*-LTR specific primers; UF-UR1 and UF-UR2: *SURE*-LTR specific primers)

	PCR primers			Mean (±SE)
	CF-CR	UF-UR1	UF-UR2	
H. annuus wild accessions				
Number of bands	60	51	50	
% Polymorphic loci	100	100	100	
Shannon's index	0.450	0.357	0.354	$0.387^{a} \pm 0.032$
Average Jaccard's similarity index	0.155	0.307	0.369	$0.277^{a,b} \pm 0.063$
H. annuus cultivars				
Number of bands	41	23	17	
% Polymorphic loci	100	82.6	64.7	
Shannon's index	0.435	0.311	0.278	$0.341^{a} \pm 0.048$
Average Jaccard's similarity index	0.532	0.737	0.405	$0.558^{b} \pm 0.097$
Helianthus species				
Number of bands	79	75	71	
% Polymorphic loci	100	100	100	
Shannon's index	0.361	0.392	0.335	$0.363^{a} \pm 0.016$
Average Jaccard's similarity index	0.111	0.106	0.109	$0.109^{a} \pm 0.001$

For Shannon's and Jaccard's indices the mean of three primer combinations are reported and two independent Tukey's tests were performed: means followed by the same letter are not significantly different at the 5% level

Jaccard's similarity indices are 0.109 (from 0 to 0.500) for interspecific fingerprints, 0.277 (from 0 to 0.714) for wild *H. annuus,* and 0.558 (from 0.222 to 1.000) for *H. annuus* cultivars (Tables 3, 4). Indeed, Jaccard's index calculated for *H. annuus* cultivars resulted significantly higher than those of wild materials (Tables 3, 4). Phylogenetic analyses using IRAP fingerprints of *Helianthus* species did not give reliable, significant phylogenetic trees (data not shown), because of large number of polymorphic, unique bands.

The comparison of genetic similarity in wild genotypes at the intraspecific and interspecific levels (Table 3) shows similar Shannon's index values. The calculation of average Jaccard's similarity indices evidence higher values for intraspecific than for interspecific comparisons, though the difference is not significant (Table 3). These data are somewhat unexpected, because genetic distances between species are usually larger than between accessions of a single species.

To analyze if such large variability among wild genotypes is at least partly accounted by variability within one genotype, the IRAP protocol was applied to eight single plants of four randomly chosen genotypes, two wild *H. annuus*, from Texas and South Dakota, and two *Helianthus* species, *H. nuttallii* and *H. petiolaris* (Fig. 3, Table 5). Though the number of bands is lower compared to the mean number of bands of wild accessions or of the *Helianthus* species, the percentages of polymorphic loci are very high in both groups, and Shannon's indices are comparable to those calculated for wild accessions or species (compare Table 5 with Table 3). Average Jaccard's similarity indices also evidence large variability, with differences between *Helicopia* and *SURE* primer combinations depending on the tested population: the index referred to *Helicopia* is lower than that to *SURE* in *H. annuus* Texas accession and in *H. nuttallii*. On the whole, these results indicate that a large portion of genetic variability is attributable to differences within populations.

Table 4 One-way ANOVA for Shannon's (above) and Jaccard's Similarity Indices (below) of three groups of genotypes (*Helianthus* species, wild *H. annuus* accessions, and *H. annuus* cultivars), calculated on the mean of three primer combinations

Source of variation	SS	Degrees of freedom	MS	*F*	*P*
Between groups	0.0031	2	0.0016	0.44	0.66^{ns}
Error	0.0213	6	0.0035		
Total	0.0244	8			
Between groups	0.3092	2	0.1546	11.54	0.0088*
Error	0.0804	6	0.0134		
Total	0.3896	8			

ns Not significant

* Significant at $P < 0.05$

Table 5 Number of bands, percentage of polymorphic loci and Shannon's Index within four genotypes of *Helianthus* (*H. annuus* acc. Texas, *H. annuus* acc. South Dakota, *H. petiolaris* ssp. *petiolaris*, and *H. nuttalli* ssp. *nuttalli*), eight individuals per genotype, measured for each primer pair used (CF-CR: *Helicopia*-LTR specific primers; UF-UR1 and UF-UR2: *SURE*-LTR specific primers)

PCR primers	CF-CR	UF-UR1	UF-UR2	CF-CR	UF-UR1	UF-UR2
H. annuus wild accessions	Texas			South Dakota		
Number of bands	61	20	22	41	29	24
% Polymorphic loci	100	95	81.8	100	100	83.3
Shannon's index	0.522	0.497	0.413	0.520	0.578	0.452
Jaccard's similarity index (mean)	0.137	0.420	0.517	0.234	0.281	0.443
Helianthus species	*H. petiolaris*			*H. nuttalli*		
Number of bands	20	20	19	19	31	15
% Polymorphic loci	95	75	73.7	100	93.5	93.3
Shannon's index	0.497	0.390	0.351	0.480	0.460	0.502
Jaccard's similarity index (mean)	0.391	0.338	0.385	0.355	0.648	0.670

The number of IRAP bands is affected both by the abundance of the retrotransposon family on which the primers are based and on the distribution of the elements or, strictly speaking, of the LTRs. To examine these factors, the copy number of the *Helicopia* LTRs was determined by dot-blot hybridization in the same genotypes. The number of *Helicopia* LTRs per haploid genome in the two wild *H. annuus* was similar, 5,742 and 5,760 copies for South Dakota and Texas accessions, respectively, and significantly higher than in *H. nuttallii* (3,720 copies) and *H. petiolaris* (4,254 copies). In parallel, the number of bands was, respectively, three times and two times higher in the Texas and South Dakota accessions than in *H. nuttallii* and *H. petiolaris*. These data suggest that, after *Helianthus* speciation, the *Helicopia* element has been more active within *H. annuus* than in other *Helianthus* species. Considering the two tested retrotransposons, the number of bands and the percentage of polymorphic bands were considerably higher using the *Helicopia* LTR than the *SURE* LTR. These two factors suggest *Helicopia* may be both younger and more active than *SURE*. Older insertions would tend both to be fixed genetically and gradually cleared from the genome by recombinational processes.

Plotting of genotypes by PCA produced no recognizable pattern among wild accessions of *H. annuus* (data not shown). On the contrary, two groups of *Helianthus* species, annuals and perennials, were clearly defined by the first, second, and third principal components, which represented 21.14% + 19.78% + 17.61% (=58.53%) of the diversity for *Helicopia* LTRs, and 23.68% + 21.19% + 15.58% (=60.45%) of the diversity for *SURE* LTRs. PCA plots of *Helianthus* species using *Helicopia*- or *SURE*-LTR fingerprints are reported in Fig. 4. The points within each group were not tightly clustered, confirming large variability among wild *H. annuus* accessions and supporting that retrotransposon activity has continued after speciation of this genus.

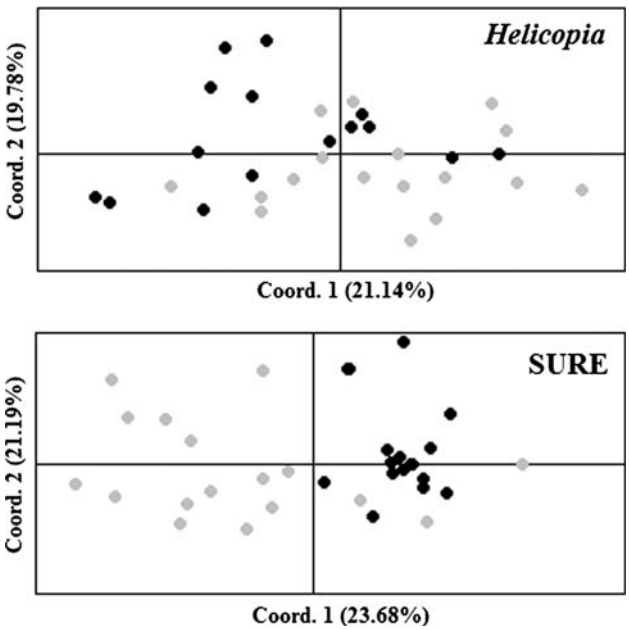

Fig. 4 PCA plot of 32 *Helianthus* species using *Helicopia*- or *SURE*-LTR fingerprints (*black dots* annual species, *grey dots* perennial species). The percentage of variation accounted by each axis is reported

Discussion

A number of LTRs were isolated in our experiments. They putatively belong to two groups of retroelements, one to a *Copia*-like retroelement (*Helicopia*) and the other to a retrotransposon of unknown nature (*SURE*). Primers designed on these LTRs allowed to evidence highly

polymorphic IRAP fingerprints in *H. annuus* wild accessions. Such variability in percentage of polymorphic loci is reduced among domesticated *H. annuus* especially considering the *SURE* element.

In addition to serving in genetic diversity and phylogenetic analyses, IRAP fingerprints can give insight into the evolution of retrotransposons and genomes such as that of the genus *Helianthus*. IRAP analyses in *H. annuus* wild accessions and cultivars produced a large number of bands with both *Helicopia* and *SURE* selected primer pairs. Recent data indicate that no predominant retrotransposon families occur in the *H. annuus* genome, with the repetitiveness of the most frequent retrotransposon family (belonging to the *Gypsy* Superfamily) being around 25,000 copies per haploid genome (Cavallini et al., submitted). Given the *H. annuus* haploid genome size of 3.2×10^9 bp (Cavallini et al. 1986), if a repetitive retrotransposon of about 10 kb were present in 25,000 copies and homogenously distributed, the individual elements would be in average 118 kb from each other, too far to be amplified by *Taq* polymerase or to be resolved in our gel system. Consequently, the occurrence of a high number of bands indicates that, in *H. annuus* as in other plant species (see SanMiguel et al. 1996), retrotransposons have a bias to form clusters.

Concerning interspecies analyses, the LTR primer pairs used in these experiments allowed amplification of numerous fragments in all species analyzed, indicating that primers are not species-specific, as expected because of the conservation of retroelement families within the *Helianthus* genus (Santini et al. 2002). This may suggest both the relatively recent divergence of this genus and the presence of both *Helicopia* and *SURE* before the radiation of its species.

Phylogenetic analyses using IRAP fingerprints of *Helianthus* species did not give reliable, significant phylogenetic trees, because of large number of polymorphic, unique bands. Phylogenetic reconstruction requires analysis of the degree of shared characters. For retrotransposons, a character is an individual insertion at a particular locus. Rigorously, this requires sequencing of polymorphic bands of similar mobility to establish the identity of the retrotransposon and its insertion site in each case. In practice, as for other fingerprinting methods such as AFLP (Meudt and Clarke 2007), bands of virtually identical mobility across many accessions are treated as shared characters. As alleles (shared bands) are rare among the analyzed accessions, assessment of identity becomes more difficult. Our failure to obtain reliable phylogenetic trees on the species level for *Helianthus* species indicates that the *SURE* and *Helicopia* retrotransposons have been turned over too rapidly for IRAP to be effective in interspecific analyses in this case.

The observed large interspecies variability in the percentage of polymorphic loci suggests that few of the ancient, pre-speciation insertions have remained fixed and/or that these two retrotransposon families have remained active throughout the *Helianthus* genus. Furthermore, the variability in copy number observed between *H. annuus*, *H. petiolaris* and *H. nuttallii* supports that propagation of these families continued after *Helianthus* species divergence, as proposed by Ungerer et al. (2006) for other retroelements in *Helianthus* species originated by interspecific hybridization. That retrotransposition is still contributing to species differentiation in this relatively young genus is also suggested by preliminary data show widespread transcription of retrotransposons in *H. annuus* (Vukich et al., in preparation).

The relative incompleteness of species differentiation within *Helianthus* is indicated by cross compatibility between *H. annuus* and annual *Helianthus* species and sometimes also between *H. annuus* and perennial species (Whelan 1978). Different degrees of genetic variability between cultivars and wild accessions of *H. annuus* was documented for microsatellites (Cheres and Knapp 1998) and protein encoding genes (Natali et al. 2003) and probably related to the time course of sunflower breeding, which started from relatively few American genotypes introduced into Europe by early Spanish explorers and in Russia by Peter the Great in the 18th century (see Natali et al. 2003). Concerning retrotransposon-related variability, the number of bands, the percentage of polymorphic loci, and the Jaccard's similarity indices in the cultivated sunflowers are reduced compared to wild accessions.

The geographic distribution of wild *H. annuus*, by far the largest of the *Helianthus* species (Rogers et al. 1982), in North America may help explain the large genetic variability of wild *H. annuus* populations. Retrotransposon distribution patterns can show eco-geographical gradients due to the underlying effect of environment and stress on retrotransposon activation (Kalendar et al. 2000). Retrotransposon-based variability in *H. annuus* could thus be related to the variation between the environments in which this widespread species lives. Because insertion or loss of retrotransposons affects the regulatory machinery of genes (Kobayashi et al. 2004), the connection between environmentally-activated retrotransposition and genetic differentiation could be of particular importance for genes involved in environmental adaptation.

Acknowledgments Research work supported by PRIN-MIUR, Italy, Project "Variabilità di sequenza ed eterosi in piante coltivate". M. V. was supported by a 6 months mobility scholarship issued by the Finnish Centre of International Mobility (CIMO).

References

Ammiraju JS, Zuccolo A, Yu Y, Song X, Piegu B, Chevalier F, Walling JG, Ma J, Talag J, Brar DS, SanMiguel PJ, Jiang N,

Jackson SA, Panaud O, Wing RA (2007) Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*. Plant J 52:342–351

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Ann Rev Microbiol 43:403–434

Cavallini A, Zolfino C, Natali L, Cionini G, Cionini PG (1986) Nuclear DNA changes within *Helianthus annuus* L.: origin and control mechanism. Theor Appl Genet 73:20–26

Cheres MT, Knapp SJ (1998) Ancestral origins and genetic diversity of cultivated sunflower: analysis of the pedigrees of public germplasm. Crop Sci 38:1476–1482

Harter AV, Gardner KA, Falush D, Lentz DL, Bye RA, Rieseberg LH (2004) Origin of extant domesticated sunflowers in eastern North America. Nature 430:201–205

Hawkins JS, Hu G, Rapp RA, Grafenberg JL, Wendel JF (2008) Phylogenetic determination of the pace of transposable element proliferation in plants: *Copia* and LINE-like elements in *Gossypium*. Genome 51:11–18

Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloë dactyloides* (Nutt.) Engelm.]. Theor Appl Genet 86:927–934

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270

Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, Ellis THN, Flavell AJ (2007) Gene-based sequence diversity analysis of field pea (*Pisum*). Genetics 177:2263–2275

Kalendar R, Schulman AH (2006) IRAP and REMAP for retrotransposon based genotyping and fingerprinting. Nature Protocols 1:2478–2484

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98:704–711

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97:6603–6607

Kalendar R, Vicient CM, Peleg O, Anamthawat-Jonsson K, Bolshoy A, Schulman AH (2004) Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. Genetics 166:1437–1450

Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O, Schulman AH (2008) Cassandra retrotransposons carry independently transcribed 5S RNA. Proc Natl Acad Sci USA 105:5833–5838

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin colour. Science 304:982

Kumar A, Bennetzen J (1999) Plant retrotransposons. Annu Rev Genet 33:479–532

Lentz DL, Pohl MD, Alvarado JL, Tarighat S, Bye R (2008) Sunflower (*Helianthus annuus* L.) as a pre-Columbian domesticate in Mexico. Proc Natl Acad Sci USA 105:6232–6237

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101:12404–12410

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends Plant Sci 12:106–117

Morse AM, Peterson DG, Islam-Faridi MN, Smith KE, Magbanua Z, Garcia SA, Kubisiak TL, Amerson HV, Carlson JE, Nelson CD, Davis JM (2009) Evolution of genome size and complexity in *Pinus*. PLoS ONE 4:e4332

Natali L, Giordani T, Cavallini A (2003) Sequence variability of a dehydrin gene within *Helianthus annuus*. Theor Appl Genet 106:811–818

Natali L, Santini S, Giordani T, Minelli S, Maestrini P, Cionini PG, Cavallini A (2006) Distribution of Ty3-*gypsy*- and Ty1-*copia*-like DNA sequences in the genus *Helianthus* and other Asteraceae. Genome 49:64–72

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY

Neumann P, Koblízková A, Navrátilová A, Macas J (2006) Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics 173:1047–1056

Orloci L (1978) Multivariate analysis in phytosociology: partition, classification and prediction. J Theor Biol 20:271–284

Peakall R, Smouse PE (2006) GENEALEX 6: genetic analysis in Excel, Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Rieseberg LH, Van Fossen C, Desrochers A (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature 375:313–316

Rogers CE, Thompson TE, Seiler GJ (1982) Sunflower species of the United States. National Sunflower Association, Bismarck, North Dakota

Rohlf FJ (2000) NTSys-pc: numerical taxonomy and multivariate analysis system Version 2.1. Exeter Software, New York

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Sabot F, Schulman AH (2006) Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97:381–388

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Santini S, Cavallini A, Natali L, Minelli S, Maggini F, Cionini PG (2002) Ty1/*Copia*- and Ty3/*Gypsy*-like DNA sequences in *Helianthus* species. Chromosoma 111:192–200

Schilling EE (1997) Phylogenetic analysis of *Helianthus* (Asteraceae) based on chloroplast DNA restriction site data. Theor Appl Genet 94:925–933

Schilling EE, Linder CR, Noyes RD, Rieseberg LH (1998) Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. Syst Bot 23:177–187

Schulman AH, Flavell AJ, Ellis THN (2004) The application of LTR retrotransposons as molecular markers in plants. Methods Mol Biol 260:145–173

Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana, IL

Sossey-Alaoui K, Serieys H, Tersac M, Lambert P, Schilling EE, Griveau Y, Kaan F, Bervillé A (1998) Evidence for several genomes in *Helianthus*. Theor Appl Genet 97:422–430

Suoniemi A, Anamthawat-Jónsson K, Arna T, Schulman AH (1996) Retrotransposon BARE-1 is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. Plant Mol Biol 30:1321–1329

Suoniemi A, Schmidt D, Schulman AH (1997) BARE-1 insertion site preferences and evolutionary conservation of RNA and cDNA processing sites. Genetica 100:219–230

Tanskanen JA, Sabot F, Vicient C, Schulman AH (2007) Life without GAG: The *BARE*-2 retrotransposon as a parasite's parasite. Gene 390:166–174

Thompson JD, Desmond G, Gibson H, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res 22:4673–4680

Ungerer MC, Strakosh SC, Zhen Y (2006) Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol 16:R872–R873

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas WTB, Powell W (1997) Genetic distribution of *BARE*-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:687–694

Whelan EDP (1978) Cytology and interspecific hybridization. In: Carter JF (ed) Sunflower Science and Technology, Am. Soc. Agronomy, Madison, Wis, pp 339–370

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell AJ, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nature Rev Genet 8:973–982

Witte CP, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 98:13778–13783

Yeh FC, Yang RC, Boyle TBJ (1999) POPGENE version 1.32, Microsoft Window-based free ware for population genetic analysis. Computer program and documentation distributed by University of Alberta and Centre for International Forestry Research, Alberta, Canada. http://www.ualberta.ca/;fyeh/index.htm