REVIEW

Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function

CrossMark

Shweta Mehrotra *, **Vinod Goyal**

Department of Botany, University of Delhi, Delhi 110007, India

Abstract Repetitive DNA sequences are a major component of eukaryotic genomes and may account for up to 90% of the genome size. They can be divided into minisatellite, microsatellite and satellite sequences. Satellite DNA sequences are considered to be a fast-evolving component of eukaryotic genomes, comprising tandemly-arrayed, highly-repetitive and highly-conserved monomer sequences. The monomer unit of satellite DNA is 150–400 base pairs (bp) in length. Repetitive sequences may be species- or genus-specific, and may be centromeric or subtelomeric in nature. They exhibit cohesive and concerted evolution caused by molecular drive, leading to high sequence homogeneity. Repetitive sequences accumulate variations in sequence and copy number during evolution, hence they are important tools for taxonomic and phylogenetic studies, and are known as "tuning knobs" in the evolution. Therefore, knowledge of repetitive sequences assists our understanding of the organization, evolution and behavior of eukaryotic genomes. Repetitive sequences have cytoplasmic, cellular and developmental effects and play a role in chromosomal recombination. In the post-genomics era, with the introduction of next-generation sequencing technology, it is possible to evaluate complex genomes for analyzing repetitive sequences and deciphering the yet unknown functional potential of repetitive sequences.

Introduction

Genomes of higher eukaryotes contain more DNA than expected when estimates are based on the length and number of coding genes in the genomes. The amount of DNA in the unreplicated genome, or the haploid genome, of a species is known as C-value or Constant-value [1,2]. The lack of

correlation between size and complexity of eukaryotic genomes, largely due to the presence of noncoding highly repetitive DNA, is termed as the C-value paradox, which is a common phenomenon observed in higher plants. It is believed that the proportion of protein-coding sequences is generally similar for different plant species, with variation in genome size mainly due to the presence of repetitive DNA [3,4] that has accumulated in the genomes during evolution, since ancestral angiosperms had been indicated to possess small genomes [5]. The term "repetitive sequences" refers to homologous DNA fragments that are present in multiple copies in the genome. Repetitive DNA sequences are present in all higher plants and can account for

* Corresponding author.
  E-mail: shwetamehrotra@rediffmail.com (Mehrotra S).

up to 90% of the genome size in some species. These repetitive DNA sequences are considered to generate major differences between genomes, which may reflect evolutionary distances between species. The repetitive sequences were once thought to be "selfish elements" or "junk DNA" [6,7], since they do not harbor genes.

Knowledge of the distribution, genomic organization, chromosomal location and evolutionary origin of repetitive DNA sequences is necessary for insight into the organization, evolution, behavior and functional potential of repetitive sequences in eukaryotic genomes [8]. In the last few decades, several repetitive sequences have been analyzed (Table 1) to gain more information on primary structure, molecular organization, evolution and function of repetitive sequences, whereas more recently genome-sequencing technology has produced an unprecedented wealth of information about origin, diversity and genomic impact of repetitive sequences [9].

Types and distribution of repetitive DNA sequences

The genomes of most eukaryotes contain a variety of repetitive DNA sequences [10,11]. Repetitive DNA may be dispersed throughout the genome or may be restricted at specific locations in a tandem configuration. According to the length of the repeated unit and array size, tandem repeated DNA sequences can be classified into three groups: (i) microsatellites with 2–5 bp repeats and an array size of the order of 10–100 units, (ii) minisatellites with 6–100 bp (usually around 15 bp) repeats and an array size of 0.5–30 kb and (iii) satellite DNA (satDNA) with a variable AT-rich repeat unit that often forms arrays up to 100 Mb. The monomer length of satDNA sequences ranges from 150–400 bp in majority of plants and animals. satDNA sequences are located at heterochromatic regions, which are found mostly in centromeric and subtelomeric regions in the chromosomes but also at intercalary positions. Figure 1 shows a diagrammatic representation of different types of repetitive sequences on a plant chromosome. Satellite DNAs are portions of the genome that can be separated as satellite peaks from primary DNA peaks [12] using the methods including restriction endonuclease digestion, colony filter hybridization with relic DNA, amplification with specific primers, colony filter hybridization and genomic self-amplification (also known as self-priming) [13].

Clustered DNA repeats can be detected in centromeric and telomeric heterochromatin, and are transcriptionally inert. Centromeric DNA (such as *CENH3*) is the most abundant tandem repeats found in both plants and animals [14]. In contrast, subtelomeric repetitive sequence families are usually genus-specific, such as TrsA and Os48 in *Oryza* species [15,16]; TrsB in *Oryza brachyantha* [17]; *Sac*I family in *Silene latifolia* [18]; and pAv34, pAc34, pRp34, pRn34 and pRs34 in *Beta* species [19], or chromosome (Chr)-specific, such as WE 35 on Chr 5B in *Triticum aestivum* [20], JNK family on Chr 2R in *Secale cereale* [21], *Afa*I family in Poaceae members [22], TRI on Chr Y in *Silene latifolia* [23] and RUSI on Chr 1 in *Rumex* species [24].

A particular sequence may be either species-specific or present in many species within a taxonomic family or various families, indicating that some repetitive sequences evolve rapidly, whereas others may be conserved [25]. For instance, the CL600 satellite repeat isolated from *Citrus limon* was detected in other Rutaceae members (*C. aurantium, C. paradisi,*

Poncirus trifoliata and *Fortunella margarita*) [26]. Similarly, the PCvKB repetitive DNA isolated from *Crocus vernus* was present in 16 species of *Crocus* analyzed and 2 species of Iridaceae, 3 species of Liliaceae and 1 species of Amaryllidaceae [27]. The *Hin*dIII repeat of *Brassica campestris* shows homology not only to that of other *Brassica* species but also to that of *Raphanus sativus, Sinapis alba, Diplotaxis muralis* and *Erucastrum* sp. [28]. Repetitive elements in cauliflower, mustard and radish, all belonging to Brassicaceae family, show 75%–80% homologies [29,30]. Mehrotra et al. [31] recently reported that the pCt*Kpn*I-I satellite repeat, initially isolated from *Carthamus tinctorius* and other species of *Carthamus* [31], is present in widely divergent families of angiosperms [32].

Various repetitive sequences from diverse taxa have been integrated into a database for easy accessibility. PlantSat, a database specialized for plant satDNA, has integrated sequence data from several resources such as NCBI and DNA Data Bank of Japan (DDBJ) [33], to provide a list of satDNA sequences for members of many plant families including Poaceae, Brassicaceae, Chenopodiaceae, Cucurbitaceae and Solanaceae [34–36], as well as many other plants. Plant satDNA sequences commonly have monomer unit lengths of 135–195 bp or 315–375 bp, which are consistent with reports that the basic monomer units of plant repetitive DNA sequences correspond to the length of DNA wrapped around a nucleosomal particle [10,33,37].

Different families of repetitive sequences show consistent presence of motifs like AA/TT dinucleotides, pentanucleotide CAAAA, *etc.* The presence of conserved motifs in unrelated repetitive sequence families suggests their significance for molecular mechanisms underlying the amplification and maintenance of tandem repeats in a genome, and the determination of specific chromatin properties of loci containing the repetitive DNA [33]. The occurrence of short, direct and inverted repeats and short palindromes is a characteristic feature of various plant satellite families. These may act as preferential sites for changes and as potential substrates for homologous recombination allowing rearrangements [38–42]. The repetitive sequences have the nearest-neighbor characteristics with high frequencies of GG, AG and GA nearest neighbors [43], which originate during the repair of heteroduplex intermediates of the exchange [44,45]. Frequent occurrence of GGT and GTT trinucleotides in the monomers of repeat sequences makes the sequence suitable substrate for the *de novo* telomere synthesis in the repairing process of broken chromosome ends [46].

Another characteristic feature of repetitive sequences is methylation. Methylation of DNA sequences is considered to trigger structural changes in DNA strands [21,22]. Methylation has been observed in satellite repeats following restriction analysis of genomic DNA with *Msp*I and *Hpa*I. Methylation has been reported to occur in a 500 bp satellite repeat family in *Arabidopsis* [47], in a JNK repeat family in Japanese rye [21], in a pCt*Kpn*I repeat family in *Carthamus* [31], and many other repeat families [26,48].

Functions of repetitive sequences

Repetitive DNA sequences are present in the heterochromatin region. Heterochromatin has been associated with several functions ranging from regulating gene expression to protecting chromosomal integrity. Heterochromatin can incur different

Table 1   Repeat DNA families in various plant species

Species	Repetitive sequence name	Repeat length* (bp)
Aegilops speltoides	spelt-1	150
Aegilops squarrosa	pAS1(Afa family)	336–337
(DD)	spelt-1	150
Allium cepa	ACSAT 1/ACSAT 2/ACSAT 3	370
	pAc074	314
Allium fistulosum	pAfi100	380
Arabidopsis thaliana	180 bp repeat/*Hin*dIII repeat/AtCon/AtCen/ pAL1/pAS1/pAtMR/ pAtHR/pAa214/AaKB27 family	180
Avena	pAm1	58
Beta vulgaris	pAv34/pAc34/pRp34/ pRn34/pRs34	334–362
	pBVI	327–328
	pHC28	149
	pHT30	140–149
	pHT49	162
	ppHC8	162
	pRN1	209–233
	pTS4.1	312
	pTS5	153–160
Brassica	pBcKB4/pBT11/ *Hin*dIII family/Canrep	175–180
	pBo1.6	203
Brassica campestris	BT4	296
	BT11	175
	CS1	88
	CT10	213
	pBcKB4	360
Brassica napus	Canrep	176
Brassica nigra	pBN4	459
	pBNE8	1732
Brassica oleracea	pBoKB1	360
Camellia sinensis	pMST11	894
Centaurea	*Hin*fI	350
Citrus limon	CL600	600
Crocus vernus	pCvKB4	270
Cucurbita pepo	350 bp satellite	349–352
Cucurbita maxima	170 bp satellite	168–170
Cucumis melo	*Hin*dIII repeat	352
Cucumis metuliferus	pMetSat	346
Cucumis sativus	Type I	182
	Type III	177
	Type IV	360
Diplotaxis erucoides	Canrep	175–180
Elaeis guineensis	pEgKB15	355
	pEgKB20	342
Elymus trachycaulus	pEt2	337–339
Glycine max	SB92	92
	STR120	120
Hordeum chilense	pHch1	2.6 kb
	pHch2	2.1 kb
	pHch3	500
	pHch4	2.6 kb
	pHch5	2.0 kb
	pHcKB6	339
Hordeum vulgare	HvRT	118
	pHvMWG2315	331
Leymus racemosus	350 bp family	350
	Lt1	380
	pLrAfa1–6	340
	TaiI	570
Lycopersicon	GR1	162
esculentum	pLEG15	168

Table 1   (continued)

Species	Repetitive sequence name	Repeat length* (bp)
Medicago truncatula	MtR1	166
	MtR2/MtR3	183, 166
Musa spp.	Radka1	685
	Radka2	409
	Radka3	808
	Radka4	605
	Radka5	742
	Radka6	193
	Radka7	596
	Radka8	337
	Radka9	334
	Radka10	689
Nicotiana tabacum	HRS60	180
	TAS49	460
Nicotiana rustica	HRS-60 family	180
Oryza australiensis	pOa237	1300
Oryza minuta	pOm1	239
	pOm4	438
	pOmA536	400
	pOmPB10	305
Oryza rufipogon	H2	615
Oryza sativa	C154	352
	C193	353
	CentO-C	126
	CentO/RCS2/TrsD	155
	Os48/TrsA	355
	OsG3	498
	OsG5	756
	TrsC	366
Olea europaea	81 bp family	81
	Oe179	179
	OeTaq80	80
	OLEU	178
	pOS218	218
Pennisetum glaucum	pPgKB19	137
Phaseolus vulgaris	PvMeso 31	3.4 kb
	PvMeso 47	1.7 kb
Pisum	PiSTR-A	211–212
	PiSTR-B	506
Potamogeton pectinatus L.	PpeRsa1	362–364
	PpeRsa2	355–359
Psathyrostachys juncea (NN)	pPjAfa1-3	340
Raphanus sativus	Canrep	175–180
	pRA5/pRB12	177
Rumex	RAE180	180–186
	RAE730 family	727–731
	RAYS1 family	922–932
	RUS1	170
Saccharum officinarum	SCEN family	140
Sorghum bicolor	pSau3A10/pCEN38	137
Secale africanum	pSaD15	887
Secale cereale (Japanese rye)	JNK family	1.2 kb
	pSc34	480
	pSc74	610
	pSc119.1	350
	pSc119.2	350
	pSc200	380
	spelt-1	150
Silene latifolia	15Ssp	159
	SacI	313
	STAR-C	43
	TRAYC	172
	X43.1	335

Table 1 (*continued*)

Species	Repetitive sequence name	Repeat length[*] (bp)
Sinapis alba	Canrep	175–180
Solanum brevidens	pSB1	322
	pSB7	167
Solanum bulbocastanum	2D8	5.9 kb
Solanum circaefolium	pSCH15	168
Solanum tuberosum	pST3	845
	pST10	121
Trifolium	TrR350	350
Triticum aestivum	dpTa1	340
	*Tai*I family	570
	WE35	320
Vicia faba	*Bam*H1	250, 1500
	Fok1	59
Vigna unguiculata	pVuKB1	488
Zea mays	Cent4	156
	CentC	156
	H2a/H2b	184–185
	MR68	410
	MR77	1.2 kb
Zingeria biebersteiniana	Zbcen1	755

Note: Plant species are listed alphabetically. [*] Repeat length is indicated in bp unless otherwise followed with "kb".

levels of expression of adjacent genes during inheritance, because of either a position effect or juxta-position of heterochromatin and highly active genes, as observed in *Drosophila* and *Saccharomyces* [49–52].

Repetitive sequences are implicated in numerous processes such as chromosome movement and pairing, centromeric condensation, chromosome recombination, sister chromatid pairing, chromosome association with the mitotic spindle, chromosome arrangement, interaction of chromatin proteins, histone binding, determination of chromosome structure, karyotypic evolution, regulation of gene expression and genome response to environmental stimuli and physiological changes. These are all considered key components of evolutionary mechanisms and karyotypic differentiation [53,32]. Therefore, repetitive sequences play an important role in the evolution of species [54], and they are all speculated to model the regulatory patterns of genes leading to phenotypic variation [55]. The occurrence of species-specific satDNA enables rapid and reliable identification of species. Grewal and Elgin [56] proposed the transcription of satDNA and its impact on heterochromatin, particularly in terms of the formation and maintenance of heterochromatin structure. Repetitive DNA sequence elements are also involved in cooperative molecular interactions for the formation of nucleoprotein complexes [57]. Repeat sequences may attract some specific nuclear proteins, and the chromatin folding code dictates the DNA–protein interactions, which may underlie the

genetic function of the tandem repeats [58]. Tandem repeats are proposed as "tuning knobs" in the evolution due to their ability to adjust the genetic hereditary traits and thereby facilitate the adaptation [59,60]. Melters et al. [14] have suggested that tandem repeats at centromeres may promote the concerted evolution of centromere DNA across chromosomes owing to their high prevalence. Centromere DNA transcripts have been reported to be involved in centromere structure and function [61]. It is speculated that if repetitive DNA is transposable, it may create novel genes [62]. The functions of satellite sequences are still not clear. In the post-genomics era, due to a wealth of resources, it is possible to gain insights into the functional potential of repetitive sequences.

Evolution of repetitive sequences

Repetitive DNA sequences exhibit cohesive and concerted evolution by mechanisms that cause continuous nuclear genome turnover and constitute molecular drive. Such concerted evolution produces high homogeneity in a repetitive DNA family where mutations are diagnostic for species, and are the origin of interspecies genetic divergence [63,64]. satDNA exhibits internal sequence variability depending on the ratio between mutation and homogenization/fixation rates within a species [65]. Levels of sequence identity between the satellite repeats can be attributed by the following factors: the rates and biases of transfer between homologous and nonhomologous chromosomes, the number and distribution of repeats, the physical constraints within the genome, the generation time, the effective population size and various biological and selective constraints [66].

Tandem repetitive sequences are considered to be generated *de novo* by the combinatorial action of molecular mechanisms such as mutations, unequal crossing over, gene conversion, slippage replication and/or rolling circle replication [44,67], which create and maintain homogeneity of satDNA sequences within species [65,68]. Among them, sequence duplication by aberrant recombination [69,70] or replication slippage [67,71,72] represents the primary event in the formation of a repeat.

Unequal crossing over is assumed to be the primary evolutionary force acting on satellite sequences [44]. Arrays of satDNA are maintained by unequal exchange and intrastrand exchange [73] and unequal crossing over accounts for the alterations in copy numbers of satellite monomers [44]. Individual repetitive units do not evolve independently; instead, the arrays evolve in concert. However, unequal crossing over, by itself, cannot create large tandem arrays of satDNA. Gene amplification and subsequent duplications also play a significant role in satDNA evolution. A few repeats may excise from a tandem array and circularize to provide a template for rolling circle replication [72], and after amplification into a linear array, the repeats may be inserted into a new location in the genome.
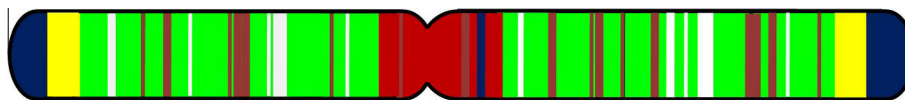

Figure 1   General distribution of repetitive sequences on plant chromosomes
Distribution of different types of repetitive sequences is represented diagrammatically on a plant chromosome with different colors. Red, centromeric tandem repeats; blue, telomeric repeats; yellow, sub-telomeric tandem repeats; green, intercalary tandem repeats; brown, dispersed repeats; white, genes and low-copy sequences.

Repetitive elements are under different evolutionary constraints as compared to genes, and are considered as fast-evolving components of eukaryotic genomes. The high evolution rates of repetitive sequences can be used to differentiate related species [35,74–78]. Hybrid polyploids are excellent models for studying the evolution of repetitive sequences [37], and variations in their repetitive sequences allow them to be used for taxonomic and phylogenetic studies [79]. Sequence homogeneity and evolution of repetitive sequences are correlated with their copy number. Repetitive sequences with a low copy number are homogeneous and evolve slowly, whereas repetitive sequences with a high copy number are more heterogeneous and evolve quickly [80,81]. Sequence divergence in satDNA proceeds in a gradual manner due to the accumulation of nucleotide substitutions. For instance, sequences were highly conserved in the repetitive satDNA of *Palorus ratzeburgii* and *Palorus subdepressus* [82,83]. In addition, high sequence conservation was also observed in human α-satDNA repeat, which is a rare and a highly-conserved repeat in the evolutionarily distant species such as chicken and zebrafish [84]. Similarly, the simple dodeca satDNA repeat is also conserved among evolutionarily distant organisms such as fruitfly, *Arabidopsis* and human [85]. Evolutionary persistence of large tandem arrays is affected not only by the balance between the rate of amplification and the rate of unequal exchange, but also by a wide range of mechanisms for recombination, replication and gene amplification. However, the amount of bias in these processes acting on satDNA remains unresolved. Nonetheless, natural selection does not have any effect on satellite repeats, because satDNA sequences are not transcribable and are consequently neutral to selection [6,7].

Evolutionary significance of repetitive sequences

Repetitive sequences are speculated to influence cytoplasmic, cellular and developmental processes [86] by increasing genome size and affecting chromosomal recombination. satDNA repeats represent recombination "hotspots" of genome reorganization [87]. The occurrence of satDNA in interstitial and telomeric heterochromatin reduces genetic recombination in adjacent regions [87]. Robertsonian chromosome fusion or fission, the joining of two telo/acrocentric chromosomes at their centromeres to form a metacentric, has been postulated to depend on sequence similarity of regions on the two chromosomes because of large amounts of centromeric, heterochromatic satDNA in eukaryotic genomes [88]. Satellite repeats in eukaryotes are likely involved in sequence-specific interactions and subsequently in epigenetic processes. Nonetheless, repetitive satDNA also has a sequence-independent role in the formation and maintenance of heterochromatin. Transcripts from tandem arrays or satellites are processed by RNA-dependent RNA polymerase (RdRP) and Dicer to produce small interfering RNAs (siRNAs) [89,90]. satDNA-derived siRNAs are probably involved in posttranscriptional gene regulation through the action of the RNA-induced silencing complex (RISC) [91]. Additionally, satDNA-derived siR-NAs also have a possible role in gene expression and heterochromatin formation on tandemly-repeated noncoding regions, and in the expression of particular genes with embedded satellite repeats [92].

Recent advances and future perspectives

A remarkable advance in the knowledge of repetitive sequences has occurred in recent years because of the introduction of next-generation sequencing technologies. These technologies can be applied to highly complex populations of repetitive elements in plant genomes, and have been used to characterize genomes and establish phylogenies based on repetitive sequences in *Silene latifolia*, *Helianthus* and *Orobanche* species [93–97]. Various strategies such as single nucleotide polymorphism (SNP) discovery and more sophisticated approaches are being developed to assemble NGS data and to analyze repeats for a better understanding of their contribution to gene function and genome evolution [95]. 454 sequencing and Illumina platforms have been extensively used to comprehensively characterize repetitive DNA in the pea and olive genome as well as the satellite sequences in *Silene*, including the detection of the most conserved regions, reconstruction of consensus sequences of repeat monomers, identification of major sequence variants and designing of hybridization probes for localization on chromosomes using fluorescence *in situ* hybridization (FISH) [93,97,98]. 454 pyrosequencing was also used to detect copy-number repeats in the soybean genome [99] and to deduce the repeat composition of genomes for nine species of Orobanchaceae [94]. Recently, Sergeeva et al. [100] presented a detailed account of repetitive sequences of wheat based on 454 sequencing. Whole genome shotgun sequencing has also been employed to identify and analyze the most abundant tandem repeats from diverse animal and plant species [14].

Various web tools have been recently introduced for the analysis of repetitive sequences. A tool called "REViewer" was developed for visualizing and analyzing repetitive elements [101]. A comprehensive toolkit "RepEx" was developed by Gurusaran et al. [102] for extracting various repeats (inverted, everted and mirror) from genomic sequences. Recently, Novák et al. [103] introduced a collection of software tools called "Repeat Explorer" for characterizing repetitive elements and identifying high- and medium-copy repeats in higher plant genomes.

Repetitive sequences are technically challenging to clone and sequence and can be better studied by combining various approaches like mapping and sequence analysis. Repetitive sequences also pose challenges in sequencing and assembling of genomes. Cytogenetics, genomics and bioinformatics tools have allowed the genomes of complex eukaryotes to be investigated. Whole genome resequencing studies, genome-wide analysis, transposon-based sequencing strategies and fine mapping of repetitive sequences can elucidate the structure, evolution, and functional potential of this enigmatic yet indispensable component of the genome. This will in turn assist in sequencing and assembly of complex eukaryotic genomes.

Competing interests

The authors have declared that no competing interests exist.

Acknowledgements

References

[1] Nagl W. Zellkern und Zellzyklen. In: Doenecke D, editor. Biochemical education. Stuttgart: Eugen Ulmer Verlag; 1976. p. 486.

[2] Nagl W, Fusenig HP. Types of chromatin organization in plant nuclei. Plant Sys Evol 1979;2:221–33.

[3] SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, et al. Nested retrotransposons in the intergenic regions of the maize genome. Science 1996;274:765–8.

[4] Pearce SR, Harrison G, Li D, Heslop-Harrison J, Kumar A, Flavell AJ. The Ty1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 1996;250:305–15.

[5] Leitch IJ, Chase MW, Bennett MD. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. Ann Bot 1998;82:85–94.

[6] Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. Nature 1980;284:604–7.

[7] Doolittle W, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature 1980;284:601–3.

[8] Plohl M. Those mysterious sequences of satellite DNAs. Period Biol 2010;112:403–10.

[9] Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive sequences in complex genomes: structure and evolution. Annu Rev Genomics Hum Genet 2007;8:241–59.

[10] Heslop-Harrison JS, Schmidt T. Genomes, genes and junk: the large-scale organization of plant genomes. Trends Plant Sci 1998;3:195–9.

[11] Jiang J, Birchler JA, Parrott WA, Dawe RK. A molecular view of plant centromeres. Trends Plant Sci 2003;8:570–5.

[12] Beridze T. Satellite DNA. Berlin: Springer-Verlag; 1986.

[13] Macas J, Pozárková D, Navrátilová A, Nouzová M, Neumann P. Two new families of tandem repeats isolated from genus *Vicia* using genomic self-priming PCR. Mol Gen Genet 2000;263:741–51.

[14] Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 2013;14:R10.

[15] Cheng Z, Stupar RM, Gu M, Jiang J. A tandemly repeated DNA sequence is associated with both knob-like heterochromatin and a highly decondensed structure in the meiotic pachytene chromosomes of rice. Chromosoma 2001;110:24–31.

[16] Ohtsubo H, Ohtsubo E. Involvement of transposition in dispersion of tandem repeat sequences (TrsA) in rice genomes. Mol Gen Genet 1994;245:449–55.

[17] Nakajima R, Noma K, Ohtsubo H, Ohtsubo E. Identification and characterization of two tandem repeat sequences (TrsB and TrsC) and a retrotransposon (RIRE1) as genome-general sequences in rice. Genes Genet Syst 1996;71:373–82.

[18] Garrido-Ramos MA, de la Herrán R, Ruiz Rejón M, Ruiz Rejón C. A subtelomeric satellite DNA family isolated from the genome of the dioecious plant *Silene latifolia*. Genome 1999;42:442–6.

[19] Dechyeva D, Schmidt T. Molecular organization of the terminal repetitive DNA in *Beta* species. Chromosome Res 2006;14:881–97.

[20] Ueng PP, Hang A, Tsang H, Vega JM, Wang L, Burton CS, et al. Molecular analyses of a repetitive DNA sequence in wheat (*Triticum aestivum* L.). Genome 2000;43:556–63.

[21] Nagaki K, Tsujimoto H, Sasakuma T. A novel repetitive sequence, termed the JNK repeat family, located on an extra heterochromatic region of chromosome 2R of Japanese rye. Chromosome Res 1999;7:95–101.

[22] Nagaki K, Kishii M, Tsujimoto H, Sasakuma T. Tandem repetitive Afa-family sequences from *Leymus racemosus* and *Psathyrostachys juncea* (Poaceae). Genome 1999;42:1258–60.

[23] Cermak T, Kubat Z, Hobza R, Koblizkova A, Widmer A, Macas J, et al. Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes. Chromosome Res 2008;16:961–76.

[24] Navajas-Pérez R, Schwarzacher T, Ruiz Rejón M, Garrido-Ramos MA. Characterization of RUSI, a telomere-associated satellite DNA, in the genus *Rumex* (Polygonaceae). Cytogenet Genome Res 2009;124:81–9.

[25] Anamthawat-Jónsson K, Heslop-Harrison JS. Species specific DNA sequences in the Triticeae. Hereditas 1992;116:40–54.

[26] De Felice B, Wilson RR, Ciarmiello L, Conicella C. A novel repetitive DNA sequence in lemon (*Citrus limon* (L.) Burm.) and related species. J Appl Genet 2004;45:15–20.

[27] Frello S, Heslop-Harrison JS. Repetitive DNA sequences in *Crocus vernus* Hill. (Iridaceae): the genomic organization and distribution of dispersed elements in the genus *Crocus* and allies. Genome 2000;43:902–9.

[28] Lakshmikumaran M, Ranade SA. Isolation and characterization of a highly repetitive DNA from *Brassica campestris*. Plant Mol Biol 1990;14:447–8.

[29] Benslimane AA, Dron M, Hartmann C, Rode A. Small tandemly repeated sequences of higher plants likely originate from a tRNA gene ancestor. Nucleic Acids Res 1986;14:8111–9.

[30] Grellet F, Delcasso D, Panabieres F, Delseny M. Organization and evolution of a higher plant alphoid-like satellite DNA sequence. J Mol Biol 1986;187:495–507.

[31] Mehrotra S, Goel S, Raina SN, Rajpal VR. Significance of satellite DNA revealed by conservation of a widespread repeat DNA sequence among angiosperms. Appl Biochem Biotechnol 2014;173:1790–801.

[32] Mehrotra S, Goel S, Sharma S, Raina SN, Rajpal VR. Sequence analysis of KpnI repeat sequences to revisit the phylogeny of the Genus *Carthamus* L. Appl Biochem Biotechnol 2013;169: 1109–25.

[33] Macas J, Mészáros T, Nouzová M. PlantSat: a specialized database for plant satellite repeats. Bioinformatics 2002;18: 28–35.

[34] Schmidt T, Jung C, Metzlaff M. Distribution and evolution of two satellite DNAs in the genus *Beta*. Theor Appl Genet 1991;82:793–9.

[35] Schmidt T, Heslop-Harrison JS. Variability and evolution of highly repeated DNA sequences in the genus *Beta*. Genome 1993;36:1074–9.

[36] Schmidt T, Heslop-Harrison JS. High-resolution mapping of repetitive DNA by *in situ* hybridization: molecular and chromosomal features of prominent dispersed and discretely localized DNA families from the wild beet species *Beta procumbens*. Plant Mol Biol 1996;30:1099–119.

[37] Kubis S, Schmidt T, Heslop-Harrison JS. Repetitive DNA elements as a major component of plant genomes. Ann Bot 1998;82:45–55.

[38] Gordenin DA, Lobachev KS, Degtyareva NP, Malkova AL, Perkins E, Resnick MA. Inverted DNA repeats: a source of eukaryotic genomic instability. Mol Cell Biol 1993;13:5315–22.

[39] Linares C, Ferrer E, Fominaya A. Discrimination of the closely related A and D genomes of the hexaploid oat *Avena sativa* L. Proc Natl Acad Sci U S A 1998;95:12450–5.

[40] Vershinin A, Svitashev S, Gummesson PO, Salomon B, von Bothmer R, Bryngelsson T. Characterization of a family of tandemly repeated DNA sequences in Triticeae. Theor Appl Genet 1994;89:217–25.

[41] Vershinin AV, Schwarzacher T, Heslop-Harrison JS. The large-scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes. Plant Cell 1995;7: 1823–33.

[42] Vershinin AV, Alkhimova AG, Heslop-Harrison JS, Potapova TA, Omelianchuk N. Different patterns in molecular evolution of the Triticeae. Hereditas 2001;135:153–60.

[43] Blake RD, Wang JZ, Beauregard L. Repetitive sequence families in *Alces alces americana*. J Mol Evol 1997;445:509–20.

[44] Smith GP. Evolution of repeated DNA sequences by unequal crossover. Science 1976;191:528–35.

[45] Friedberg EC, Walker GC, Siede W. DNA repair and mutagenesis. Washington, DC: ASM Press; 1995.

[46] Tsujimoto H. Molecular cytological evidence for gradual telomere synthesis at the broken chromosome ends in wheat. J Plant Res 1993;106:239–44.

[47] Simoens CR, Gielen J, Van Montagu M, Inzé D. Characterization of highly repetitive sequences of *Arabidopsis thaliana*. Nucl Acids Res 1988;16:6753–66.

[48] Barragán MJ, Martínez S, Marchal JA, Bullejos M, Díaz de la Guardia R, Sánchez Bullejos A. Highly repeated DNA sequences in three species of the genus *Pteropus* (Megachiroptera, Mammalia). Heredity (Edinb) 2002;88:366–70.

[49] Zhimulev IF, Belyaeva ES, Bgatov AV, Baricheva EM, Vlassova IE. Cytogenetic and molecular aspects of position effect variegation in *Drosophila melanogaster*. Chromosoma 1986;94: 492–504.

[50] Burgess-Beusse B, Farrell C, Gaszner M, Litt M, Mutskov V, Recillas-Targa F, et al. The insulation of genes from external enhancers and silencing chromatin. Proc Natl Acad Sci U S A 2002;99:16433–7.

[51] Noma K, Allis CD, Grewal SI. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science 2001;293:1150–5.

[52] Donze D, Kamakaka RT. RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in *Saccharomyces cerevisiae*. EMBO J 2001;20:520–31.

[53] Silva DM, Pansonato-Alves JC, Utsunomia R, Daniel SN, Hashimoto DT, Oliveira C, et al. Chromosomal organization of repetitive DNA sequences in *Astyanax bockmanni* (Teleostei, Characiformes): dispersive location, association and co-localization in the genome. Genetica 2013;141:329–36.

[54] Britten RJ. Transposable element insertions have strongly affected human evolution. Proc Natl Acad Sci U S A 2010;107: 19945–8.

[55] Knight JC. Allele-specific gene expression uncovered. Trends Genet 2004;20:113–6.

[56] Grewal SI, Elgins SC. Transcription and RNA interference in the formation of heterochromatin. Nature 2007;447:399–406.

[57] Shapiro JA, von Sternberg R. Why repetitive DNA is essential to genome function. Biol Rev Camb Philos Soc 2005;80:227–50.

[58] Vogt P. Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". Hum Genet 1990;84: 301–36.

[59] Kashi Y, King D, Soller M. Simple sequence repeats as a source of quantitative genetic variation. Trends Genet 1997;13: 74–8.

[60] Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. Trends Genet 2006;22:253–9.

[61] Plohl M, Meštrović N, Mravinac B. Centromere identity from the DNA point of view. Chromosoma 2014;123:313–25.

[62] Rao SR, Trivedi S, Emmanuel D, Merita K, Hynniewta M. DNA repetitive sequences-types, distribution and function: a review. J Cell Mol Biol 2010;7(2) & 8(1):1–11.

[63] Dover G. Molecular drive. Trends Genet 2002;18:587–9.

[64] Heslop-Harrison JS, Schmidt T. Plant nuclear genome composition. Chichester: John Wiley & Sons Ltd; 2012.

[65] Dover G. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 1986;2:159–65.

[66] Ohta T, Dover GA. The cohesive population genetics of molecular drive. Genetics 1984;108:501–21.

[67] Stephan W. Tandem-repetitive noncoding DNA: forms and forces. Mol Biol Evol 1989;6:198–212.

[68] Okumura K, Kiyama R, Oishi M. Sequence analyses of extrachromosomal *Sau*3A and related family DNA: analysis of recombination in the excision event. Nucleic Acids Res 1987;15:7477–89.

[69] Kruger J, Vogel F. Population genetics of unequal crossing over. J Mol Evol 1975;4:201–47.

[70] Ohta T. Evolution and variation of multigene families. New York: Springer-Verlag; 1980.

[71] Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 1987; 4:203–21.

[72] Walsh JB. Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics 1987;115:553–67.

[73] Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 1994;371: 215–20.

[74] Flavell RB. Sequence amplification, deletion, and rearrangement: major sources of variation during species divergence. In: Dover GA, Flavell RB, editors. Genome evolution. London: Academic Press; 1982. p. 301–23.

[75] Lapitan NLV. Organization and evolution of higher plant nuclear genomes. Genome 1992;35:171–81.

[76] Hemleben V, Zentgraf U, King K, Borisjuk N, Schweizer G. Middle repetitive and highly repetitive sequences detect polymorphisms in plants. In: Kahl K, Appelhans H, Kompf J, Driesel AJ, editors. DNA-polymorphisms in eukaryotic genomes. Bio Tech Forum (BFT), Adv Mol Genet, vol. 5. Heidelberg: Huethig Verlag; 1992. p. 157–70.

[77] Hemleben V. Repetitive and highly repetitive DNA components as molecular markers for evolutionary studies and in plant breeding. Curr Top Mol Genet (Life Sci Adv) 1993;1:173–85.

[78] King K, Jobst J, Hemleben V. Differential homogenization and amplification of two satellite DNAs in the genus *Cucurbita*. J Mol Evol 1995;41:996–1005.

[79] Smith DB, Flavell RB. The relatedness and evolution of repeated nucleotide sequences in the genomes of some Gramineae species. Biochem Genet 1974;12:243–56.

[80] Stadler M, Stelzer T, Borisjuk N, Zanke C, Schilde-Rentschler L, Hemleben V. Distribution of novel and known repeated elements of *Solanum* and application for the identification of somatic hybrids among *Solanum* species. Theor Appl Genet 1995;91: 1271–8.

[81] Helm M, Hemleben V. Characterization of a new prominent satellite of *Cucumis metuliferus* and differential distribution of satellite DNA in cultivated and wild species of *Cucumis* and in related genera of Cucurbitaceae. Euphytica 1997;94:219–26.

[82] Mravinac B, Plohl M, Mestrović N, Ugarković D. Sequence of PRAT satellite DNA "frozen" in some Coleopteran species. J Mol Evol 2002;54:774–83.

[83] Mravinac B, Plohl M, Ugarković D. Preservation and high sequence conservation of satellite DNAs indicate functional constraints. J Mol Evol 2005;61:542–50.

[84] Li YX, Kirby ML. Coordinated and conserved expression of alphoid repeat and alphoid repeat-tagged coding sequences. Dev Dyn 2003;228:72–81.

[85] Abad JP, Carmena M, Baars S, Saunders RD, Glover DM, Ludena P, et al. Dodeca-satellite: a conserved G + C-rich satellite from the centromeric heterochromatin of *Drosophila melanogaster*. Proc Natl Acad Sci U S A 1992;89:4663–7.

[86] Cavalier-Smith T. The evolution of genome size. Chichester: Wiley; 1985.

[87] Miklos GLG. Sequencing and manipulating highly repeated DNA. In: Flavell FB, editor. Genome evolution. London: Academic Press; 1982.

[88] White MJD. Animal cytology and evolution. 3rd ed. New York: Cambridge University Press; 1973. p. 3–8.

[89] Martienssen RA. Maintenance of heterochromatin by RNA interference of tandem repeats. Nat Genet 2003;35:213–4.

[90] Plohl M, Luchetti A, Mestrović N, Mantovani B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. Gene 2008;409:72–82.

[91] Hammond SM, Boettcher S, Caudy AA, Kobayasi R, Hannon GJ. Argonaute2, a link between genetic and biochemical analyses of RNAi. Science 2001;293:1146–50.

[92] Ugarkovic D. Functional elements residing within satellite DNAs. EMBO Rep 2005;6:1035–9.

[93] Macas J, Kejnovský E, Neumann P, Novák P, Koblížková A, Vyskot B. Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. PLoS One 2011;6:e27335.

[94] Piednoël M, Aberer AJ, Schneeweiss GM, Macas J, Novak P, Gundlach H, et al. Next-generation sequencing reveals the impact of repetitive DNA in phylogenetically closely related genomes of Orobanchaceae. Mol Biol Evol 2012;29:3601–11.

[95] Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 2011;13:36–46.

[96] Natali L, Cossu RM, Barghini E, Giordani T, Buti M, Mascagni F, et al. The repetitive component of the sunflower genome as revealed by different procedures for assembling next generation sequencing reads. BMC Genomics 2013;14:686.

[97] Macas J, Neumann P, Navrátilová A. Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 2007;8:427.

[98] Barghini E, Natali L, Cossu MR, Giordani T, Pindo M, Cattonaro F, et al. The peculiar landscape of repetitive sequences in the olive (*Olea europaea* L.) genome. Genome Biol Evol 2014;6:776–91.

[99] Swaminathan K, Varala K, Hudson ME. Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. BMC Genomics 2007;8:132.

[100] Sergeeva EM, Afonnikov DA, Koltunova MK, Gusev VD, Miroshnichenko LA, Vrána J, et al. Common wheat chromosome 5B composition analysis using low-coverage 454 sequencing. Plant Genome 2014;7:1–16.

[101] You RN, Kim WC, Lee KH, Lee YK, Shin KS, Cho K, et al. REViewer: a tool for linear visualization of repetitive elements within a sequence query. Genomics 2013;102:209–14.

[102] Gurusaran M, Ravella D, Sekar K. RepEx: repeat extractor for biological sequences. Genomics 2013;102:403–8.

[103] Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 2013;29:792–3.