Current Plant Biology 1 (2014) 45–54

Contents lists available at ScienceDirect

Current Plant Biology

journal homepage: www.elsevier.com/locate/cpb

The repetitive component of the sunflower genome

T. Giordani, A. Cavallini, L. Natali *

Department of Agriculture, Food, and Environment, University of Pisa, Via del Borghetto 80, I-56124 Pisa, Italy

ARTICLE INFO

Article history:
Received 1 February 2014
Received in revised form 19 May 2014
Accepted 24 May 2014

Keywords:
Genome composition
Helianthus
LTR-retrotransposons
Repetitive DNA
Sunflower

ABSTRACT

The sunflower (*Helianthus annuus*) and species belonging to the genus *Helianthus* are emerging as a model species and genus for a number of studies on genome evolution. In this review, we report on the repetitive component of the *H. annuus* genome at the biochemical, molecular, cytological, and genomic levels. Recent work on sunflower genome composition is described, with emphasis on different types of repeat sequences, especially LTR-retrotransposons, of which we report on isolation, characterisation, cytological localisation, transcription, dynamics of proliferation, and comparative analyses within the genus *Helianthus*.

Contents

1. Introduction

1.1. The repetitive component of plant genomes

Eukaryotes show large variation in genome size, especially in higher plants. Angiosperm genome size (1C) ranges from 63 Mbp in *Genlisea margaretae* to 150 Gbp in *Paris japonica* [1,2]. Such differences arise from two main processes: polyploidy and amplification of transposons and related sequences. Transposon amplification has resulted in the accumulation of many repeated sequences (i.e., sequences that are identical or similar to sequences elsewhere in the genome but whose copy number is much larger than that possibly achieved through polyploidisation). Some repeats are considered to be non-functional, whereas others have played key roles in the evolution of species [3]. For example, the mutagenic action of transposons provides substantial increases in genetic variability [4]. Transposons also create novel functions by fine-tuning gene activity, resulting in phenotypic variation [5,6]. Although changes in the repetitive component have played major roles in the evolution of plant genomes, large datasets of repetitive DNA are available for such monocots as maize, rice, and barley, and a comprehensive analysis

* Corresponding author. Tel.: +39 0502216665; fax: +39 0502216661.
E-mail addresses: lucia.natali@unipi.it, lnatali@agr.unipi.it (L. Natali).

http://dx.doi.org/10.1016/j.cpb.2014.05.001

of the repetitive component is still lacking in many genera of dicots.

Among dicotyledonous species, sunflowers (*Helianthus* spp.) are emerging as a model system for genomic research on adaptation and speciation [7,8]. Given the extensive characterisation of their ecology and genomes, sunflowers are very suitable for establishing the occurrence of different types of repeats, their timing of amplification or reduction, and the role(s) of different repetitive sequences in the maintenance of a unique genome. Exploring the diversity and evolutionary dynamics of transposable elements in the sunflower (*H. annuus* L.) is of considerable importance for understanding the evolutionary history of this species, given its large genome size (3500 Mbp) [9] and the well-documented cases of retrotransposon amplification within the genus [10,11].

1.2. The Helianthus genus

The Asteraceae family is the largest plant family on Earth, with more than 24,000 described species, corresponding roughly to 10% of all angiosperms [12]. Asteraceae species live in a number of different environments, including forests, grasslands, deserts, wetlands, mountain tops, salt marshes, lawns, and agricultural fields [13]. They include economically important crops, wildflowers, valuable medicinals, invasive plants, and rangeland weeds [14]. The most important crop is the cultivated sunflower (*H. annuus* L.), which ranked 14th in 2012 among the world's food crops in terms of area harvested (http://www.fao.org/).

Helianthus includes 49 species, which are widespread in the continental United States [15], although other ecotypes are assuming the status of the species [16]. These species differ in many phenotypic traits, including reproductive timing, branching patterns, height [17,18], and especially habitat preferences. In some cases, species coexist in the same environment, and interbreeding between species is very common [19], despite large-scale karyotypic differences [20,21] and high levels of pollen and seed non-viability in the hybrids [22].

The described sunflower species, native to diverse environments throughout North America, include examples of allo- and autopolyploids [23], ecologically isolated sympatric and allopatric species [15], karyotypically divergent species [20,21,24], allopatric species with weak barriers to gene flow [15], and several homoploid hybrids [25]. Such diversity of speciation mechanisms and barriers to gene flow make sunflower an ideal model for understanding speciation and species divergence [26,27].

2. The *Helianthus* genome structure

Sunflower genomic DNA was first studied by thermal denaturation and analysis of reassociation kinetics [28,29]. Fig. 1A shows the melting profile and the first derivative curve of the DNA extracted from roots of seedlings of a selfed sunflower line. The T_m value indicates a GC content around 40%. Clear-cut shoulders were observed both on the light and the heavy sides of the derivative curves, indicating that minor specific DNA repeat families occur in the genome [28,29].

Analysis of the reassociation curves (Fig. 1B) revealed that the genome is organised into three main fractions according to their redundancy. Highly repeated (HR) sequences account for 5% of the genome, medium repeated (MR) sequences for around 60%, and the so-called unique sequences comprise the remaining 35% of the genome [28,29]. The same analyses, performed on different sunflower genotypes, showed differences in the repetitive fractions (either HR or MR), reflecting variations in the genome size [28,29].

Biochemical analyses did not consider DNA sequence but only denaturation and reassociation kinetics of DNA. Therefore, the sequence composition of the isolated fraction was not studied. Moreover, those analyses could not evaluate the occurrence in the "unique" fraction of the genome of rare forms of repeats, such as retrotransposon remnants, that were excluded from the repetitive component.

The repetitive fraction of the sunflower genome was characterised at the molecular level using a Sanger-sequenced small insert library [30]. That library provided a first set of sequences (1638, for a total of 954,517 bp) that were used, in combination with slot-blot hybridisation and fluorescent *in situ* hybridisation (FISH), to analyse the composition of the genome in terms of repeat types and

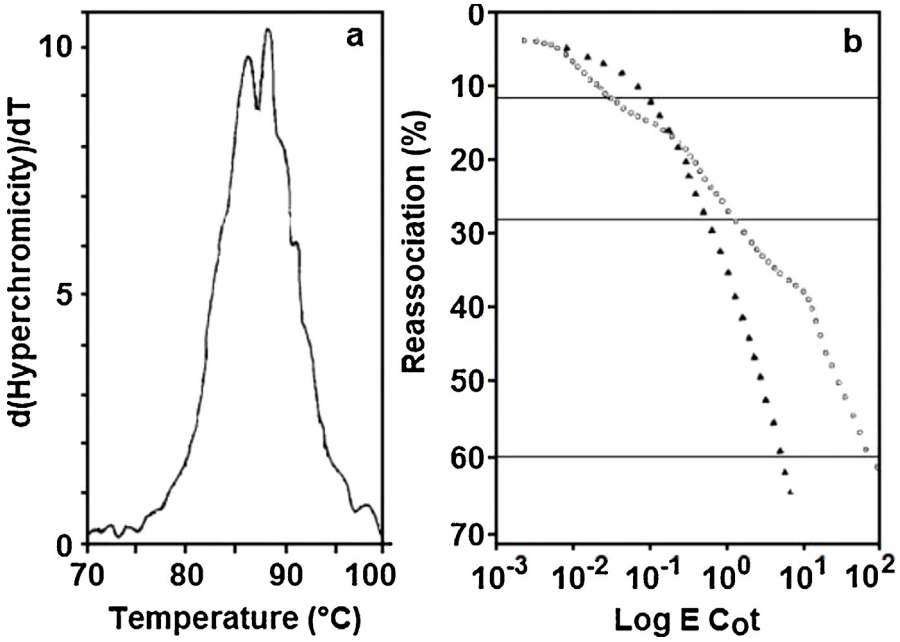

Fig. 1. (a) First derivative curves of the melting profile of sunflower genomic DNA. Small shoulders indicate specific A-T or G-C rich families of repeats. (b) Reassociation kinetics of the same DNA (circles) and of DNA of *E. coli* (triangles). Horizontal lines separate groups of sunflower sequences according to their redundancy. Redrawn from [29].

Table 1
Percentage distribution of different functional classes of repeats in the sunflower genome, based on data of the small insert Sanger-sequenced library [30] and mapping data on the Illumina plus 454 reads assemblies [40]. The observed differences are mainly related to the huge amount of sequence data published between 2010 and 2013: actually, the percentage of sequences classified as repeats by Natali et al. [40] is nearly two-fold that by Cavallini et al. [30].

Sequence type		Percentage in the small insert library [31]	Percentage in the Illumina plus 454 assemblies [39]
Class I (retrotransposons)	LTR-*Copia*	3.54	19.22
	LTR-*Gypsy*	15.57	49.22
	LTR-unknown	0.37	9.90
	Non-LTR	0.43	0.71
	Other/unclassified	–	0.48
Class II	DNA transposons	0.73	2.56
Tandem repeats	Ribosomal DNA	1.16	0.35
	Other tandem repeats	0.37	0.60
Unknown repeats		25.58	7.90
Total (classified as repeats)		47.75	90.94

abundance. About 62% of the sequences of the library belonged to the repetitive fraction, while putative functional genes accounted for 4%. The largest component was made of long terminal repeat (LTR) retrotransposons (REs), especially of the *Gypsy* superfamily.

It appeared that no transposon family was amplified to very high levels in the sunflower unlike in other species, in which specific families can account for large fractions of the genome [31–34]. All known types of repetitive elements were found in the library, although some classes (i.e., Class II transposable elements) were scarcely represented. One repeat family of unknown nature (the so-called Contig 61) was shown to be the most repeated in the sunflower genome [30].

Due to dramatic advancements of DNA sequencing technology in the past decade that have made sequencing and assembling a new genome more practical and much less expensive [35], the sunflower genome is currently being sequenced, despite its large size [36]. Next-generation sequencing technologies were used to characterise the repetitive component of the sunflower genome. Staton et al. [37] have analysed approximately 25% of the genome using 454 random sequencing and showed that the sunflower genome is composed of more than 81% transposable elements, 77% of which were LTR-REs. The LTR-REs were also studied in bacterial artificial chromosomes (BAC) clones, which resulted disproportionately composed of *Gypsy* LTR-REs containing a chromodomain ("chromoviruses") [38], with the majority of the intact chromoviruses showing chromodomain tandem duplications [37].

In other experiments [39,40], Illumina and 454 sequencing technologies were used to produce a whole genome set of sequences of sunflower by different computational assembly approaches. Mapping a large sample of Illumina reads to this set of sequences (composed of 283,300 contigs) provided a picture of sunflower genome composition [40]. Considering that Illumina reads are sampled without bias for particular sequence types, the percentage of reads that matched to a repeat class should indicate the proportion of that class in the genome. Mapping results, and, hence, sunflower genome composition, are summarised in Table 1, in which a comparison to results previously obtained by Cavallini et al. [30] is also reported. The different results obtained from analysing the small insert library are mainly related to the huge amount of sequence data published from 2010 to 2013; specifically, the percentage of sequences classified as repeats by Natali et al. [40] is nearly two-fold that by Cavallini et al. [30].

Mapping results confirmed that LTR-REs were by far the most abundant class of sequences in the sunflower genome, accounting for at least 79.53% of the reads matching the contigs. The ratio between *Gypsy* and *Copia* RE frequencies in the whole genome amounted to 2.29. This ratio is generally species-specific. The *Gypsy* to *Copia* frequency ratio is even higher in papaya (5:1) [41], sorghum (4:1) [42], and rice (3:1) [43] compared to sunflower. In other cases, such as in maize [31] and olive [44], a similar abundance of the two superfamilies was observed. In grapevine, an opposite trend was found, with *Copia* elements representing two-fold more than *Gypsy* [45].

3. Repeat types in the *Helianthus* genome

3.1. The SUNREP database

A database of repetitive sequences of sunflower was obtained by subdividing the previously described whole genome set of assembled sequences [40] according to the average coverage of each sequence. The database (SUNREP) is available at the Department of Agriculture, Food, and Environment of Pisa University website (http://www.agr.unipi.it/ricerca/plant-genetics-and-genomics-lab/sequence-repository.html).

The distribution of different sequence types in SUNREP is reported in Table 2. Around 11% of sequences did not find any hits in the public databases used for annotation. Among the annotated sequence types, REs and especially LTR-REs were clearly the most represented. Of these, elements belonging to the *Gypsy* superfamily were 2.3-fold more represented than those belonging to the *Copia* superfamily.

The assembled sequences forming SUNREP were also analysed to estimate the occurrence of families (i.e., composed of at least two sequences) within each repeat class or superfamily. The most represented family, belonging to the *Gypsy* LTR-RE superfamily, included only 96 sequences, and only four *Gypsy* families were composed of

Table 2
Functional percentage distribution of the sequences in the SUNREP database, as reported in Natali et al. [40].

Sequence type		Number (%)
Class I (retrotransposons)	Unclassified	192 (0.40)
	LTR-*Copia*	8605 (17.96)
	LTR-*Gypsy*	19,726 (41.16)
	LTR-unknown	5636 (11.76)
	Non-LTR	261 (0.54)
	Pararetrovirus	11 (0.02)
Class II (DNA transposons)	Unclassified	373 (0.78)
	Tc1 Mariner	5 (0.01)
	hAT	67 (0.14)
	Mutator	101 (0.21)
	PIF-Harbinger	18 (0.04)
	CACTA	64 (0.13)
	Helitron	324 (0.68)
	MITE	382 (0.80)
Tandem repeats	rDNA	84 (0.18)
	Other tandem repeat	385 (0.80)
Putative genes		483 (1.01)
Unknown repeats	Unclassified	4739 (9.89)
	Contig 61-type	957 (2.00)
No hits found		5511 (11.50)
Total		47,924

more than 50 sequences. Considering the 30 most numerous LTR-REs families, the vast majority belonged to the *Gypsy* superfamily. Among the 30 most numerous DNA transposons, the most common families belonged to the helitron class, followed by putative miniature inverted-repeat transposable elements (MITEs). Overall, the number of sequences comprising a family was generally low, confirming that there are not prominent repeat families in this species [30,37].

3.2. LTR-retrotransposons

The first RE sequence fragment of sunflower was isolated by PCR using the universal primers designed from the *Copia* retrotranscriptase domain [46]. The RE component of the sunflower genome was analysed in detail using two sequences isolated from a genomic DNA library, pHaS13 and pHaS211: the former representing portions of the integrase gene of a *Gypsy* retroelement and the latter corresponding to an RNase-H gene of a *Copia* retroelement [47,48]. Southern blotting patterns obtained by hybridising the two probes to restricted genomic DNA from different *Helianthus* species and from other Asteraceae species showed pHaS13 and pHaS211 were parts of dispersed repeats at least 8 and 7 kb in length, respectively.

In situ hybridisation experiments showed that sequences of both *Gypsy* and *Copia* superfamilies were dispersed throughout all chromosomes in *H. annuus* [30,47]. However, *Gypsy* sequences were localised preferentially at the centromeric regions, whereas *Copia* sequences were less represented or absent around the centromeres; only one *Copia* sequence was plentiful at the chromosome ends.

Although *Gypsy* and *Copia* LTR-REs are by far the major fraction of sunflower repetitive DNA, large numbers of sequences were identified as LTR-REs, of which the superfamily could not be determined [40,49]. Such elements, called TRIMs or LARDs [50,51], are non-autonomous, usually species-specific, and lack any distinctive protein-coding sequences.

The large number of observed REs indicates they have been actively replicating during the evolution of *H. annuus*. The large abundance of *Gypsy* elements compared to *Copia* can be explained by three hypotheses, not mutually excluding: (i) *Gypsy* elements have been more active during sunflower evolution; (ii) they have been active more recently, so that are more easily recognisable by similarity searches, having been subjected to fewer mutations; and (iii) they have been less prone to DNA removal.

To study RE dynamics in detail, it is necessary that full-length elements are available. Full-length REs have a built-in molecular clock useful for estimating their insertion times, based on sister-LTR divergence. In fact, when an RE inserts into the genome, its LTRs are usually 100% identical [52]. Mutations then occur within the two LTRs, and as more time passes since the insertion, the larger the genetic distance between LTRs. Hence, the RE insertion time can be estimated using a nucleotide substitution rate suitable for such elements that is assumed to be higher than that of gene regions [53–55].

Sunflower full-length LTR-REs were identified analysing large sequences belonging to BAC libraries [37,49,55]. In particular, Buti et al. [49] performed a fine annotation of three BAC clones, identifying 18 full-length LTR-REs. In some cases, REs formed nested structures, with one element inserted into another, similar to those commonly found in maize [53]. Analysis of the insertion time of full-length REs showed that they inserted relatively recently (during the past 3 million years). *Gypsy* elements were generally younger than *Copia*, though some *Copia* elements were relatively young, as well (Fig. 2) [49]. Staton et al. [37] confirmed and extended these findings, showing that most intact LTR-REs likely have inserted since the origin of *H. annuus*, providing further evidence that biased

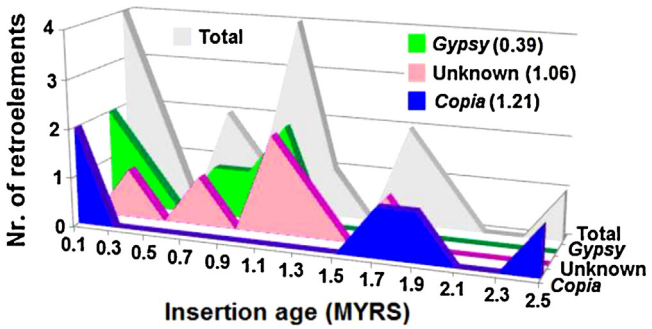

Fig. 2. Distributions of *Copia*, *Gypsy* and unknown full-length elements identified in three sequenced BAC clones according to their estimated insertion ages (millions of years, MYRS). Mean insertion age for each superfamily is reported in parentheses. Redrawn from [49].

LTR-RE activity has played a major role in shaping the DNA landscape of the sunflower genome.

That retrotransposition in sunflower has been and is probably still occurring is also indicated by recent studies showing that sunflower LTR-REs are transcriptionally active [30,56–58]. For example, RT-PCR data showed that both *Copia* and *Gypsy* REs transcription occurred in all analysed sunflower organs (embryos, leaves, roots, and flowers). Transcription was apparently not induced by environmental factors or culture conditions. Analyses made on progenies of a selfed line showed one out of 64 individuals exhibited the integration of a new element into the genome [57].

Similar results were obtained by Kawakami et al. [58]. In their transcriptional assays, multiple lineages of *Gypsy* and *Copia* elements were found to be active in natural populations of *H. annuus* and *H. petiolaris* and in their natural interspecific hybrids. In these species, transcriptional activity was not associated with copy number increases, suggesting the occurrence of strong post-transcriptional mechanisms of repression that reduce the mutagenic potentiality of RE transposition [59]. In fact, REs can inactivate genes after inserting within or nearby gene sequences, as observed in sunflower genes encoding a 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase [60] and a lipid transfer protein [49].

It is known that LTR-REs are subjected to both amplification and removal [61–63]. The RE removal is driven in plants by DNA rearrangements, illegitimate recombination, and unequal homologous recombination via a number of mechanisms, such as the repair of double-strand breaks (non-homologous end-joining) and slipstrand mispairing [33,56,64–67].

The occurrence of unequal homologous recombination typically produce the so-called solo-LTRs. Mapping Illumina reads to a set of intact LTR-REs indicated the occurrence of numerous solo-LTRs for many of the tested RE families (Fig. 3). Natali et al. [40] suggested that massive amplification of these elements in the sunflower genome was partly counterbalanced by substantial DNA loss, especially related to *Gypsy* elements. However, in other studies on sunflower repeats, solo-LTRs have been found for *Copia* elements, as well [30,37].

3.3. The so-called "Contig 61"

Clustering analysis of sequences in the short insert library [30] produced many contigs representing repetitive DNA sequences. Of these, the longest contig (Contig 61) was identified as the most repeated in the sunflower genome, with a redundancy of 27,000 copies per haploid genome, estimated by slot-blot and hybridisation.

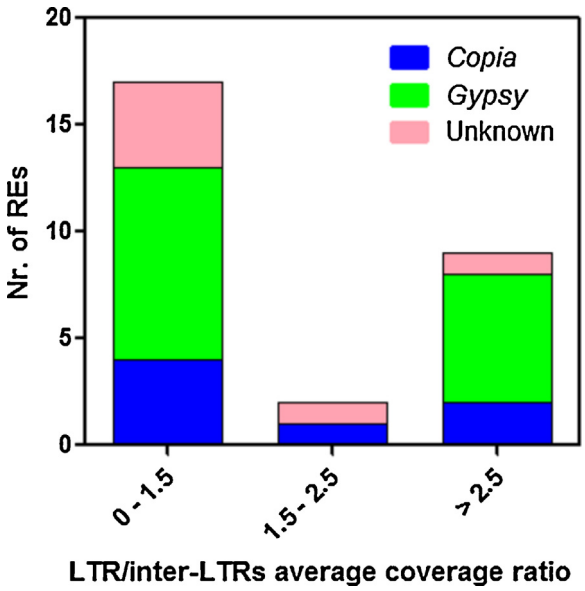

Fig. 3. Distribution of sunflower LTR-retrotransposons according to the ratio between redundancies (measured by average coverage) of LTRs and inter-LTR regions (data include 9 retrotransposons analysed by Cavallini et al. [30] and 19 analysed by Natali et al. [40]).

Many contigs sharing high sequence similarity to Contig 61 were also found in the SUNREP database. Mapping these contigs with a large set of Illumina reads established that this repeat accounts for 2.81% of the genome [40].

Southern analysis confirmed high redundancy of Contig 61 repeats, with heavy labelling and many bands. The use of isoschizomeres with different sensitivities to cytosine methylation in the target site showed that this repeated element is highly methylated [30].

When hybridising sequences belonging to Contig 61 to metaphase chromosomes of *H. annuus*, scattered labelling throughout all chromosomes was observed, indicating wide dispersal of DNA sequences [30].

The occurrence and redundancy of sequences belonging to Contig 61 were also studied in species belonging to the genus *Helianthus*, both annual and perennial, and to other Asteraceae by slot-blot hybridisation [30]. In *Helianthus*, the hybridisation signals were as strong as those observed in sunflower; in perennial species, they were stronger than in annuals. These results indicate that amplification of this sequence should have occurred in the progenitor of the *Helianthus* genus, and, after splitting between annuals and perennials, amplification has occurred in the perennial ancestor and/or loss of sequences has occurred in the annual ancestor. Regarding the other Asteraceae, some of the sequences belonging to Contig 61 also showed strong hybridisation signals to *Viguiera multiflora* genomic DNA. The genus *Viguiera* is considered the closest relative to *Helianthus*. Because Contig 61 is also apparently highly repeated in *Viguiera*, it is presumable that the initial amplification of this repeat predates the origin of the genus *Helianthus*.

Sequence analysis of Contig 61 and of related sequences obtained after Illumina and 454 assemblies revealed no structural feature that could help in the classification of this family of sequences, which, therefore, awaits further characterisation and whose nature remains unknown.

3.4. Tandem repeats

Only a few tandem repeats have been characterised in the sunflower genome. In fact, only a small number of contigs containing tandem repeats were found in SUNREP. Mapping these contigs with a large set of Illumina reads established that ribosomal DNA (rDNA) and other tandem repeats account for only 0.35% and 0.60% of the genome, respectively [40].

The rDNA was localised in chromosomes by *in situ* hybridisation [68]. Four chromosome pairs were labelled at the end of their short arm. Of these, three pairs showed secondary constrictions and strong hybridisation signals. Two other signals were punctiform at the end of the short arm of a chromosome pair where a satellite was never observed. These findings are in agreement with those obtained by Schrader et al. [69] regarding the number of satellite pairs and those that bear rDNA but contrast with their results regarding the labelling intensity, since they observed four strong and four weak signals.

Two other tandem repeats, the sequences HAG004N15 and HAG002P01, were isolated in the small insert library and characterised. HAG004N15 was 1071 bp in length and was made up of three tandemly arranged repeats having a length of 368 bp, which falls into the typical range of satellite DNAs [70]. The CAAAA or GAAAA motifs, previously associated with the amplification of repeated sequences [71,72], were found between repeats. Southern hybridisation with HAG004N15 as a probe produced the expected ladder pattern after digestion of genomic DNA with cytosine methylation-sensitive enzymes, showing that this sequence is highly methylated. Its copy number per haploid (1C) genome was estimated to be 7800.

After *in situ* hybridisation of HAG004N15 repeats [68], all the chromosomes were labelled at discrete regions (Fig. 4). Using an image analyser, the labelling was precisely localised in the metaphase chromosomes, despite their number ($2n = 34$), small size, and similar morphology (Fig. 4). The FISH signals were observed at the ends of both chromosome arms in four pairs and at the end of only one arm in eight other pairs. HAG004N15 repeats were not found in the centromeric chromosomal regions, except for one pair, where the hybridisation signals reached the centromere in the short arm, which was entirely labelled. Signals were also observed at the intercalary (mostly sub-telomeric) regions in all the pairs, in both arms in eight pairs, and in only one arm in the other nine pairs.

The chromosomal localisation of rDNA and HAG004N15 repeats formed a pattern that allowed all of the complement pairs to be distinguished from each other [68]. This hybridisation pattern was the same in different sunflower genotypes.

The redundancy of HAG004N15 and the chromosomal localisation of HAG004N15 and of rDNA were also studied in annual and perennial species of *Helianthus* and in other Asteraceae [73]. HAG004N15 sequences, which did not show amplification in other Asteraceae except *Viguiera multiflora*, were redundant in all the *Helianthus* species tested, but their frequency was significantly higher in perennials compared to annuals. These sequences were located at the ends and intercalary regions of all chromosome pairs of annual species. A similar pattern was found in the perennials, but a metacentric pair in their complement was not labelled. Ribosomal genes were carried on two chromosome pairs in perennials and on three pairs in annuals, except for *H. annuus*, where rDNA loci were on four pairs. These findings support the hypothesis that the separation between annual and perennial *Helianthus* species occurred through interspecific hybridisation involving at least one different parent [74]. However, genomic *in situ* hybridisation in *H. annuus* (using DNA from the perennial *H. giganteus* as blocking DNA) failed to reveal different genomic assets in annual and perennial species.

The other tandem repeat isolated in the small insert library [30], HAG002P01, was estimated to be present in 2600 copies per haploid genome. Using this sequence as a probe, FISH revealed an interspersed distribution with possible enhanced hybridisation at the centromeric regions of many chromosomes, indicating

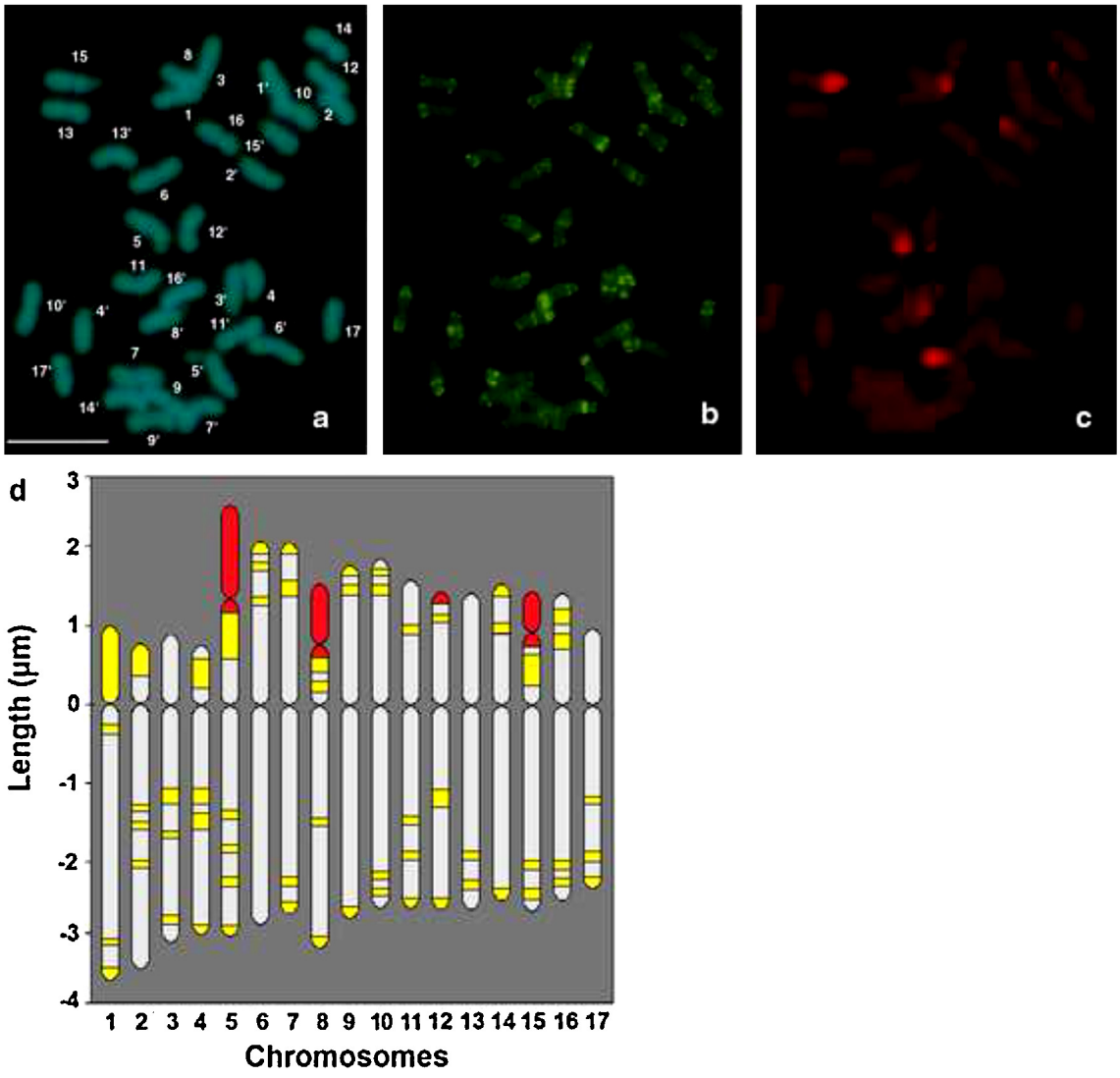

Fig. 4. (a–c) Metaphase plate of *H. annuus* line HA89 after DAPI staining (a) and hybridisation with HAG004N15 repeats (b; fluorescein) or pTa71 ribosomal probe (c; Cy3). Numerals indicate members of each chromosome pair. Bar = 10 µm (from Ceccarelli et al. [68], Characterisation of the chromosome complement of *Helianthus annuus* by *in situ* hybridisation of a tandem repeated DNA sequence. Genome, 50: 429–434, ©Canadian Science Publishing or its licensors); (d) Idiogram of the haploid chromosome complement showing the distribution of HAG004N15-related tandem repeats (yellow) and of ribosomal cistrons (red). The chromosomes are subdivided into two groups, four acrocentric and thirteen meta- to submetacentric pairs, and arranged within each group according to the total length, according to Raicu et al. [88] classification. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that HAG002P01 may have preferential paracentromeric localisation.

3.5. Other repeat types

The non-LTR-REs were poorly represented in the repetitive fraction of the sunflower genome, as frequently observed in other plants. Putative DNA transposons accounted for a number of sequences. However, a fraction of such sequences were classified as DNA transposons only by virtue of a sequence similarity to the short domain of the transposase gene. Among DNA transposons, a prevalence of MITEs [75] and helitrons [5] was observed [40].

Interestingly, according to BLAST analysis, many repetitive sequences of the SUNREP database showed similarity to putative protein-coding genes (Table 2) [40]. The most redundant gene family encodes the NBS-LRR class of proteins, receptors that recognise highly variable pathogen effectors [76]. Another redundant gene family encode DNAJ proteins, which function in association with Hsp70 molecular chaperones to facilitate protein folding and play an active role in regulating normal cellular events like protein degradation, morphogenesis, and cell cycle progression [77]. The third redundant gene family is very heterogeneous, encoding proteins with unspecified protein-kinase domains that are involved in the transduction of signals to binding factors, centromeres, and other effectors. In addition to non-specific kinases, serine/threonine/tyrosine kinases also were found among redundant gene families. Finally, another redundant gene family encoding F-box motif-containing proteins was identified by the presence of protein interaction domains that bind ubiquitination targets [78].

The occurrence of putative genes among repetitive DNA sequences of the SUNREP database deserves additional comments. In some cases, such sequences showed similarity to gene families already known to be repeated in plant genomes, such as NBS-LRR genes [76]. In other cases, it is likely that gene portions encoding functional domains, and not complete, full-length genes, are apparently redundant. For example, F-box proteins include a large variety of types that share a short motif [78]. It also could be that for

other assembled contigs, a gene (or a gene fragment) lies close to a repeated sequence, and the redundancy of that contig is related to the repeated sequence and not to the gene sequence. It should be noted that SUNREP is composed of a relatively high number of putative helitrons (Table 2) known to include exon fragments in their sequences [5], which might be partially responsible for the relatively high frequency of gene fragments in the database.

4. Retrotransposons in the evolution of the *Helianthus* genus

Southern blotting patterns obtained by hybridising the first isolated *Gypsy* and *Copia* RE fragments (pHaS13 and pHAS211, respectively) [47] to restricted genomic DNA from different *Helianthus* species and from other Asteraceae showed that these elements were conserved in all *Helianthus* species studied and, to a lesser extent, in *Tithonia rotundifolia*, a species belonging to a genus strictly close to *Helianthus* [47,48]. Comparable hybridisation patterns were obtained in all *Helianthus* species using the *Gypsy* pHaS13 as a probe. Conversely, the patterns obtained by hybridising the *Copia* pHAS211 clearly differentiated annual species from perennials [47], attesting a different amplification history of the two superfamilies of LTR-REs in the evolution of the genus *Helianthus* [30].

As observed in *H. annuus*, FISH analysis of *Gypsy* and *Copia* sequences confirmed scattered labelling throughout all metaphase chromosomes also in different species of *Helianthus* [48]. However, preferential localisation of *Gypsy* sequences at centromeric chromosome regions was observed in all of the species. Conversely, *Copia* sequences showed preferential localisation at the chromosome ends only in *H. annuus*.

The evolution of the repetitive component in the *Helianthus* genus and in other Asteraceae was studied by comparative analysis of the hybridisation of genomic DNAs isolated from these species to the sunflower small insert library (Fig. 5), which revealed some similarity only between *Helianthus* species and *Viguiera multiflora*. Regarding REs, comparable hybridisation patterns were observed among *Helianthus* species, while no hybridisation occurred using other Asteraceae. Such results indicate the specificity of RE families to the *Helianthus* genus.

An interesting example of how REs have accompanied the evolution of species within the genus *Helianthus* concerns the massive amplification of transposable elements after interspecific hybridisation occurred in three species of sunflowers, *H. anomalus*, *H. deserticola*, and *H. paradoxus* [10]. All of them originated by hybridisation between two ancestral species, *H. annuus* and *H. petiolaris* [25] and have the same chromosome number ($n = 17$) and genome sizes at least 50% larger than their parental species. Such genome size increase is partially explained by proliferation of *Gypsy* and, to a lesser extent, of *Copia* LTR-REs in hybrids [10,11].

In particular, *Gypsy* sequences were first shown to be much more redundant in the three hybrids [10]. Ungerer et al. [79] later demonstrated that *Gypsy* REs exist as multiple, well-supported sublineages in both the parental and hybrid derivative species and that these sequences underwent proliferation in each hybrid species' genome, occurring approximately 0.5–1 million years ago.

In contrast to *Gypsy*, the burst of transposition of *Copia* REs varied among hybrid species and was most pronounced in *H. paradoxus*, a species adapted to saline environments (*H. anomalus* and *H. deserticola* are found in desert-like habitats). Although external stress factors may contribute to de-repression of transposable elements [80], the association between habitat and RE frequency may be purely coincidental. *Copia* RE lineages underlying the burst are ancient and predate the origin of the sunflower group. However, the majority (70%) of sequences were derived from a single lineage of elements, which implicates a recent proliferation event [79].

5. Retrotransposons as molecular markers in sunflower

The LTR-REs in plant genomes have been employed to generate molecular markers [81,82]. In fact, the ubiquity, abundance, dispersion, and dynamism of these elements make them excellent tools to explore genetic variability even within species [83]. The methods generally rely on PCR amplification between a conserved RE feature, most often the LTR, and another abundant,

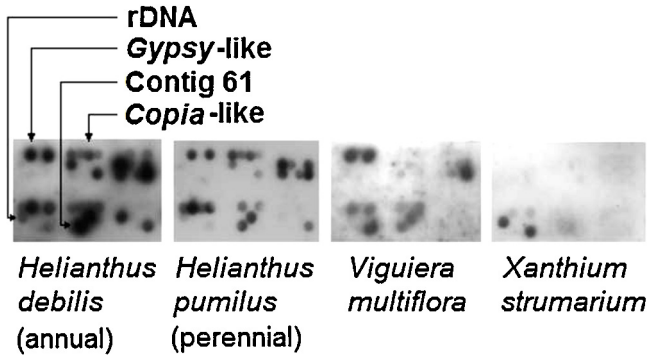

Fig. 5. Examples of hybridisation patterns of labelled genomic DNA from different species to 36 clones of the sunflower small-insert library, spotted in a non-regular duplicate arrangement. For four clones, the putative annotation is reported: for three of these, hybridisation is observed in the DNA of the two *Helianthus* species and in *Viguiera multiflora*; rDNA containing clone is labelled also in *Xanthium strumarium* DNA. For the technical procedure see [30].

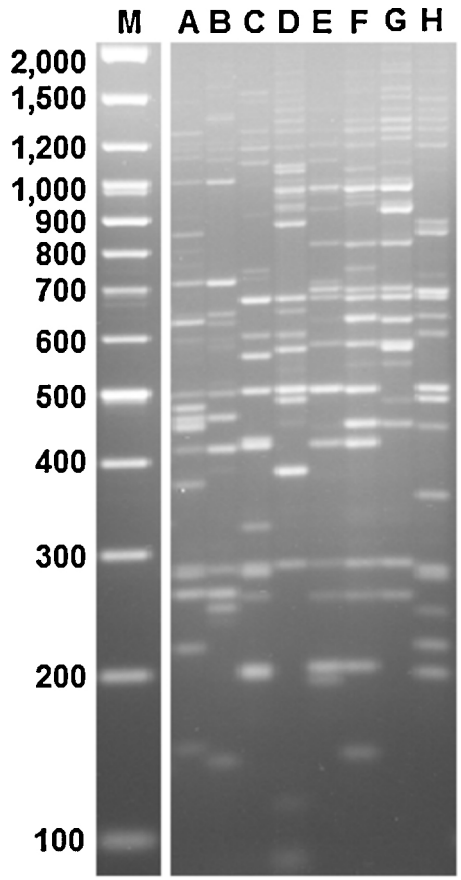

Fig. 6. IRAP fingerprints obtained with LTR-*Copia* specific primers [86] in eight individuals (A–H) belonging to a wild accession (USDA-PI 597907) of *H. annuus*. Molecular weight marker (M, Gene Ruler DNA Ladder Mix, Fermentas) was also loaded.

dispersed, conserved feature in the genome. This second site may be a restriction site adapter in sequence-specific amplified polymorphism (SSAP) [84], a microsatellite in RE-microsatellite amplified polymorphism (REMAP) [85], or another RE in inter-RE amplified polymorphism (IRAP) [85]. These methods identify genetic variants related to the insertion/loss of a retroelement or to DNA sequence variations (nucleotide substitutions or indels), which are common within REs because of the error-prone mode of retrotranscription used by these elements and extensive DNA methylation within RE-rich genome regions.

The IRAP protocol was applied within the genus *Helianthus* to assess intraspecific variability based on RE sequences among 36 wild accessions and 26 cultivars of *H. annuus* and interspecific variability among 39 species of *Helianthus* [86]. Two groups of LTRs, one belonging to a *Copia* retroelement and the other to an RE of unknown nature, were isolated and sequenced, and primers were designed to obtain IRAP fingerprints. An example of IRAP fingerprints in plants of a wild accession of *H. annuus* is reported in Fig. 6. The number of polymorphic bands and RE variability among *H. annuus* wild accessions are as high as among different *Helianthus* species. Conversely, RE-related variability was reduced among domesticated *H. annuus*.

Large RE-related variability was also found among sunflower inbred lines [87]. It was observed that between-line genetic distance correlated to heterosis in hybrids between those lines, suggesting that variations in the repetitive component of the genome, especially LTR-REs, affect the display of heterosis.

6. Conclusions and perspectives

Repetitive DNA and especially transposons apparently have a central role in the biology and evolution of plants. These sequences can also be employed to generate molecular markers that have proved efficient in distinguishing even similar genotypes.

The repetitive component of the *H. annuus* genome has been studied at the biochemical, cytological, molecular, and genomic levels. Data are now available that indicate a genome made of a large number of LTR-REs, especially of the *Gypsy* superfamily. The LTR-REs are apparently expressed, and the RE burst seems to be unexhausted as shown by the relatively recent insertion times of many REs. It is also worth noting the large number of LTR-REs of unknown nature, which have been (and presumably are) active because of enzymes produced by intact retroelements.

The consequences of such RE mobilisation are evident when observing the evolution of *Helianthus* interspecific hybrids that probably have completed their speciation because of chromosomal differentiation from parental species due to insertions of transposable elements.

Many data are also available on other repeated sequences, such as helitrons, whose redundancy is relatively high compared to that found in other species. All these data can be combined to produce a collection of sequences that will be useful in the annotation of the sunflower genome sequence when it is complete. Until then, the enormous mass of available data will be useful for comparative studies on genome evolution in a genus (*Helianthus*) that has been studied at the evolutionary and ecological levels and especially for its protein-coding components. The completion of the *H. annuus* whole genome sequence in conjunction with the availability of new sequencing techniques allowing the production of very large DNA sequences (i.e., Pac-Bio technology) will be useful for comparative studies among *Helianthus* species and accessions and for clarifying many aspects of genome evolution in this genus.

Acknowledgements

This work was supported by PRIN-MIUR, Italy, Project "SUNREP: caratterizzazione molecolare della componente ripetitiva del genoma di girasole".

References

[1] J. Greilhuber, T. Borsch, K. Müller, A. Worberg, S. Porembski, W. Barthlott, Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size, Plant Biol. 8 (2006) 770–777.
[2] J. Pellicer, M.F. Fay, I.J. Leitch, The largest eukaryotic genome of them all? Bot. J. Linnean Soc. 164 (2010) 10–15.
[3] R.J. Britten, Transposable element insertions have strongly affected human evolution, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 19945–19948.
[4] M. Morgante, E. De Paoli, S. Radovic, Transposable elements and the plant pangenomes, Curr. Opin. Plant Biol. 10 (2007) 149–155.
[5] M. Morgante, S. Brunner, G. Pea, K. Fengler, A. Zuccolo, A. Rafalski, Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize, Nat. Genet. 37 (2005) 997–1002.
[6] R.K. Slotkin, R. Martienssen, Transposable elements and the epigenetic regulation of the genome, Nat. Rev. Genet. 8 (2007) 272–285.
[7] D.A. Rasmussen, M.A.F. Noor, What can you do with 0.1× genome coverage? A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae), BMC Genomics 10 (2009) 382.
[8] R.M. Lee, J. Thimmapuram, K.A. Thinglum, G. Gong, A.G. Hernandez, C.L. Wright, et al., Sampling the waterhemp (*Amaranthus tuberculatus*) genome using pyrosequencing technology, Weed Sci. 57 (2009) 463–469.
[9] E.J. Baack, K.D. Whitney, L.H. Rieseberg, Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species, New Phytol. 167 (2005) 623–630.
[10] M.C. Ungerer, S.C. Strakosh, Y. Zhen, Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation, Curr. Biol. 16 (2006) R872–R873.
[11] T. Kawakami, S.C. Strakosh, Y. Zhen, M.C. Ungerer, Different scales of Ty1/*copia* like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species, Heredity 104 (2010) 341–350.
[12] P. Stevens, Angiosperm Phylogeny, 2010, Available from http://www.Mobot.Org/mobot/research/apweb/
[13] V. Funk, A. Susanna, T. Stuessy, R. Bayer, Systematics, Evolution and Biogeography of the Compositae, International Association of Plant Taxonomy, Vienna, Austria, 2009.
[14] H. Dempewolf, L.H. Rieseberg, Q.C. Cronk, Crop domestication in the Compositae: a family-wide trait assessment, Genet. Resour. Crop Evol. 55 (2008) 1141–1157.
[15] C.B. Heiser, D.M. Smith, S. Clegenger, W.C. Martin, The North American sunflowers (*Helianthus*), Mem. Torrey Bot. Club 22 (1969) 1–218.
[16] B.T. Moyers, L.H. Rieseberg, Divergence in gene expression is uncoupled from divergence in coding sequence in a secondarily woody sunflower, Int. J. Plant Sci. 174 (2013) 1079–1089.
[17] D.M. Rosenthal, A.E. Schwarzbach, L.A. Donovan, O. Raymond, L.H. Rieseberg, Phenotypic differentiation between three ancient hybrid taxa and their parental species, Int. J. Plant Sci. 163 (2002) 387–398.
[18] D.M. Rosenthal, L.H. Rieseberg, L.A. Donovan, Re-creating ancient hybrid species' complex phenotypes from early-generation synthetic hybrids: three examples using wild sunflowers, Am. Nat. 166 (2005) 26–41.
[19] L.H. Rieseberg, S.J.E. Baird, A.M. Desrochers, Patterns of mating in wild sunflower hybrid zones, Evolution 52 (1998) 713–726.
[20] J.M. Burke, Z. Lai, M. Salmaso, T. Nakazato, S. Tang, A. Heesacker, et al., Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*, Genetics 167 (2004) 449–457.
[21] Z. Lai, T. Nakazato, M. Salmaso, J.M. Burke, S. Tang, S.J. Knapp, et al., Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species, Genetics 171 (2005) 291–303.
[22] L.H. Rieseberg, M.J. Kim, G.J. Seiler, Introgression between the cultivated sunflower and a sympatric wild relative, *Helianthus petiolaris* (Asteraceae), Int. J. Plant Sci. 160 (1999) 102–108.
[23] R.E. Timme, B.B. Simpson, R.C. Linder, High-resolution phylogeny for *Helianthus* (Asteraceae) using the 18S-26S ribosomal DNA external transcribed spacer, Am. J. Bot. 94 (2007) 1837–1852.
[24] J.M. Chandler, C.C. Jan, B.H. Beard, Chromosomal differentiation among the annual *Helianthus* species, Syst. Bot. 11 (1986) 354–371.
[25] L.H. Rieseberg, Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes, Am. J. Bot. 78 (1991) 1218–1237.
[26] L.H. Rieseberg, C.R. Linder, G.J. Seiler, Chromosomal and genic barriers to introgression in *Helianthus*, Genetics 141 (1995) 1163–1171.
[27] N.C. Kane, J.M. Burke, L. Marek, G. Seiler, F. Vear, G. Baute, S.J. Knapp, et al., Sunflower genetic, genomic and ecological resources, Mol. Ecol. Res. 13 (2013) 10–20.
[28] A. Cavallini, C. Zolfino, G. Cionini, R. Cremonini, L. Natali, O. Sassoli, et al., Nuclear DNA changes within *Helianthus annuus* L.: cytophotometric, karyological and biochemical analyses, Theor. Appl. Genet. 73 (1986) 20–26.

[29] L. Natali, A. Cavallini, G. Cionini, O. Sassoli, P.G. Cionini, M. Durante, Nuclear DNA changes within *Helianthus annuus* L.: changes within single progenies and their relationships with plant development, Theor. Appl. Genet. 85 (1993) 506–512.
[30] A. Cavallini, L. Natali, A. Zuccolo, T. Giordani, I. Jurman, V. Ferrillo, et al., Analysis of transposons and repeat composition of the sunflower (*Helianthus annuus* L.) genome, Theor. Appl. Genet. 120 (2010) 491–508.
[31] B.C. Meyers, S.V. Tingey, M. Morgante, Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome, Genome Res. 11 (2001) 1660–1676.
[32] C.M. Vicient, R. Kalendar, A.H. Schulman, Variability, recombination, and mosaic evolution of the barley BARE-1 retrotransposon, J. Mol. Evol. 61 (2005) 275–291.
[33] P. Neumann, A. Koblizkova, A. Navratilova, J. Macas, Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement, Genetics 173 (2006) 1047–1056.
[34] C. Vitte, J.L. Bennetzen, Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 17638–17643.
[35] E.R. Mardis, The impact of next-generation sequencing technology on genetics, Trends Genet. 24 (2008) 133–141.
[36] N.C. Kane, N. Gill, M.G. King, J.E. Bowers, H. Berges, J. Gouzy, et al., Progress towards a reference genome for sunflower, Botany 89 (2011) 429–437.
[37] S.E. Staton, B.H. Bakken, B.K. Blackman, M.A. Chapman, N.C. Kane, S. Tang, et al., The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements, Plant J. 72 (2012) 142–153.
[38] D. Kordis, A genomic perspective on the chromodomain-containing retrotransposons: chromoviruses, Gene 347 (2005) 161–173.
[39] R.M. Cossu, L. Natali, E. Barghini, T. Giordani, M. Buti, F. Mascagni, et al., Using different procedures for assembling next generation sequencing reads allows to obtain a complete description of the repetitive component of the sunflower genome, in: Plant Genome Evolution 2013 – A Current Opinion Conference, Amsterdam, 2013, Abstract Book P040.
[40] L. Natali, R.M. Cossu, E. Barghini, T. Giordani, M. Buti, F. Mascagni, M. Morgante, et al., The repetitive component of the sunflower genome as revealed by different procedures for assembling next generation sequencing reads, BMC Genomics 14 (2013) 686.
[41] R. Ming, S. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte, H. Albert, et al., The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus), Nature 452 (2008) 991–997.
[42] A.H. Paterson, J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, et al., The *Sorghum bicolor* genome and the diversification of grasses, Nature 457 (2009) 551–556.
[43] The International Rice Genome Sequencing Project, The map based sequence of the rice genome, Nature 436 (2005) 793–800.
[44] E. Barghini, L. Natali, R.M. Cossu, T. Giordani, M. Pindo, F. Cattonaro, et al., The peculiar landscape of repetitive sequences in the olive (*Olea europaea* L.) genome, Genome Biol. Evol. 6 (2014) 776–791.
[45] O. Jaillon, J.M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, et al., The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla, Nature 449 (2007) 463–467.
[46] D.F. Voytas, M.P. Cummings, A. Konieczny, F.M. Ausubel, S.R. Rodermel, *Copia*-like retrotransposons are ubiquitous among plants, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 7124–7128.
[47] S. Santini, A. Cavallini, L. Natali, S. Minelli, F. Maggini, P.G. Cionini, Ty1/*copia*- and Ty3/*gypsy*-like DNA sequences in *Helianthus* species, Chromosoma 111 (2002) 192–200.
[48] L. Natali, S. Santini, T. Giordani, S. Minelli, P. Maestrini, P.G. Cionini, et al., Distribution of Ty3-*gypsy*- and Ty1-*copia*-like DNA sequences in the genus *Helianthus* and other Asteraceae, Genome 49 (2006) 64–72.
[49] M. Buti, T. Giordani, F. Cattonaro, R.M. Cossu, L. Pistelli, M. Vukich, et al., Temporal dynamics in the evolution of the sunflower genome as revealed by sequencing and annotation of three large genomic regions, Theor. Appl. Genet. 123 (2011) 779–791.
[50] C.P. Witte, Q.H. Le, T. Bureau, A. Kumar, Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 13778–13783.
[51] R. Kalendar, C.M. Vicient, O. Peleg, K. Anamthawat-Jonsson, A. Bolshoyb, A.H. Schulman, Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes, Genetics 166 (2004) 1437–1450.
[52] A. Kumar, J.L. Bennetzen, Plant retrotransposons, Annu. Rev. Genet. 33 (1999) 479–532.
[53] P. SanMiguel, A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, et al., Nested retrotransposons in the intergenic regions of the maize genome, Science 274 (1996) 765–768.
[54] P. SanMiguel, J.L. Bennetzen, Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons, Ann. Bot. 82 (1998) 37–44.
[55] M. Buti, T. Giordani, M. Vukich, L. Gentzbittel, L. Pistelli, F. Cattonaro, et al., HACRE1, a recently inserted *copia*-like retrotransposon of sunflower (*Helianthus annuus* L.), Genome 11 (2009) 904–911.
[56] J. Ma, J.L. Bennetzen, Rapid recent growth and divergence of rice nuclear genomes, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 12404–12410.
[57] M. Vukich, T. Giordani, L. Natali, A. Cavallini, *Copia* and *Gypsy* retrotransposons activity in sunflower (*Helianthus annuus* L.), BMC Plant Biol. 9 (2009) 150.
[58] T. Kawakami, P. Dhakal, A.N. Katterhenry, C.A. Heatherington, M.C. Ungerer, Transposable element proliferation and genome expansion are rare in contemporary sunflower hybrid populations despite widespread transcriptional activity of LTR retrotransposons, Genome Biol. Evol. 3 (2011) 156–167.
[59] D. Lisch, Epigenetic regulation of transposable elements in plants, Ann. Rev. Plant Biol. 60 (2009) 43–66.
[60] S. Tang, C.G. Hass, S.J. Knapp, Ty3/*gypsy*-like retrotransposon knockout of a 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase is non-lethal, uncovers a cryptic paralogous mutation, and produces novel tocopherol (vitamin E) profiles in sunflower, Theor. Appl. Genet. 113 (2006) 783–799.
[61] K.M. Devos, J.K.M. Brown, J.L. Bennetzen, Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis, Genome Res. 12 (2002) 1075–1079.
[62] J. Ma, K.M. Devos, J.L. Bennetzen, Analyses of LTR retrotransposon structures reveal recent and rapid genomic DNA loss in rice, Genome Res. 14 (2004) 860–869.
[63] C. Grover, J. Hawkins, J. Wendel, Phylogenetic insights into the pace and pattern of plant genome size evolution, in: J.N. Volff (Ed.), Plant Genomes. Genome Dynamics, vol. 4, Karger, Basel, CH, 2008, pp. 57–68.
[64] R. Kalendar, J. Tanskanen, S. Immonen, E. Nevo, A.H. Schulman, Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 6603–6607.
[65] J.S. Ammiraju, A. Zuccolo, Y. Yu, X. Song, P. Piegu, F. Chevalier, et al., Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*, Plant J. 52 (2007) 342–351.
[66] J.S. Hawkins, G. Hu, R.A. Rapp, J.L. Grafenberg, J.F. Wendel, Phylogenetic determination of the pace of transposable element proliferation in plants: *Copia* and LINE-like elements in *Gossypium*, Genome 51 (2008) 11–18.
[67] A.M. Morse, D.G. Peterson, M.N. Islam-Faridi, K.E. Smith, Z. Magbanua, et al., Evolution of genome size and complexity in *Pinus*, PLoS ONE 4 (2009) e4332.
[68] M. Ceccarelli, V. Sarri, L. Natali, T. Giordani, A. Cavallini, A. Zuccolo, et al., Characterization of the chromosome complement of *Helianthus annuus* by *in situ* hybridization of a tandemly repeated DNA sequence, Genome 50 (2007) 429–434.
[69] O. Schrader, R. Ahne, J. Fuchs, I. Schubert, Karyotype analysis of *Helianthus annuus* using Giemsa banding and fluorescence *in situ* hybridization, Chromosome Res. 5 (1997) 451–456.
[70] J. Macas, T. Meszaros, M. Nouzova, PlantSat: a specialized database for plant satellite repeats, Bioinformatics 18 (2002) 28–35.
[71] R. Appels, L.B. Moran, J.P. Gustafson, Rye heterochromatin: studies on clusters of major repeating sequences and the identification of a new dispersed repetitive sequence element, Can. J. Genet. Cytol. 28 (1986) 3389–3402.
[72] A. Katsiotis, M. Hagidimitriov, A. Douka, P. Hatzopolus, Genomic organization, sequence interrelationship and physical localization using *in situ* hybridization of tandemly repeated DNA sequences in the genus *Olea*, Genome 41 (1998) 527–534.
[73] L. Natali, M. Ceccarelli, T. Giordani, V. Sarri, A. Zuccolo, I. Jurman, et al., Phylogenetic relationships between annual and perennial species of *Helianthus*: evolution of a tandem repeated DNA sequence and cytological hybridization experiments, Genome 51 (2008) 1047–1053.
[74] K. Sossey-Alaoui, H. Serieys, M. Tersac, P. Lambert, E.E. Schilling, Y. Griveau, et al., Evidence for several genomes in *Helianthus*, Theor. Appl. Genet. 97 (1998) 422–430.
[75] S.R. Wessler, T.E. Bureau, S.E. White, LTR-retrotransposons and MITEs: important players in the evolution of plant genomes, Curr. Opin. Genet. Dev. 5 (1995) 814–821.
[76] L. McHale, X. Tan, P. Koehl, R.W. Michelmore, Plant NBS-LRR proteins: adaptable guards, Genome Biol. 7 (2006) 212.
[77] G. Frugis, G. Mele, D. Giannino, D. Mariotti, MsJ1, an alfalfa DnaJ-like gene, is tissue-specific and transcriptionally regulated during cell cycle, Plant Mol. Biol. 40 (1999) 397–408.
[78] J. Jin, T. Cardozo, R.C. Lovering, S.J. Elledge, M. Pagano, J.W. Harper, Systematic analysis and nomenclature of mammalian F-box proteins, Genes Dev. 18 (2004) 2573–2580.
[79] M.C. Ungerer, S.C. Strakosh, K.M. Stimpson, Proliferation of Ty3/*gypsy*-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data, BMC Biol. 7 (2009) 40.
[80] M.A. Grandbastien, Activation of plant retrotransposons under stress conditions, Trends Plant Sci. 3 (1998) 181–187.
[81] A.H. Schulman, A.J. Flavell, T.H.N. Ellis, The application of LTR retrotransposons as molecular markers in plants, Methods Mol. Biol. 260 (2004) 145–173.
[82] R. Kalendar, A.H. Schulman, IRAP and REMAP for retrotransposon based genotyping and fingerprinting, Nat. Protoc. 1 (2006) 2478–2484.
[83] C. D'Onofrio, G. De Lorenzis, T. Giordani, L. Natali, A. Cavallini, G. Scalabrelli, Retrotransposon-based molecular markers for grapevine species and cultivars identification, Tree Genet. Genomes 6 (2010) 451–466.
[84] R. Waugh, K. McLean, A.J. Flavell, S.R. Pearce, A. Kumar, W.T.B. Thomas, et al., Genetic distribution of BARE-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP), Mol. Gen. Genet. 253 (1997) 687–694.

[85] R. Kalendar, T. Grob, M. Regina, A. Suoniemi, A.H. Schulman, IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques, Theor. Appl. Genet. 98 (1999) 704–711.
[86] M. Vukich, A.H. Schulman, T. Giordani, L. Natali, R. Kalendar, A. Cavallini, Genetic variability in sunflower (*Helianthus annuus* L.) and in the *Helianthus* genus as assessed by retrotransposon based molecular markers, Theor. Appl. Genet. 119 (2009) 1027–1038.
[87] M. Buti, T. Giordani, M. Vukich, C. Pugliesi, L. Natali, A. Cavallini, Retrotransposon-related genetic distance and hybrid performance in sunflower (*Helianthus annuus* L.), Euphytica 192 (2013) 289–303.
[88] P. Raicu, V. Vranceanu, A. Mihailescu, C. Popescu, M. Kirillova Motz, Research of the chromosome complement in *Helianthus* L. genus, Caryologia 29 (1976) 307–316.