BMC Genomics

RESEARCH

Open Access

Large tandem repeats of grass frog (*Rana temporaria*) in silico and in situ

Marina A. Popova[1,2*], Aleksey S. Komissarov[3], Dmitrii I. Ostromyshenskii[1], Olga I. Podgornaya[1] and Aleksandra O. Travina[1*]

Abstract

Background Genomes of higher eukaryotes contain a large fraction of non-coding repetitive DNA, including tandem repeats (TRs) and transposable elements (TEs). The impact of TRs on genome structure and function and the importance of TR transcripts have been described for several model species. Amphibians have one of the most diverse genome sizes among vertebrates, attributed to the abundance of repetitive non-coding DNA. Consequently, amphibians are good models for the analysis of repetitive sequences, including TRs. However, few studies have focused on amphibian genomes.

Results Bioinformatic analyses were performed to characterise the content and localisation of TRs in the sequenced grass frog *Rana temporaria* genome. By applying different bioinformatic approaches, 76 TR families and 314 single TR arrays (not grouped into families) were identified. Each TR was characterised on the basis of chromosomal position, monomer length and variability and GC content. Bioinformatic analysis revealed a great diversity of TRs, with a clear predominance of TRs with short monomers (< 100 bp), although TRs with long monomers (> 1000 bp) also exist. The six most abundant TRs were successfully mapped by fluorescence in situ hybridization (FISH), which highlighted the presence of specific TR sequences in strategic chromosomal regions, i.e., the pericentromeric regions. A comparison of the results of in situ and in silico TR mapping revealed some inaccuracies in the assembly of heterochromatic regions. A putative new non-autonomous TE called "FEDoR" (Frog Element Dispersed organised Repeat) is also described. FEDoR is ~3.5 kb in length, has no significant similarity to any known TE family, contains multiple internal TR motifs, and is flanked on both sides by pairs of inverted repeat sequences (IRSs) and target site duplications (TSDs).

Conclusion Characterisation of TRs in this frog species has provided some insights regarding TR biology in Anuran amphibians.

Keywords Tandem repeats, Satellite DNA, Amphibian, FISH, Transposable elements

*Correspondence:
Marina A. Popova
marinaalexpopova@yandex.ru; Marina.Popova2@skoltech.ru
Aleksandra O. Travina
alotra1234@gmail.com
[1]Institute of Cytology RAS, Saint-Petersburg 194064, Russia
[2]Center for Molecular and Cellular Biology, Skolkovo Institute of Science and Technology, Moscow 121205, Russia
[3]Applied Genomics Laboratory, SCAMT Institute, ITMO University, Saint Petersburg 197101, Russia

Background

Genomes of higher eukaryotes contain a large fraction of non-coding repetitive DNA sequences, which are classified according to their structure and genomic organisation, as either dispersed and tandem repeats (TRs) [1]. Dispersed repeats are embedded in DNA as separate sequences which are repeated many times throughout the genome. Dispersed repeats are mainly represented by transposable elements (TEs) [2], which are classified as either retrotransposons or DNA transposons (DNA-TEs), each of which has orders, superfamilies and families (reviewed in [3]). TR DNA sequences are organised as multiple copies of sequences of a certain size (repeat unit or monomer) arranged in a head-to-tail pattern to form tandem arrays which can span up to several megabases [4]. Historically, some TRs have been referred to as satellite DNA (satDNA). The term "satellite DNA" was derived from CsCl density gradient centrifugation experiments that identified satellite bands of DNA separated from the bulk genomic DNA due to their skewed GC content [5]. In the current study, the term "TR" is used. This term can be formalised using the bioinformatic approach, as all TR features (monomer length, chromosome position, array variability, and GC content) have numerical expressions [6].

TR DNA constitutes a large portion of the genome, sometimes exceeding 50% of total DNA [4]. TRs are mostly found in constitutive heterochromatin blocks [7], predominantly in the centromeric (CEN), pericentromeric (periCEN) and subtelomeric (subTEL) chromosomal regions, although they have also been reported in euchromatic regions [8, 9]. TRs are among the most rapidly evolving genomic elements, with the majority being genus- or even species-specific [10]. High TR DNA divergence could be important for speciation and has been suggested to be an essential force triggering reproductive isolation [11, 12]. TRs have been implicated in a variety of important cellular functions, including kinetochore formation, spatial genome organisation, chromosomal rearrangements, segregation during cell division, homologous chromosome pairing and others [13–16]. TRs are not transcriptionally inert [17, 18]. They are highly transcribed during embryogenesis, and their transcripts are crucial for the overall 3D structure of the genome. Murine periCEN major satellite (MaSat) transcripts are responsible for the reorganisation of periCEN DNA into chromocenters [19]. Transcription of TRs was also detected in amphibian and avian oocytes at the lampbrush stage of oogenesis [18]. Despite the functional importance of TR sequences, genome-wide analyses of TRs have only been performed in a limited number of species (mostly mammals and some model organisms). Up to now, most studies of TRs and their transcripts employ probes designed from cloned monomers or derivatives after genomic DNA restriction rather than those based on in silico genome analyses.

Amphibians are good models for TR studies due to their extraordinary biodiversity, wide range of genome sizes [20, 21], high amounts of TRs [22] and recognition of TR transcription during their oogenesis [18]. The order Anura accounts for over 80% of extant amphibian species.

However, information concerning the TRs of these species remains sparse. For example, the only cloned genus-specific repeat S1 is known for European brown frogs *Rana* [23, 24]. There have been almost no genomic studies of TRs in amphibians. Until recently, the large genome size and high repetitive sequence content of many amphibian species have led to a marked delay in reference genome development in amphibians compared to other vertebrate species [25–28]. The sequenced among the first genomes of *Xenopus* species [29, 30] are best studied, however, the extent to which *Xenopus* species are representative of anurans remains in question because the Pipidae group, which contains 41 species [31], diverged early from the ancestral anurans [32]. Recent technological advances in DNA sequencing, along with reduced sequencing costs, have led to an increase in the number of amphibian genome assemblies [27, 28]. Ranidae, a large family of frogs in the order Anura, referred to as the true frogs, is one of the most studied amphibian families, along with Pipidae and Bufonidae. The family contains several genera and 455 species [31]. To date, 12 Ranidae genomes have been published (NCBI genome database records accessed on 19 February 2025), including that of the grass frog *Rana temporaria* [33], which is native to Europe. Amphibians, particularly *Xenopus* and *Rana*, are model organisms of developmental biology [34–38] because they offer significant advantages due to their large oocytes [39], the large physical size of their chromosomes [40] and free-living embryos [41, 42]. The paucity of information on TR DNA in amphibian species limits development in this area.

Oogenesis and embryonic development of the grass frog *R. temporaria* are well studied at the morphological level [34, 35, 37, 43, 44]. Repetitive non-coding DNA sequence analysis can open up new perspectives for understanding the molecular basis of development. The current study aimed to search for TRs in the raw sequencing data of the *R. temporaria* genome and in the genome assembly, to verify the bioinformatic TR predictions by in situ hybridization (FISH), to map TRs on pseudochromosomes in silico and to compare with in situ data.

Methods

Raw reads and genome assembly

Publicly available Illumina reads from the Bioproject on *R. temporaria* low-coverage genome sequence (PRJNA294436) of the sample *R. temporaria* isolate RtempB1-S22F (common frog) were used for raw reads.

aRanTem1.1 from the NCBI database with accession number GCF_905171775.1 and its annotation [33] were used for assembly analysis. Metrics provided by the NCBI Assembly Database were used for assembly quality metrics. The aRanTem1.1 assembly has a size of 4.1 Gb, an N50 of 481.8 Mb, and an L50 of 4. It is organised into 13 pseudochromosomes and 554 scaffolds. BUSCO analysis against the tetrapoda_odb10 dataset (5310 BUSCOs) indicated a completeness of 95.2% (S:92.8%, D:2.4%), with 0.6% fragmented and 4.1% missing BUSCOs.

Search for tandem repeats in raw reads

Tandem Repeat Analyser (TAREAN) [45] and extracTR (https://github.com/aglabx/extracTR) and raw reads of *R. temporaria* (accession: PRJNA294436; 2015) were used for TR analysis. Prior to analysis, reads were trimmed from adapters using Trimmomatic v.0.39 [46] with default parameters. All trimmed reads (147539853 read pairs) were used for extracTR. k-mers (k = 23) were extracted from the dataset of raw reads of *R. temporaria* (accession: PRJNA294436; 2015) after removing the Illumina adapters and were sorted according to their frequency of occurrence. Employing a de Bruijn graph methodology, circular paths were then assembled from these k-mers. The most frequently occurring k-mers within each circular path were identified and sequentially arranged. To verify extracTR results, an older approach with TAREAN software [45] was also used after adapter trimming. A random subset of one million read pairs was used for TAREAN, run on the RepeatExplorer Galaxy server (https://repeatexplorer-elixir.cerit-sc.cz/) with default parameters. The nomenclature for TR families combines "Rtem" representing *R. temporaria*, the length of the most common monomer (if unspecified), and a distinguishing letter for families with the same monomer length. The species specification "Rtem" was omitted from all figures and tables, which concerned only one species. Since the known S1A satDNA of *R. temporaria* is not present in the Dfam v.3.7 database [47], we used the BLAST program v.2.12.0 [48]. with default parameters and the core nucleotide (core_nt) database to identify S1A from the nucleotide sequences obtained.

Table 1 The most abundant families of TRs

Name	The most frequently occurring k-mers (FISH probe)	extracTR (reads per million)	TAREAN (reads per million)
35A	ATAGTGGTATAGTGATGTCATAG	5103	5728
32A	AGATAGATAGGGAAAGAGAGAGA	2787	2964
47A	CCATCAAACGCAGCCACTGTGCC	1615	Not found
494A (S1A)	AACTTGGGGAGCATCTTCCTGAA	1606	5719
219A (5S rDNA)	TATCTTCAAGAGAGTTAAGGACA	995	2121
35B	TGCAGGGTGCTGTGTAATGTAAA	615	Not found

Probe design

The most frequently occurring k-mers with k = 23 were selected for the probe design for the six most abundant TRs. Short single-stranded oligonucleotide probes were designed using the extracTR software (https://github.com/aglabx/extracTR). Primer3 [49] v.4.1.0 was used to check for possible discrepancies (the secondary structure and GC content in the oligonucleotides). The six probes were synthesised (DNA Synthesis, Russia) as DNA oligonucleotides with 3′ ends labeled with biotin. The sequences of the probes are listed in Table 1 according to their abundance in the genome (based on the results of the extracTR). As the 494A TR contains a relatively long monomer size and corresponds to the previously described S1A [24], the additional PCR-amplified probe was also used (see 2.5).

R. temporaria chromosome plates

Frogs were collected from their natural habitats in the Leningrad region, Russia. Four male and one female *R. temporaria* were anaesthetised by submersion in a 1% solution of tricaine methanesulfonate (MS 222; Sigma-Aldrich, Germany) and subsequently euthanised by decapitation. This procedure complies with the international principles for the humane treatment of laboratory animals. Metaphase chromosomes were obtained from intestinal epithelial cells and testes according to the published protocol [50]. The resulting cell suspension was dropped on slides, heated on the surface of a water bath (55 °C), and air-dried.

DNA isolation and PCR amplification

DNA was extracted from frog liver using a standard method [23]. Primers (S1A_F 5′-AACTTGGGGAGCATCTTCCT-3′, S1A_R 5′-TCCCATGTTAAACGGTCCAT-3′) for amplification of ~250 bp fragment of S1A TR were designed on the basis of 21 cloned and sequenced *R.temporaria* S1A isolates [24]. The S1A FISH-probe was amplified in the presence of biotin-digoxigenin-11-dUTP (Roche, Germany) by PCR: 95 °C 2 min and then 95 °C 30 s, 57 °C 30 s, 72 °C 30 s, 30 circles.

Fluorescence in situ hybridization (FISH)

FISH was performed with single-stranded probes according to the standard method [51], with some modifications. Slides with metaphase spreads were treated with

RNAse (Sigma-Aldrich, R6513, Merck KGaA, Darmstadt, Germany) stock solution (10 µg/mL) diluted 1:200 with 2x SSC buffer (0.3 M NaCl, 0.03 M sodium citrate) for 45–60 min at 37 °C and washed 3 times for 5 min with 2x SSC at room temperature. Metaphase spreads were denatured in solution (70% formamide, 2xSSC) for 3–5 min at 72 °C and dehydrated in an ethanol series at – 20 °C. Then, slides were incubated with the biotinylated probe diluted in Hybrisol (Molecular Probes, Eugene, OR, USA), for 16–18 h, at 37 °C. The S1A probe resolved in Hybrisol buffer was denatured at 80 °C for 7 min and cooled by immediately placing the mixture on ice for at least 10 min. Commercially synthesised probes did not require this denaturation step. After 3x post-hybridisation washes in 2x SSC, the slides were incubated with Alexa-488-conjugated streptavidin (Thermo Fisher Scientific, Waltham, MA, USA). Biotinylated anti-streptavidin (Vector Laboratories, Burlingame, CA, USA) was used to amplify the signal, and then Alexa-488-conjugated streptavidin was used again for labelling; all reagent concentrations were according to the manufacturer's protocol. Finally, the slides were mounted in Prolong Gold Antifade with DAPI (Thermo Fisher Scientific, Waltham, MA, USA) and stored refrigerated in the dark.

Microscopy and image acquisition

The preparations were analysed using a Leica TCS SP5 (Leica Microsystems, Wetzlar, Germany) laser scanning confocal microscope in the Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia. Chromosome identification was performed according to Guillemin [52]. Approximately 15 metaphase spreads were analysed for each TR probe.

Search for tandem repeats in genome assembly

TR sequences were extracted from the aRanTem1.1 reference genome (NCBI accession: GCF_905171775.1) [33]. Initially, the widely utilised Tandem Repeat Finder (TRF) tool [53] v.4.09, was applied with the following parameters: match (2), mismatch (5), delta (7), PM (80), PI (10), minscore (50) and maxperiod (2000). To focus on large TRs, those with an array length of ≥ 10 kb were retained, and TRs within protein-coding genes were excluded.

The monomers of all TRs were converted to dimer form, and a blastn analysis [48] (v.2.12.0., parameters: -outfmt 7 -evalue 0.000001 -word_size 10 -perc_identity 0.75 -qcov_hsp_perc 0.45 -dust no -soft_masking false) was employed for clustering into families.

The DFAM database v.3.7 [47], encompassing only the curated section, was utilised to determine TR similarity to transposable elements.

The visualisations of the results ((1) an overview of the TRs according to their characteristics and a (2) TR arrays distribution on pseudochromosomes) were performed using Python's Plotly library v.5.22.0 [54].

Compare 219A arrays and 5S ribosomal DNA (rDNA) loci

The 219A probe, 219A monomer and *R. temporaria* annotated 5S rDNA gene sequences were aligned using BLAST program v.2.12.0 [48]. This allowed us to identify the sequence, which we then used to design the probe for the 5S rDNA gene. To compare longer sequences, we aligned 219A arrays with annotated genes based on their coordinates using a custom Python script, and visualised the results using Python's Plotly library v.5.22.0 [54]. For comparison in situ, the probe (5′-ATCATTCTGAAAGC GCCCGATCT-3′) corresponding to the part of the annotated *R. temporaria* 5S rDNA gene labelled at both ends with FAM (fluorescein) was used.

Search for 47A probe in the genome

Several approaches were used to analyse the distribution of the 47A probe in the genome. TRF [53] v.4.09 (parameters as in 2.8) was used to identify TR arrays ≥ 1 kb. The subsequent part of the work was done using custom C++ scripts or Python scripts (part of the https://pypi.org/project/RepeatRanger/ framework). A cus tom C++ script was used to identify all entries of the 47A probe sequence within the genome. To visualise the spatial distribution of the 47A probe, we performed a density analysis using 10 Mb windows across all chromosomes. All occurrences of the probes for the six most abundant TRs, both in the genome and within the TR arrays (≥ 1 kb), were calculated. The 47A probe sequence entries outside of TR arrays were identified, and the most abundant nucleotides to the left and right of each entry were determined. The sequences were then iteratively extended to obtain highly conserved long sequences. These sequences were then searched across the genome. The resulting sequences were aligned with the 1000 bp flanking regions on either side. Non-conserved regions were identified and trimmed using a custom Python script. BLAST search [48] (v.2.12.0., parameters: -outfmt 7 -evalue 0.000001 -word_size 10 -perc_identity 0.75 -qcov_hsp_perc 0.45 -dust no–soft masking false) was then performed to identify similar sequences.

The structures of these sequences were analysed using Inverted Repeats Finder (IRF) v.3.08 [55] to identify inverted repeats, TRF [53] v.4.09 (parameters as in 2.8) to detect internal TR motifs, and SINEBase v.1.1 [56] to identify target site duplications (TSDs). Similarities with TEs were assessed using CENSOR from a server on the GIRI website (http://www.girinst.org/censor/index.php) [57].

Visualisation was performed using Python's Plotly library v.5.22.0 [54].

Results

TRs search in raw reads

The six most abundant TRs were found in the full dataset of raw reads from *R. temporaria* (accession: PRJNA294436; 2015), and consensus sequences for these TRs were generated using the extractTR pipeline (Table 1). The 494A TR corresponded to the previously described S1A [24]. In addition to S1A (494 bp), the homology sequence S1B (285 bp) is present in most brown frog species but not in *R. temporaria* [24]. In agreement with the literature, the S1B sequence was also not found in our analysis. To verify the extracTR results, an older approach using TAREAN software [45] was also used. To estimate the abundance of the detected TRs, extracTR counted the reads with an exact match to the most abundant k-mers and TAREAN calculated the size of the extracted graph community. Two families were detected by extracTR but not by TAREAN, and the 494A (S1A) reads per million estimate was lower in extracTR than in TAREAN. These variations can be attributed to differences in the input data: extracTR processes all raw reads, whereas TAREAN uses only a fraction (a random subset of one million reads or 2.46% of expected genome size). For further characterisation of the identified TR families and verification of metaphase chromosomal localisation, DNA probes were constructed using the consensus sequences generated by extractTR (Table 1).

The most abundant TRs from the dataset of raw reads of *R. temporaria* genome (accession: PRJNA294436; 2015) identified by the extracTR and TAREAN tools are presented in descending order of abundance. TR abundances were analyzed using two approaches: (1) extracTR processed all raw sequencing reads, with counts recalculated as reads per million; (2) TAREAN analyzed a random one million reads, reported as reads per million. The most abundant k-mers (k = 23) detected by extracTR and subsequently used as probes are shown in 5'→3' orientation. The names of the homology sequences detected by BLASTN in the core_nt database are given in brackets.

FISH chromosomal mapping of major TRs

The karyotype of *R. temporaria* is 2n = 26, with 5 pairs of large metacentric or submetacentric chromosomes and 8 pairs of small metacentric or submetacentric chromosomes [52, 58]. The six TR probes produced positive signals on several chromosomes (Fig. 1). All probes showed intense signals predominantly within the primary constrictions of all large and some small pairs of the karyotype. Signals from the 219A probe were also located within the primary constrictions of all large and two small (subtelocentric and submetacentric) chromosomes (Fig. 1). Detailed analysis of the 219A and 5S rDNA loci became possible after in silico mapping (see below, Sect. 3.4).

A distinct pattern was observed for the PCR-produced 494A (S1A) probe (Fig. 1). The signals were located on the short arms of chromosomes 2–5 and on the long arms of chromosomes 7 and 9 and, in contrast to the oligo probe, on both arms of chromosome 1, revealing two TR arrays (Fig. 1, 494A (PCR)). The same pattern of FISH signals on both arms of some chromosome pairs from PCR-produced S1 probes has been observed in other brown frog species and has been interpreted as periCEN [59, 60]. The current TR probes were also periCEN.

TR DNA is predominantly associated with heterochromatin regions; however, it has also been found in euchromatin [8, 9]. Indeed, second-order hybridisation signals were visible on chromosome arms with some probes, for example, 32A and 35B. The most prominent signals on the arms were obtained with the 47A probe. Ten out of 13 chromosome pairs displayed additional signals along the arms (Fig. 1, 47A).

The availability of the assembled genome allowed to determine the position of the TR arrays and probes in silico. However, the first step was to classify the TRs in the assembly in order to compare them with the set extracted from the raw read data.

Search for TRs in assembly and mapping on pseudochromosomes

The search for TRs is possible on the raw reads (paragraph 3.1), but the advantage of genome assemblies is TR arrays detection. 10.23% of the *R. temporaria* assembled genome was identified as tandemly repetitive DNA by TRF. The cumulative length of all large TR arrays (over 10 kb) was only 0.32% of the total genome. This small proportion of large TRs is related to a limitation of the approach (see 4.1) and obviously does not reflect the real TR content of the genome. A total of 768 TR arrays of more than 10 kb were successfully identified, 415 of which clustered into 76 distinct families (Additional file 1, Table S1). The remaining TR arrays (314) were classified as "SING", denoting singletons — individual monomers that failed to form clusters with any other monomer. Notably, 129 TR arrays were located within unplaced scaffolds, while 56 TR arrays were located within unlocalised scaffolds spanning different chromosomes. The remaining 583 TR arrays were distributed across 13 assembled chromosomes. The total length of all TR arrays was 13.3 Mb. The longest TR array was 100,622 bp (27A). Repeat unit (TR monomer) lengths ranged from 4 to 1930 bp. Four out of 6 most abundant families of TRs in raw reads were among the first 12 families of TRs in the genome assembly (Table 2, bold italic).

Some discrepancies were noted between the FISH probes (Table 1) and the new TR classification; the probe name and the TR family name sometimes did not correspond. In cases where a FISH probe was detected in more

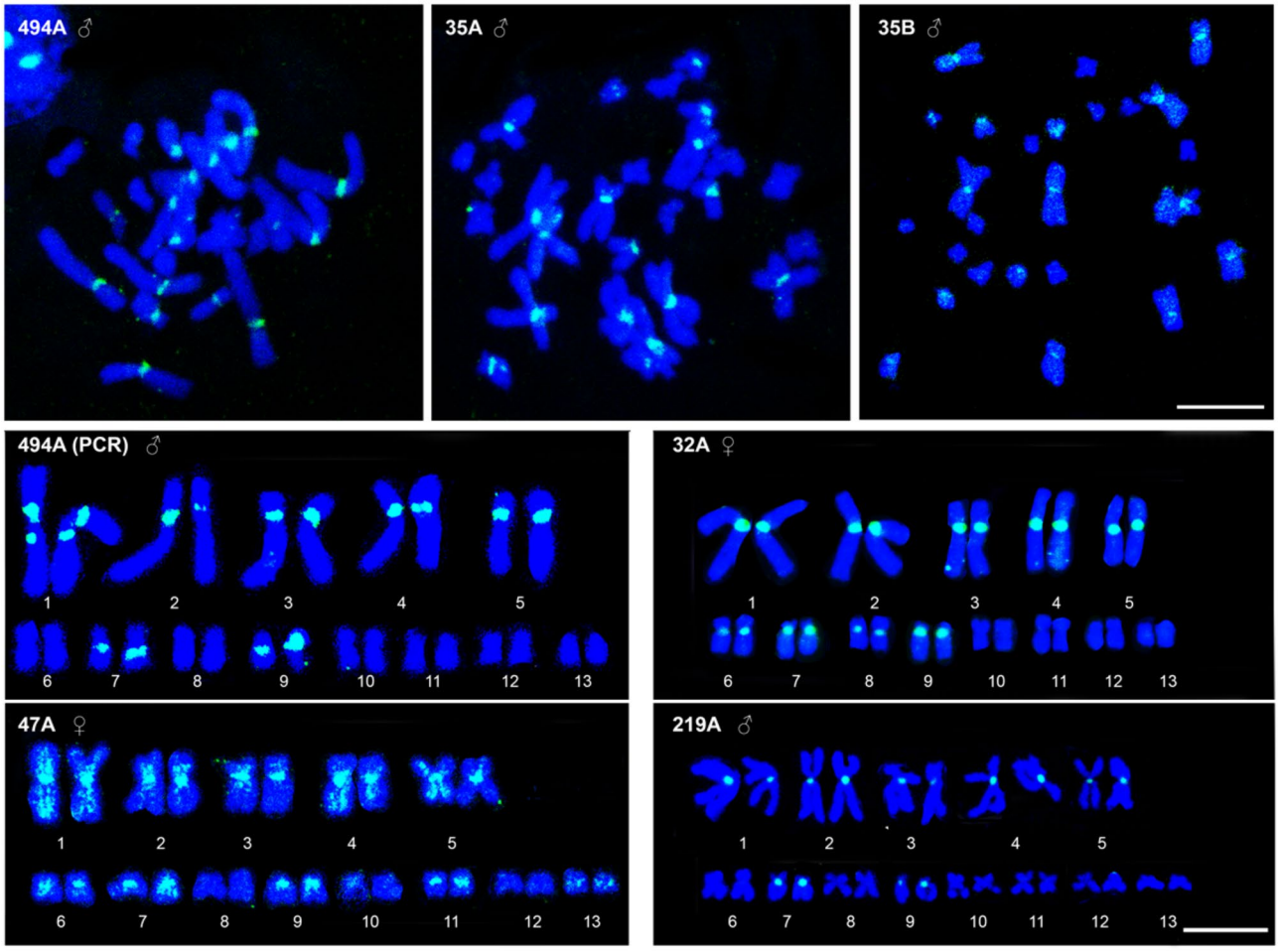

Fig. 1 *R. temporaria* metaphase plates and karyotypes after FISH with the indicated TR probes. The most frequently occurring k-mers corresponding to the most conserved parts of the monomer sequences for the six most abundant TRs were used as labelled oligonucleotide probes for in situ mapping (Table 1). The additional probe (494A PCR) was prepared from genomic DNA by PCR because of its relatively long monomer (494 bp). The probe names are given in the upper left corner for each image. Hybridisation signals (green) from all TR probes were observed in the periCEN regions of some chromosome pairs, and the chromosomes were counterstained with DAPI (blue). Bar 10 µm

Table 2 First 12 families of TRs in the *R. temporaria* genome assembly

N	TR	Number of arrays	Mean GC, %	Max_array length, bp	Chromosome	Dfam TE	% in genome
1	1149A	62	37.21	24,814	All but 9, Unpl		0.022
2	138A	35	46.36	62,988	1, 2, 3, 4, 5, 8, 9, Unpl	U5 snRNA, U4 snRNA	0.018
3	27A	29	48.4	100,622	All but 1, 3, 5, 9, Unpl		0.015
4	*35A*	*27*	*38.04*	*62,152*	*2, 4, 11, 13, Unpl*		*0.014*
5	*494A*	*21*	*46.83*	*57,784*	*1, 2, 3, 4, 5, 8, 9, Unpl*		*0.012*
6	*219A*	*20*	*50.75*	*57,710*	*8, 13, Unpl*	*5S rRNA*	*0.013*
7	44A	13	47.2	38,448	1, 2, 3, 5, 8, 10, 13, Unpl		0.006
8	31A	10	47.36	17,669	1, 2, 4, 8, 12, Unpl		0.004
9	140A	9	56.47	19,368	1, 2, 3, 4, 5, 12, 13		0.003
10	44B	9	41.74	44,868	1, 4, 5, 6, 8, 9, 10		0.004
11	1099A	7	48.83	32,032	3, 4, Unpl	tRNA	0.003
12	*32A*	*7*	*44.55*	*77,222*	*1, 3, 4, 6, Unpl*		*0.004*

GC% - average GC content in TRs in percent; TR similarity by Dfam database; TR chromosome positions predicted in silico. TRs for which FISH was performed on metaphase plates (Fig. 1) are shown in *bold italic*. Unpl – unplaced scaffolds

than half of the TR arrays within a family, the entire family was named accordingly. The 47A probe occurred in three TR arrays (≥ 10 kb) across different families – 27A, 54A and "SING". Within these families, the 47A probe occurred in either single or double copies per TR array. Consequently, naming these families based on the probe was impractical. Similarly, the 35B probe was detected in a single TR array, albeit with a significant copy number of 62. The formation of a family around a single TR array also seemed to be impractical. The search in the Dfam database did not reveal any similarity between TR consensus monomers and TEs. Some TR families showed chromosome specificity in silico. But most of them were also located in the unplaced scaffold and their localisation on other chromosomes cannot be ruled out.

An overview of the identified TRs is shown in a 3D-graph by monomer length, the degree of monomer similarity within the family and GC content (Fig. 2). An interactive image with a description of each family can be found in Additional file 2. Note that the majority of TR arrays were located in the region of relatively short monomer units, although some were scattered throughout the entire plot, including the region of long monomer units (Fig. 2). Some of the arrays in the region of the long monomer units corresponded to TRs of n higher order repeats (HORs), i.e. copies consisting of TRs of n head-to-tail monomers (for example, see 44A and 27A in Additional file 2). TR monomers within the families varied considerably in length, GC content and nucleotide composition (Fig. 2). The 494A (S1A) arrays were also not uniform (Fig. 2, 494A).

The assembly was expected to have many gaps, which are usually associated with large TRs [61, 62]. Indeed, numerous gaps were observed in this context, as visually depicted (Fig. 3a). The largest number of gaps was observed in terminal regions of chromosomes, indicating an enrichment of subTEL regions with repetitive sequences. An incorrect orientation of the chromosome arms also cannot be ruled out.

The TR 10 kb arrays were mapped on chromosome scaffolds to analyze the distribution of different TR families along the genome (Fig. 3b and c, Additional file 3 Figure S1, see Additional file 4 for interactive image). Terminal (subTEL) regions of almost all chromosomes were enriched with TRs (Additional file 3 Figure S1, Additional file 4), as expected. Most of these were variable TRs classified as "SING". Some TR arrays were also localised throughout the chromosome arms.

TRs localised by FISH (Fig. 1) were mapped separately on the assembly (Fig. 3c). 47A and 35B were not mapped on pseudochromosomes because corresponding TR families with arrays longer than 10 kb were not detected (see above). The main discrepancies between in situ and in silico observations were the clear periCEN localisation of all mapped TR families and an excess of periCEN signals for almost all TR families in situ as compared to in silico predictions. For example, the 35A and 32A probes were located in situ on all large chromosomes (Fig. 1). The in situ mapped TR families that were found in the genome assembly were also located in silico in the unplaced scaffold (Table 2), and their location on other chromosomes was quite expected. In some cases, signals predicted in silico were not observed in situ. For example, signals from the 219A probe were observed on seven chromosomes in situ but not on chromosome 13 (Fig. 1, 219A), whereas in silico signals were predicted on only two chromosomes, including chromosome 13. These results may reflect incomplete and incorrect genome assembly, especially for chromosomes 2 and 13.

PeriCEN position

The large TR arrays are generally extremely poorly represented in reference genomes, including the *R. temporaria* genome, reflecting the complexity of assembling these types of sequences. The assembly with many gaps

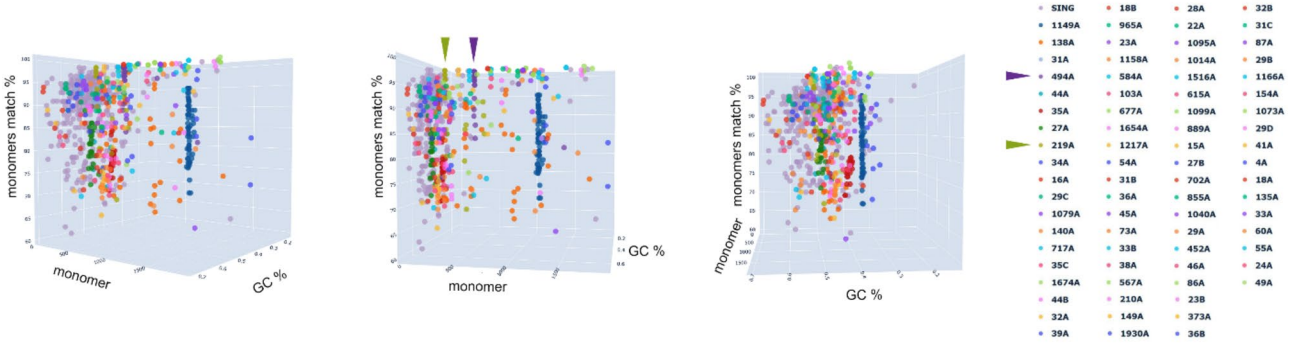

Fig. 2 Three-dimensional analysis of *R. temporaria* TR array properties. Graphical representation shows: X - monomer length (bp); Y – GC content (%); Z - degree of monomer similarity in the array (match %). Each data point represents a TR array, with color-coding by TR family. The colour legend is on the right. Arrows highlight cloned 494A (S1A) TR, the only previously characterized TR in *Rana* species (purple) and 219A TR family (lime green)

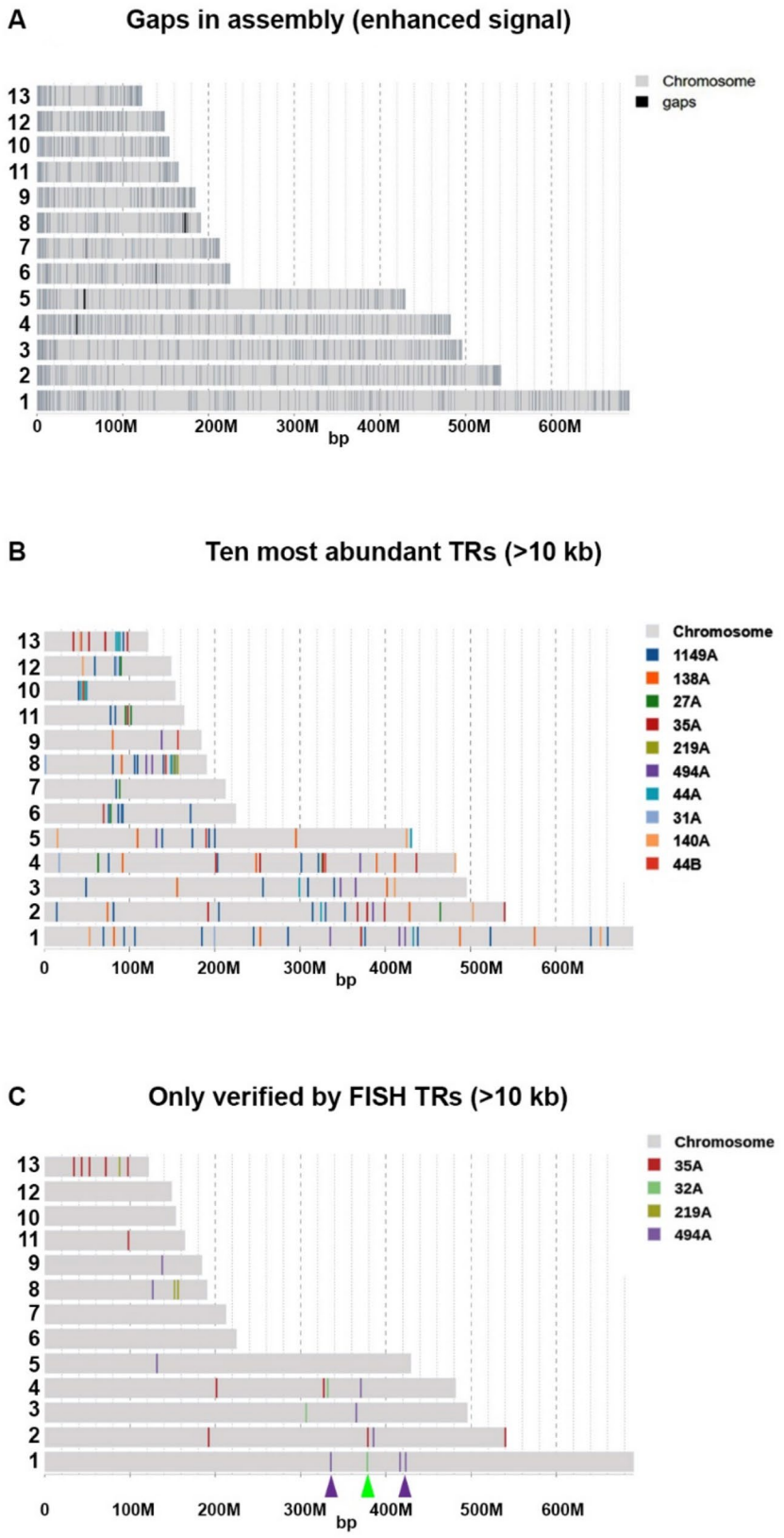

Fig. 3 Genomic landscape of TRs across *R.temporaria* pseudochromosomes. Each horizontal row represents one of 13 pseudochromosomes, showing: **a** - Assembly gaps: black lines indicate gaps in assembly; **b** - Ten most abundant TR families: colored lines denote TR arrays (see color legend at right); **c** - FISH-verified TRs: experimentally confirmed TRs (excluding 47A and 35B, which lacked arrays > 10 kb). Arrows (color-coded by TR family) mark periCEN position on pseudochromosome 1

(Fig. 3a) provided limited data for the repetitive sequence analysis. Nevertheless, it was still possible to make some predictions. PeriCEN regions were enriched for TRs, which were identifiable in genomic assemblies by analysing sequence features such as high repeat density, e.g., chromosomes 1, 6–8 and 10–13 (Additional file 3 Figure S1). No significant enrichment of TRs in the putative periCEN regions was observed on the remaining chromosomes. These results indicate an insufficiency of genome assembly, as the periCEN regions of all chromosomes are the most prominent C-bands in the grass frog [58].

Since the signal from the long 494A probe was located on the short arm of chromosome 5 in situ (Fig. 1, 494A (PCR)), it was assumed that pseudochromosome 5 was upside down. An in silico prediction also located several 494A arrays to the small chromosomes 8 and 9, whereas in situ signals were located in chromosome 7 and 9. These chromosomes are similar in size and could not be separated during sorting. The same was true for 219A: in silico analysis localised the arrays to chromosome 8, while the in situ signals were localised to small chromosomes 7 and 9. Taking into account the discrepancies between in situ (Fig. 1) and in silico (Fig. 3c) signals for 494A and 219A probes (Fig. 1), as well as the position of periCEN markedly enriched with TRs (Additional file 3 Figure S1) on pseudochromosome 8, it can be assumed that pseudochromosome 8 corresponds to chromosome 7 of the karyotype.

Discrepancies between in silico and in situ data at poorly assembled areas are expected. The only case of complete agreement between in silico and in situ data was observed for the long 494A probe (all large and two small chromosomes were labeled with the probe). TR 32A was located in between two 494A arrays (Fig. 3, b, c). Based on the in silico prediction (Fig. 3), verified by in situ FISH (Fig. 1), it can be concluded that the periCEN of chromosome 1 is localised in the region of 338–422 Mb (Fig. 3, c. arrows).

219A and 5S rDNA attitude

In light of the similarity between the 219A and 5S rDNA sequences (Tables 1 and 2), the sequences were compared in detail. The 5S rDNA consisted of multiple copies of a highly conserved 119-bp coding sequence and a nontranscribed spacer (NTS) of variable length and nucleotide composition (Fig. 4a). The 219A probe, defined as the most abundant k-mer of the 219A TR (Table 1), was placed within the NTS region (Fig. 4a, red). The 219A probe-mapped signals were observed in silico on pseudochromosomes 8 (chromosome 7 of the karyotype) and 13 (Fig. 3c). Comparison of the long arrays of 219A in the genome assembly with the annotated 5S rDNA genes showed that these sequences were located in the same regions of the pseudochromosomes and partially overlapped (Fig. 4b, see Additional files 5 and 6 for interactive images). The 219A array on pseudochromosome 8 largely overlapped with numerous annotated 5S rDNA

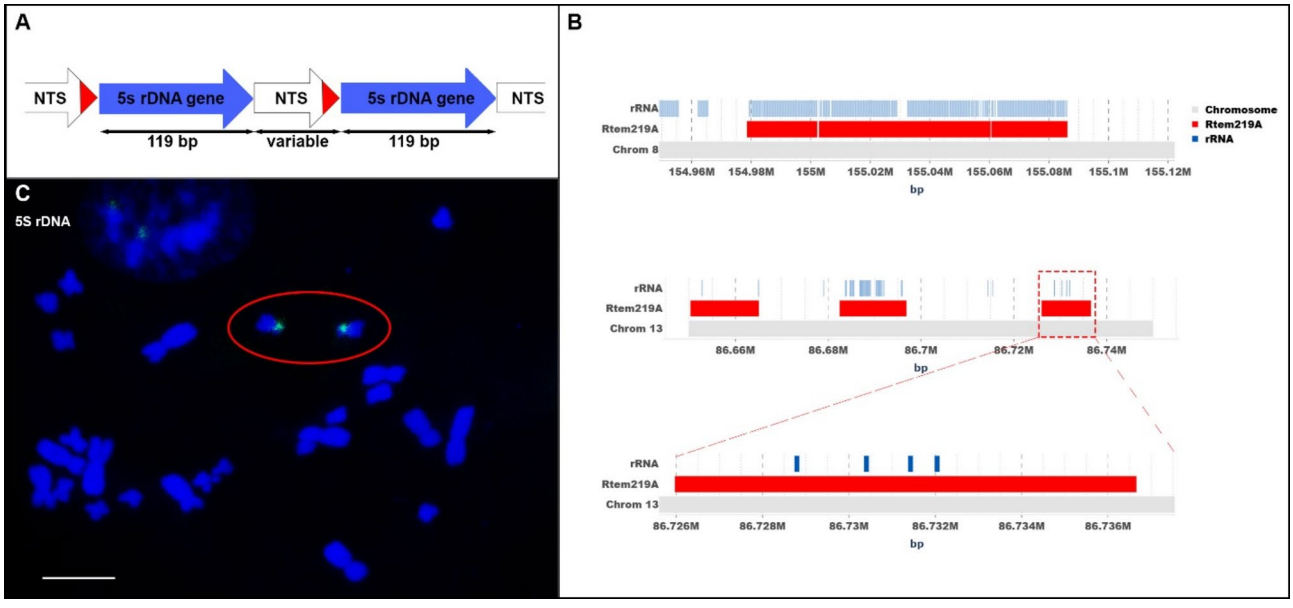

Fig. 4 Genomic relationship between 219A TR and 5S rDNA in *R. temporaria* genome. **a** – 5S rDNA organization: diagram of tandem repeat structure showing 119-bp coding sequences (blue arrows) and nontranscribed spacers (NTS, white arrows). The 219A probe target within NTS is highlighted (red); **b** - Comparative genomic mapping: gray rows represent pseudochromosomes with 5S rDNA loci (blue) and 219A arrays (red) (see legend); **c** - Cytogenetic validation: FISH of 5S rDNA gene probe (green signals) on metaphase chromosomes (DAPI counterstain, blue). The single chromosomal pair containing 5S rDNA is indicated (red oval). Scale bar: 10 μm

genes, whereas the 219A arrays located on pseudochromosome 13 differed significantly from the standard 5S rRNA sequence, but contained the NTS portion (the 219A probe) and a smaller amount of 5S rDNA gene in this region (Fig. 4b).

Since earlier cytological studies detected the single 5S rDNA site using the probe containing the potentially transcribed region of the 5S rDNA [63], localisation was tested using an oligonucleotide corresponding to the 5S rDNA gene. This 5S rDNA probe stained the single chromosome pair (Fig. 4c), in agreement with the previous finding. However, the hybridisation pattern was not consistent with that obtained with the 219A probe (Fig. 1). It is assumed that the signals from the 5S rRNA probe detected in previous studies [63] and in the present study (Fig. 4c) correspond to the long arrays detected in in silico on pseudochromosome 8 (chromosome 7 of the karyotype) (Figs. 3b and c and 4b). In contrast, the 219A TR family with similarity to 5S rDNA (Tables 1 and 2) hybridised to the periCEN regions of several chromosomes - all large and two small (Fig. 1, 219A), i.e. the 219A probe showed a similar pattern to ordinary TR (Fig. 1). PeriCEN regions, with the exception of chromosome 7, were not detected by the 5S rDNA gene probe (Fig. 4c). Thus, despite the observed similarity between the 219A sequence and 5S rDNA (Tables 1 and 2), the 219A arrays belong to the other TR sequence type.

47A is a part of a new dispersed element - frog element dispersed organised repeats (FEDoR)

An intriguing phenomenon related to the 47A probe was observed in TR arrays derived from the genome assembly. Although the probe produced a strong signal in FISH (Fig. 1, 47A) and was identified as one of the most represented TRs in raw reads (Table 1), it was present in only three long TR arrays and only one or two TR array copies. Searching the genome assembly for the 47A probes revealed that they occur at a high number of loci dispersed throughout the genome (Fig. 5a). The average number of copies per 10 Mb was calculated to be 106, with a minimum of 4 and a maximum of 589. The highest densities of 47A probes were found in the terminal regions (subTEL) of the majority of chromosomes (Fig. 5a).

As the 47A probe was detected at very low copy numbers in large TR arrays (≥10 kb), a search for shorter arrays (≥1 kb) containing 47A probe was then performed. Short tandem repeat arrays (shTR47) were mapped on pseudochromosomes (Fig. 5b, red). 27 arrays were longer than 3 kb and some of which were located in the periCEN regions (Additional file 3 Figure S2). Short arrays were also scattered throughout the genome, with notable enrichment in subTEL regions (Fig. 5b, red). PeriCEN localisation was also observed on 7 of the 13 pseudochromosomes (Fig. 5b, red), in agreement with the periCEN signals (Fig. 1, 47A).

The identified short arrays with the 47A probe could still not explain its wide distribution of the 47A probe across the genome (Fig. 5a). Examination of the frequency of probes in the genome and their occurrence in TR arrays ≥1 kb identified probe 47A inside TR arrays in only 20% of the cases and outside TR arrays in 80% of the cases (Table 3). The 35B probe showed the same tendency, albeit with a less dramatic dichotomy. Both probes showed additional hybridisation signals outside periCEN regions (Fig. 1, 47A, 35B). Thus, fractions of 47A and 35B sequences (~20–40%) form TR arrays and produce periCEN signals, while the remainder of the probes labelled an unknown element scattered throughout the genome.

Probe 47A was recognised as one of the most abundant in the genome (Table 1) and poorly represented in TR arrays (Table 3). Conserved sequences of ~1 kb in length were found by analysing nucleotide neighbourhoods characteristic of the genomic environment in which 47A probes were localised. Regions of homology were revealed by alignment of the ~1 kb flanking regions on either sides. Overall, 43 highly conserved sequences were identified (Additional file 3 Figure S3). The average size of an element, which was termed Frog Element Dispersed organised Repeats (FEDoR), was ~3.5 kb. Almost all of the elements displayed similar organisations (Fig. 6 and Additional file 3 Figure S4). All FEDoRs contained four inner TR motifs. The 1st and the 2nd repeats were formed of a shTR47 family monomer and were separated from each other by a short insert (Fig. 6). On average, a 22-bp-long TR monomer occurred 41 times in the 1st repeat and 8 times in the 2nd repeat. The 3rd and the 4th repeats differed from shTR47 and from each other; their monomers occurred on average 20 and 5 times, respectively. These repeats have no similarities with any of the TR families found with arrays larger than 10 kb (Table S1). The FEDoR core consisted of internal TR motifs flanked on either side by pairs of inverted repeat sequences (IRSs) 91 and 49-bp-long, respectively. Among the 43 FEDoR sequences, the starting IRSs had a similarity of 94%, and the ending IRSs of 84%. The starting IRSs were close together, while the ending ones were typically separated by a distance of 292 bp. Additionally, TSDs were identified at both ends of the FEDoR sequences; TSDs were not the same for the different elements and may reflect the difference of the insertion sites. Still, TSDs were identified at both termini of all FEDoR elements. In most FEDoRs, 47A probes with complete matches occurred only once in each shTR47 repeat. The distance between 47A probes inside the TR array was identical for all elements. The short insertion in the shTR47 array was also of

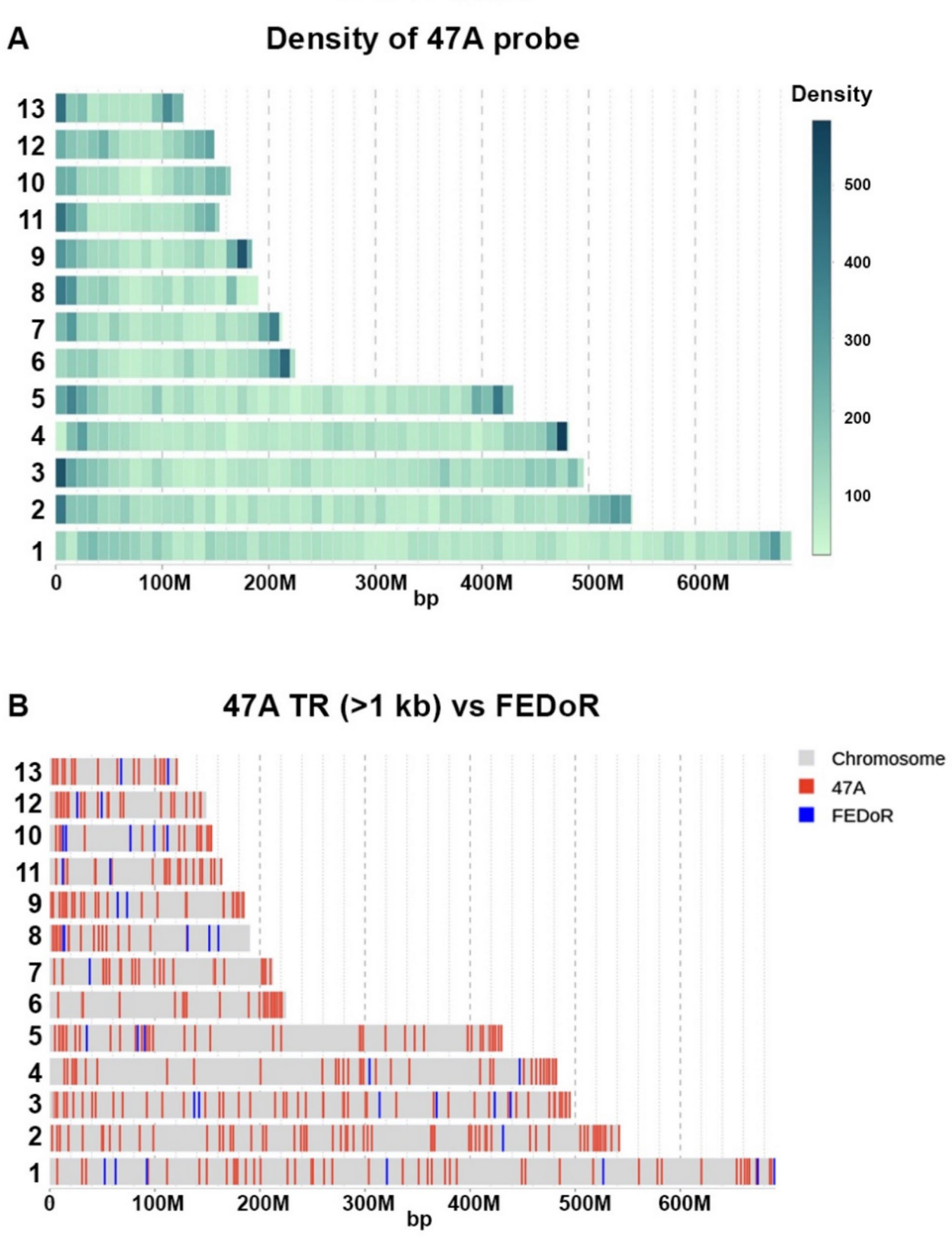

Fig. 5 The location of 47A probe in the genome. **a** - Density distribution: 47A probe frequency shown in 10 Mb windows across pseudochromosomes (horizontal rows). Color gradient indicates density from low (pastel mint) to high (navy blue), with exact counts per window shown in legend. **b** - Repetitive DNA elements containing 47A probe. Short 47A arrays ≥ 1 kb (red), DNA-TE FEDoR with internal TR motifs (blue)

Table 3 Frequency of TR probes in TR arrays ≥ 1 kb

Name	Number of occurrences in the genome, counts	Number of occurrences in TR arrays ≥ 1 kb, counts	% of probes in TR arrays
494A	1052	1031	98.00
219A	2971	2962	99.70
32A	2774	2746	98.99
35A	10,185	9894	97.14
35B	33,369	16,099	48.25
47A	44,003	9107	20.70

Quantification of probe occurrences for the most abundant tandem repeats (TRs) identified in raw sequencing reads: (1) total occurrences in the genome assembly, (2) occurrences within TR arrays (≥ 1 kb), and (3) percentage of probes located within TR arrays

constant length (Fig. 6, seq0) and AT-rich (72%); while the FEDoR AT content was 52%, and AT content for the whole genome was 56%. In some cases, the short insert was recognised as a reverse IRS similar to the initial reverse IRS (Additional file 3 Figure S5); in the remaining elements, the inserts are degenerate reverse IRSs with different degrees of similarity. Such a structure may indicate a composite FEDoR origin.

Alignment of all elements revealed the same structure and orientation of flanking regions (Additional file 3 Figure S6) and sequences in between TR motifs (Additional file 3 Figures S7, S8, S9), highlighting the regularity of the organisation pattern. No extended regions with significant similarity to any known TE were found in FEDoR (Additional file 1 Table S2). In addition, even an incomplete open reading frame (ORF) was also absent. In the flanking region, there were two short areas of similarity with DNA-TEs: ~ 150 bp with hAT and ~200 bp with Mariner (Additional file 1 Table S2). FEDoR is the definite dispersed element, as shown in silico (Fig. 5a) and in situ (Fig. 1), with a distinct structure enriched with TR motifs and with TDSs and IRSs at precise positions (Fig. 6), which could be the evidence of a former transposition.

Discussion

Genome-wide detection of TRs

Repetitive DNA sequences are the most difficult part of the genome to assemble and annotate [61, 62, 64, 65]. TR DNA is characterised by tandemly arranged repeat copies that form contiguous arrays up to megabases in length that cannot be efficiently resolved with short-read sequencing [64–66]. Even the rapid development of genome sequencing and assembly techniques has not provided a complete picture of the composition of CEN/periCEN and subTEL regions [67–70]. These are usually underrepresented in genome assemblies [62, 71, 72]. Tandemly repetitive DNA occupies 10.23% of the *R. temporaria* genome as identified by TRF. This proportion is consistent with the amounts typically observed in other amphibian genomes using bioinformatic approaches. Tandemly repetitive DNA makes up approximately 9.47% of the genome of the closely related species *R. kukunoris* [73]. However, this proportion includes low complexity and simple repeat sequences, short arrays, tandemly or segmentally duplicated genes. The cumulative length of all large TR arrays (over 10 kb) was only 0.32% of the total genome. Many of large TR arrays are found on the unplaced scaffold (Table 2, Table S1) or near gaps in pseudochromosomes (Fig. 3a, b), highlighting the difficulty of accurately assembling the regions rich in repetitive sequences.

Nowadays, the assemble-free approaches, including TAREAN [45], k-mer frequency statistics from unassembled sequence reads are widely used [22, 74]. These approaches are more robust for detection of highly abundant TRs and for monomer reconstruction, but are not suitable for detection of low-copy TRs, simple repeats and TE-based short dispersed TR arrays. Additional challenges are posed by low genome coverage. TAREAN requires only 0.01-0.50X genome coverage to detect TRs. It is noted that TAREAN is sensitive to abundant TR sequences, while it may miss less common

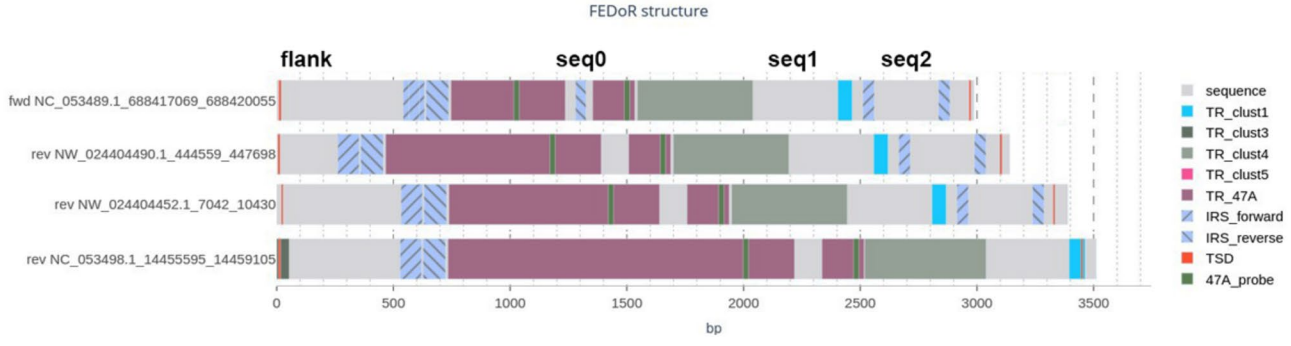

Fig. 6 Structure of the FEDoR elements. The FEDoR elements contain internal TR motifs of 47A (purple, 100% similarity with the 47A probe is shown in green), two other internal TR motifs (khaki and blue), two pairs of inverted repeat sequences (IRS) (shaded blue) and target site duplication (TSD) (red). Other conserved regions of FEDoR elements, namely flanking regions (flank), sequences between TR motifs (seq0 and seq1), and sequences between terminal IRSs (seq2), are labelled at the top of the figure

or difficult-to-detect repeats formed by short tandem arrays [45]. The requirement for minimal input data is offset by the high computational resources and time. In contrast, extracTR can process the entire batch of reads within a reasonable amount of time and may provide more accurate results. This work demonstrated that some TR sequences (47A, 35B) were undetectable by the TAREAN tool (Table 1), probably due to low input data. Using the entire batch of reads as input for the extracTR tool enabled the detection of two TRs that were undetected in the TAREAN analysis (Table 1). The reliability of the bioinformatics-based identification of TRs from raw data was experimentally confirmed by successful FISH detection of all 6 TRs on *R. temporaria* chromosomes (Fig. 1). However, a large number of *R. temporaria* TR families identified in the genome assembly (Table S1) were not detectable by assemble-free approaches, likely due to their monomer size and low copy number. This highlights the need to use all available tools to exhaustively characterise repetitive elements in a new genome.

R. temporaria TRs

A total of 76 TR families and 314 single TR arrays were identified in the grass frog genome (Fig. 2; Table S1). TRs with shorter monomers (< 100 bp) generally predominate over long ones (> 1000 bp) (Table S1, Fig. 2). In *Proceratophrys boiei*, one of the very few Anura species for which a genomic analysis of TR content and subsequent FISH mapping were performed, the monomer sizes of highly abundant CEN/periCEN TRs were about 170–180 bp [22]. A more recent study of this research group has shown numerous other TRs with short monomers (< 100 bp) in the *P. boiei* genome [75]. Therefore, TRs with short monomer lengths may be a distinctive feature of Anuran genomes.

According to the classical view of TR, the size of the monomers can correspond to the size of a nucleosomal DNA (about 140–170 bp) or two nucleosomal unit lengths (about 340 bp) [76, 77]. However, it is not a universal rule, as the sizes of TR monomers are not conserved [78]. Most TR families identified in the current study did not have the monomer size of the centromere protein A (CENP-A) nucleosome, except for 138A, 140A and three low-copy TRs (149A, 154A, 373A). In fact, none of the TRs with the possible CEN localisation (Fig. 1) had a monomer size typical for a single nucleosome. The FISH technique did not enable clear identification of probe localisation in metaphase chromosomes (distinguish CEN from periCEN locations) and none of the probes produced signals on all chromosomes; therefore, their positions are described as periCEN. All in situ-mapped TRs stained only a subset of chromosomes (Fig. 1). A similar situation was shown for CEN RrS1-like satDNA of water frogs (genus *Pelophylax*) [79, 80] and for centromere repeat 1 (Fcr1) in *X. laevis* [81]. It is known that some species, such as *Drosophila*, chicken and *X. laevis*, do not contain CEN sequences common to all chromosomes [82–84]. In *X. tropicalis*, a 205-bp consensus monomer was placed as a CEN TR in all chromosomes. A 205-bp TR was annotated as CEN based on its ability to precipitate with CENP-A and central chromosome position in silico; no functional test with artificial chromosomes has been reported [85]. The distinction between CEN and periCEN TR requires a separate study.

The monomer lengths of the 14 frog TRs exceeded 1000 bp (Fig. 2; Table S1). Long monomers often originate from parts of TEs, such as the LTRs (long terminal repeats) and UTRs (untranslated regions) [86]. The presence of TRs with large monomers (> 1000 bp) is a feature of avian genomes [87, 88] and has been observed in other groups, such as Tenebrionidae beetles [89], some plant species [90], *Megaleporinus elongatus* fish [91]. The largest monomer size was identified for BamHI-800 sequences in *Bufotes* [92]. TR monomers of longer length had not been identified in amphibians before the current study.

None of the identified TRs showed any homology with TEs (Table S1). However, some TR families shared a certain degree of similarity with transfer RNA (tRNA) and small nuclear RNA (snRNA) (Table 2, Table S1). Most non-autonomous non-LTR retrotransposons SINEs (short interspersed nuclear elements) are derived from tRNA, 7SL RNA (signal recognition particle RNA), or 5S rRNA [93]. SINEs derived from either the U1 or U2 snRNA have been reported among crocodilians [94]. Some amphibian TRs with similarities to tRNAs are known (e.g. OAX repeat in *Xenopus* [95], Rana/polIII in *Rana esculenta* [96], PolIII/TAN in the newt *Cynops pyrrhogaster* [97]) and their SINE-derived origin has been suggested [95, 96] but not yet confirmed. Interestingly, SINEs seem to be underrepresented in amphibian genomes (0.01–2.69% of genome) [98]. SINE-derived TRs have been described in scaled reptiles, but no significant correlation has been found between the abundance of SINEs and the presence of SINE-derived TRs in genomes [99]. However, SINEs are indeed more abundant in scaled reptile genomes (1.4–6.9%) [100] than in amphibians. In addition, it is worth noting that, the identification of new SINEs is challenging due to their weak structural signals and rapid sequence diversification and requires specialised approaches [56, 101, 102], and therefore the actual SINE content in genomes may be underestimated by tools such as RepeatMasker [102]. Therefore, TE-based TRs in the grass frog genome cannot be excluded, and the careful verification of such TRs existence is the subject of future work.

In eukaryotic genomes, the multigene families for rRNA genes are tandemly arrayed in clusters. In the *R.*

temporaria genome, the only 5S rDNA site located close to the CEN on the short arm of chromosome 7 was found by FISH, with a probe that contained the potentially transcribed region of the 5S rDNA [63]. The current work also confirmed this observation by FISH with a probe that corresponded to the annotated 5S rDNA gene sequence (Fig. 4c). However, in this study we found that the signals from the 219A probe corresponding to the NTS sequence occur on more chromosomes (Fig. 1). Based on the FISH results (Fig. 1, 219A) and the comparison of 219A arrays with annotated 5S rDNA genes (Fig. 4b), we suggest that 219A is a satDNA (non-coding TR) derived from 5S rDNA. SatDNA sequences derived from rDNA have been described for some plant [103, 104], fish [105, 106] and frog species [107]. In frogs, PcP190 satDNA derived from 5S rDNA is assumed to be ancient and has been described in Leptodactylidae, Hylodidae (Hyloidea) [107–109]. The PcP190 repeat unit consists of a conserved region that is highly similar to the 5S rDNA gene, and a hypervariable region [108, 109]. No similarity was found between any NTS of 5S rDNA and hypervariable regions of PcP190 satDNA [108]. But probe 219A based on the most conservative part of the TR family (see Materials and Methods 2.4), corresponded to the NTS portion (Fig. 4, a, b). Taken together, the 219A associated with the periCEN regions is other satDNA (TR) derived from 5S rDNA loci.

The chromosome assignments (Table 2, Table S1) and the locations of the TR arrays (except for arrays from unplaced scaffolds) on the chromosomes (Fig. 3b, Additional file 3 Figure S1) were determined for all TR families. However, it should be noted that several variants of *R. temporaria* karyotypes have been published [52, 58]. Differences between karyotypes concerned chromosomes 7–9. These chromosomes are of a similar size and are therefore difficult to resolve by sorting, and some pseudochromosomes may not correspond to karyotype chromosomes. For example, pseudochromosome 8 corresponds to chromosome 7 of the karyotype (see Sect. 3.4 for details).

A significant number of TR families were predicted in the distal subTEL regions in silico (Fig. 3b). The subTEL region is supposed to be the repository of TRs to be spread over the periCEN region during speciation [110]. A clear preferences of TRs in subTEL regions was observed (Fig. 3b, Additional file 3 Figure S1), although these TRs were not among the most abundant in raw reads. *X. tropicalis* featured a high density of TR in subTEL regions and unusually long subTEL, which may indicate that unequal crossing over during meiotic recombination mediates TR expansions in these highly recombinogenic chromosome regions [85]. The enrichment of TRs in subTEL regions is likely a common phenomenon in the amphibian karyotypes.

In silico analysis of the genome assembly revealed some TRs in the chromosome arms, i.e., euchromatin part of the genome not corresponding to the heterochromatic regions detected by C-banding (Fig. 3b). Euchromatic TRs have been found in several species [9, 111, 112] and have been suggested to play a role in gene expression modulation [14, 113]. The functional significance of the TRs along chromosome arms requires in depth research [14].

FEDoR

In contrast to mammalian genomes, DNA-TEs predominate over retrotransposons in amphibian genomes [29], including in *R. temporaria* (Additional file 3 Figure S10). However, our knowledge of TEs in amphibians remains fragmentary and progressing slowly, as can be seen from the abundance of unknown TEs (Additional file 3 Figure S10, gray).

We suppose, that the FEDoR observed in the current work is the new non-autonomous DNA-TE. Its structure is reminiscent of that described for the miniature inverted-repeat transposable elements (MITEs). MITEs were first described in the maize genome and are the most abundant group of DNA-TEs [114]. MITEs tend to be small in size (~100 to 800 bp), lack protein coding potential, are interspersed and can reach high copy numbers with high uniformity between copies [115]. MITEs are non-autonomous, truncated versions of autonomous DNA-TEs and their transposition requires the activity of a class II DNA transposase acting via a cut-and-paste mechanism. However, this mechanism cannot account for the high copy number of MITEs in genomes and the mechanism by which MITEs amplify remains unknown. MITEs have the structural characteristics of a typical DNA-TE, with conserved IRSs flanked by TSDs. A peculiar Tc1/mariner MITE called miDNA4-Xt from *X. tropicalis* exhibits typical MITE features but is unique in that it contains a TR motif. Most (61%) miDNA4s contain only one TR motif, although some contain multiple TRs [116]. Notably, miDNA4 sequences with multiple TR motifs localize to subTEL regions, similar to the 47A probe (Fig. 5a and b).

Despite the similarity of some of its features to those of MITE, FEDoR is suggested here to be an entirely separate new DNA-TE element. Its peculiarities include its size (~3.5 kb), which exceeds the size of miDNA4 by no less than 5 times (300–600 bp). In addition, it is not AT-rich, with 52% AT as compared to 56% AT in the *R. temporaria* genome, while miDNA4 is comprised of 63.0% AT compared to a whole-genome AT content of ~60% in the *X. tropicalis* genome. Further, FEDoR could not be attributed to Tc1/mariner or any other known TE family, as the extended regions of significant similarity are absent. Moreover, FEDoR bears three conservative regions

around TR arrays which lack similarity with any known elements. Future analysis of these sequences may lead to identification of new autonomous DNA-TE, as the non-autonomous DNA-TEs of the same class generally contain the remnants of the helper. Lastly, the principal organisation scheme of MITE is [IRS forward - $(TR)_n$ – IRS reverse], while FEDoR is built of [IRS forward – IRS reverse - $(TR)_n$- IRS forward – IRS reverse]. Namely, the essential difference lies in the duplication and proximity of the forward and reverse IRS sites. This feature could influence the transposition mechanism.

DNA-TEs containing TR motifs lead to consideration in the evolutionary realm, regarding the TR enrichment in genomes [116, 117]. The links between DNA-TEs and TRs are supported by the current work. The combination of a DNA-TE and TRs could be a driving mechanism to accumulate DNA and increase genome size. Both MITE and FEDoR can use similar mechanism for the dissemination of TRs.

Conclusions

This study characterised the TRs content in *R. temporaria*, a representative of one of the largest Anuran families, Ranidae, for which a complete dataset had never previously been explored. The results confirmed that the grass frog genome has a great diversity of TRs. Current work provides number of TR monomer consensus sequences, which will be useful for the future investigations into the significance, origin, and evolution of TRs in anurans during speciation. Despite the high quality of the assembly, several distinct inaccuracies were noted. The identification of FEDoR, in addition to the previously described miDNA4-Xt (MITE), provides further evidence of relationships between TEs and TRs in amphibians. Future studies investigating associations between TRs and TEs across species could reveal some new aspects of the mechanisms of genome evolution and function and could provide some new explanations for the variation in genome size in amphibians.

Abbreviations

7SL RNA	Signal recognition particle RNA
CEN	Centromere
CENP-A	Centromere protein A, centromere-specific histone H3 variant
DNA-TEs	DNA transposons
FAM	Fluorescein
FEDoR	Frog Element Dispersed organised Repeat
FISH	Fluorescence in situ hybridization
HOR	Higher order repeat
IRF	Inverted Repeats Finder
IRSs	Inverted repeat sequences
LTRs	Long terminal repeats
MaSat	Murine major satellite
MITE	Miniature inverted-repeat transposable element
NTS	Non-transcribed spacer
ORF	Open reading frame
periCEN	Pericentromere
rDNA	Ribosomal DNA
satDNA	Satellite DNA
shTR	Short tandem repeat array
SINEs	Short Interspersed Nuclear Elements
SING	Singleton
snRNA	Small nuclear RNA
subTEL	Subtelomere
TAREAN	Tandem Repeat Analyser
TEs	Transposable elements
TRF	Tandem repeat finder
tRNA	Transfer RNA
TRs	Tandem repeats
TSDs	Target site duplications
Unpl	Unplaced scaffolds
UTRs	Untranslated regions

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11643-5.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Acknowledgements

Technical resources of the Group of Confocal Microscopy and Image Analysis at the Institute of Cytology RAS were used. We are grateful to T.K. Yakovleva, R.A. Pasynkova and S.N. Litvinchuk for the helpful tips for preparing metaphase chromosomes and to V.A. Dikaya for the initial TE analysis.

Author contributions

OP conceived the study. MP and AK developed software and performed bioinformatics analysis. AT performed the wet laboratory work. AT, OP and MP interpreted the data and prepared the manuscript and Supporting Information. DO critically read and commented on the article. AT, OP, DO and MP contributed to revising the manuscript for resubmission. All authors read and approved the final version of the manuscript.

Funding

This work was supported by the Russian Science Foundation (project no. 24-24-00480).

Data availability

The following open access data were used in this study: the raw sequence data from the National Library of Medicine (NCBI) with BioProject accession number PRJNA294436 and the assembled genome of R. temporaria from the National Library of Medicine (NCBI) with NCBI RefSeq assembly accession number GCF_905171775.1. All datasets generated and analysed during the present study, namely (1) tandem repeat (TR) sequences, (2) FEDoR sequences, and (3) FEDoR alignments, (4) custom scripts are available on GitHub (https://github.com/non-coding-DNA-lab/TR_Rana_temporaria) and on Zenodo (https://doi.org/10.5281/zenodo.15251988). extracTR used in the current study is also available on GitHub: https://github.com/aglabx/extracTR.

Declarations

Ethics approval

All the applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All the procedures performed in the studies involving animals were consistent with the principles of the Basel Declaration and the position of the Animal Ethics Committee of the Institute of Cytology RAS (Assurance Identification number F18-00380; approval date: 8 August 2023, protocol # 14/23). This study did not involve endangered or protected species, and all the specimens were collected outside of natural reserve areas in the Russian Federation.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 27 July 2024 / Accepted: 25 April 2025
Published online: 06 May 2025

References

1. Biscotti MA, Olmo E, Heslop-Harrison JS. Repetitive DNA in eukaryotic genomes. Chromosome Res. 2015;23:415–20. https://doi.org/10.1007/s10577-015-9499-z
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921. https://doi.org/10.1038/35057062
3. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82. https://doi.org/10.1038/nrg2165
4. Plohl M, Luchetti A, Meštrović 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. https://doi.org/10.1016/j.gene.2007.11.013
5. Kit S. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. J Mol Biol. 1961;3:711–IN2. https://doi.org/10.1016/S0022-2836(61)80075-2
6. Komissarov AS, Gavrilova EV, Demin SJ, Ishov AM, Podgornaya OI. Tandemly repeated DNA families in the mouse genome. BMC Genomics. 2011;12:531. https://doi.org/10.1186/1471-2164-12-531
7. Podgornaya OI, Ostromyshenskii DI, Enukashvily NI. Who needs this junk, or genomic dark matter. Biochem Mosc. 2018;83:450–66. https://doi.org/10.1134/S0006297918040156
8. López-Flores I, Garrido-Ramos MA. The repetitive DNA content of eukaryotic genomes. Repetitive DNA. 2012;7:1–28. https://doi.org/10.1159/000337118
9. Rico-Porras JM, Mora P, Palomeque T, Montiel EE, Cabral-de-Mello DC, Lorite P. Heterochromatin is not the only place for SatDNAs: the high diversity of SatDNAs in the euchromatin of the beetle Chrysolina Americana (Coleoptera, Chrysomelidae). Genes. 2024;15:395. https://doi.org/10.3390/genes15040395
10. Macas J, Mészáros T, Nouzová M. PlantSat: a specialized database for plant satellite repeats. Bioinformatics. 2002;18:28–35. https://doi.org/10.1093/bioinformatics/18.1.28
11. Ferree PM, Barbash DA. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. PLoS Biol. 2009;7:e1000234. https://doi.org/10.1371/journal.pbio.1000234
12. Ivanova NG, Ostromyshenskii D, Podgornaya O. Tandem repeat-based probes support the loop model of pericentromere packing. Cytogenet Genome Res. 2021;161:93–102. https://doi.org/10.1159/000513228
13. Nakagawa T, Okita AK. Transcriptional silencing of centromere repeats by heterochromatin safeguards chromosome integrity. Curr Genet. 2019;65:1089–98. https://doi.org/10.1007/s00294-019-00975-x
14. Podgornaya OI. Nuclear organization by satellite DNA, SAF-A/hnRNPU and matrix attachment regions. Semin Cell Dev Biol. 2022;128:61–8. https://doi.org/10.1016/j.semcdb.2022.04.018
15. Rudd MK, Willard HF. Analysis of the centromeric regions of the human genome assembly. Trends Genet. 2004;20:529–33. https://doi.org/10.1016/j.tig.2004.08.008
16. Saifitdinova AF, Derjusheva SE, Malykh AG, Zhurov VG, Andreeva TF, Gaginskaya ER. Centromeric tandem repeat from the chaffinch genome: isolation and molecular characterization. Genome. 2001;44:96–103. https://doi.org/10.1139/gen-44-1-96
17. Enukashvily NI, Dobrynin MA, Chubar AV. RNA-seeded membraneless bodies: role of tandemly repeated RNA. Adv Protein Chem Struct Biol. 2021;126:151–93. https://doi.org/10.1016/bs.apcsb.2020.12.007
18. Trofimova I, Krasikova A. Transcription of highly repetitive tandemly organized DNA in amphibians and birds: a historical overview and modern concepts. RNA Biol. 2016;13:1246–57. https://doi.org/10.1080/15476286.2016.1240142
19. Probst AV, Okamoto I, Casanova M, El Marjou F, Le Baccon P, Almouzni G. A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. Dev Cell. 2010;19:625–38. https://doi.org/10.1016/j.devcel.2010.09.002
20. Gregory TR. Genome size evolution in animals. Evol. Genome, Elsevier;; 2005. pp. 3–87. https://doi.org/10.1016/B978-012301463-4/50003-6
21. Gregory TR. Animal genome size database. Anim Genome Size Database. 2024. http://www.genomesize.com (accessed May 20, 2024).
22. Da Silva MJ, Fogarin Destro R, Gazoni T, Narimatsu H, Pereira Dos Santos PS, Haddad CFB, et al. Great abundance of satellite DNA in *Proceratophrys* (Anura, Odontophrynidae) revealed by genome sequencing. Cytogenet Genome Res. 2020;160:141–7. https://doi.org/10.1159/000506531
23. Cardone DE, Feliciello I, Marotta M, Rosati C, Chinali G. A family of centromeric satellite DNAs from the European brown frog Rana graeca italica. Genome. 1997;40:774–81. https://doi.org/10.1139/g97-800
24. Feliciello I, Picariello O, Chinali G. The first characterisation of the overall variability of repetitive units in a species reveals unexpected features of satellite DNA. Gene. 2005;349:153–64. https://doi.org/10.1016/j.gene.2004.12.001
25. Kwon T, AmphiBase. A new genomic resource for non-model amphibian species. Genesis. 2017;55:e23010. https://doi.org/10.1002/dvg.23010
26. Sun Y-B, Zhang Y, Wang K. Perspectives on studying molecular adaptations of amphibians in the genomic era. Zool Res. 2020;41:351–64. https://doi.org/10.24272/j.issn.2095-8137.2020.046
27. Kosch TA, Torres-Sánchez M, Liedtke HC, Summers K, Yun MH, Crawford AJ, et al. The amphibian genomics consortium: advancing genomic and genetic resources for amphibian research and conservation. BMC Genomics. 2024;25:1025. https://doi.org/10.1186/s12864-024-10899-7
28. Kosch TA, Crawford AJ, Lockridge Mueller R, Wollenberg Valero KC, Power ML, Rodríguez A, et al. Comparative analysis of amphibian genomes: an emerging resource for basic and applied research. Mol Ecol Resour. 2025;25:e14025. https://doi.org/10.1111/1755-0998.14025
29. Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, et al. The genome of the Western clawed frog *Xenopus tropicalis*. Science. 2010;328:633–6. https://doi.org/10.1126/science.1183670
30. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. Genome evolution in the allotetraploid frog Xenopus laevis. Nature. 2016;538:336–43. https://doi.org/10.1038/nature19840
31. Frost DR. Amphibian species of the world: an online reference. Version 6.2. 2024. https://amphibiansoftheworld.amnh.org/index.php. (accessed May 10, 2024).
32. Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju SD, Guillaume K, et al. Global patterns of diversification in the history of modern amphibians. Proc Natl Acad Sci. 2007;104:887–92. https://doi.org/10.1073/pnas.0608378104
33. Streicher JW, Wellcome Sanger Institute Tree of Life programme. The genome sequence of the common frog, Rana temporaria Linnaeus 1758. Wellcome Open Res. 2021;6:286. https://doi.org/10.12688/wellcomeopenres.17296.1
34. Bogolyubov DS, Shabelnikov SV, Travina AO, Sulatsky MI, Bogolyubova IO. Special nuclear structures in the germinal vesicle of the common frog with emphasis on the so-called karyosphere capsule. J Dev Biol. 2023;11:44. https://doi.org/10.3390/jdb11040044
35. Gruzova MN, Parfenov VN. Ultrastructure of late oocyte nuclei in *Rana temporaria*. J Cell Sci. 1977;28:1–13. https://doi.org/10.1242/jcs.28.1.1
36. Gurdon JB, Hopwood N. The introduction of Xenopus laevis into developmental biology: of empire, pregnancy testing and ribosomal genes. Int J Dev Biol. 2000;44:43–50.
37. Ilicheva N, Podgornaya O, Bogolyubov D, Pochukalina G. The karyosphere capsule in Rana temporaria oocytes contains structural and DNA-binding proteins. Nucleus. 2018;9:516–29. https://doi.org/10.1080/19491034.2018.1530935
38. Ilicheva NV, Pochukalina GN, Podgornaya OI. Actin depolymerization disrupts karyosphere capsule integrity but not residual transcription in late oocytes of the grass frog Rana temporaria. J Cell Biochem. 2019;120:15057–68. https://doi.org/10.1002/jcb.28767
39. Scheer U, Dabauvalle M-C. In: Browder LW, Oogenesis, editors. Functional organization of the amphibian oocyte nucleus. Boston, MA: Springer US; 1985. pp. 385–430. https://doi.org/10.1007/978-1-4615-6814-8_9
40. Zlotina A, Dedukh D, Krasikova A. Amphibian and avian karyotype evolution: insights from Lampbrush chromosome studies. Genes. 2017;8:311. https://doi.org/10.3390/genes8110311
41. Manning MJ, Collie MH. Thymic function in amphibians. In: Hildemann WH, Benedict AA, editors. Immunol. Phylogeny. Volume 64. Boston, MA: Springer US; 1975. pp. 353–62. https://doi.org/10.1007/978-1-4684-3261-9_35

42. Tata JR. Chapter 3 Protein Synthesis During Amphibian Metamorphosis. Curr. Top. Dev. Biol., vol. 6, Elsevier; 1971, pp. 79–110. https://doi.org/10.1016/S0070-2153(08)60638-9
43. Dabagyan NV, Sleptsova LA. The common frog Rana temporaria. In: Dettlaff TA, Vassetzky SG, editors. Anim. Species Dev. Stud. Boston, MA: Springer US; 1991. pp. 283–305. https://doi.org/10.1007/978-1-4615-3654-3_10
44. Beattie RC, Aston RJ, Milner AGP. A field study of fertilization and embryonic development in the common frog (Rana temporaria) with particular reference to acidity and temperature. J Appl Ecol. 1991;28:346. https://doi.org/10.2307/2404134
45. Novák P, Ávila Robledillo L, Koblížková A, Vrbová I, Neumann P, Macas J. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res. 2017;45:e111–111. https://doi.org/10.1093/nar/gkx257
46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170
47. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA. 2021;12:2. https://doi.org/10.1186/s13100-020-00230-y
48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2
49. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115–115. https://doi.org/10.1093/nar/gks596
50. Schmid M. Chromosome banding in amphibia: I. Constitutive heterochromatin and nucleolus organizer regions in Bufo and Hyla. Chromosoma. 1978;66:361–88. https://doi.org/10.1007/BF00328536
51. Tagarro I, Wiegant J, Raap AK, Fernandez-Peralta G-AJJ. Assignment of human satellite 1 DNA as revealed by fluorescent in situ hybridization with oligonucleotides. Hum Genet. 1994;93. https://doi.org/10.1007/BF00210595
52. Guillemin C. Caryotypes de Rana temporaria (L.) et de Rana dalmatina (Bonaparte). Chromosoma. 1967;21:189–97. https://doi.org/10.1007/BF00343644
53. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80. https://doi.org/10.1093/nar/27.2.573
54. Plotly Technologies Inc. Collaborative Data Science. Montreal, QC: Plotly Technologies Inc. 2015. Available online: https://plot.ly
55. Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G. Inverted repeat structure of the human genome: the X-Chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res. 2004;14:1861–9. https://doi.org/10.1101/gr.2542904
56. Vassetzky NS, Kramerov DA. SINEBase: a database and tool for SINE analysis. Nucleic Acids Res. 2013;41:D83–9. https://doi.org/10.1093/nar/gks1263
57. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and censor. BMC Bioinformatics. 2006;7:474. https://doi.org/10.1186/1471-2105-7-474
58. Spasić-Bošković O, Tanić N, Blagojević J, Vujošević M. Comparative cytogenetic analysis of European brown frogs: Rana temporaria, R. dalmatina and R. graeca. Caryologia. 1997;50:139–49. https://doi.org/10.1080/00087114.1997.10797393
59. Feliciello I, Picariello O, Chinali G. Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from Rana dalmatina: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. Gene. 2006;383:81–92. https://doi.org/10.1016/j.gene.2006.07.016
60. Picariello O, Feliciello I, Bellinello R, Chinali G. S1 satellite DNA as a taxonomic marker in brown frogs: molecular evidence that Rana graeca graeca and Rana graeca italica are different species. Genome. 2002;45:63–70. https://doi.org/10.1139/g01-125
61. Ostromyshenskii DI, Chernyaeva EN, Kuznetsova IS, Podgornaya OI. Mouse chromocenters DNA content: sequencing and in silico analysis. BMC Genomics. 2018;19:151. https://doi.org/10.1186/s12864-018-4534-z
62. Peona V, Weissensteiner MH, Suh A. How complete are complete genome assemblies?—An avian perspective. Mol Ecol Resour. 2018;18:1188–95. https://doi.org/10.1111/1755-0998.12933
63. Vitelli L, Batistoni R, Andronico F, Nardi I, Barsacchi-Pilone G. Chromosomal localization of 18S + 28S and 5S ribosomal RNA genes in evolutionarily diverse Anuran amphibians. Chromosoma. 1982;84:475–91. https://doi.org/10.1007/BF00292849
64. Miga KH. Completing the human genome: the progress and challenge of satellite DNA assembly. Chromosome Res. 2015;23:421–6. https://doi.org/10.1007/s10577-015-9488-2
65. Thakur J, Packiaraj J, Henikoff S, Sequence. Chromatin and evolution of satellite DNA. Int J Mol Sci. 2021;22:4309. https://doi.org/10.3390/ijms22094309
66. Tørresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarnot P, et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. Nucleic Acids Res. 2019;47:10994–1006. https://doi.org/10.1093/nar/gkz841
67. Lin Y, Ye C, Li X, Chen Q, Wu Y, Zhang F, et al. QuarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic Res. 2023;10:uhad127. https://doi.org/10.1093/hr/uhad127
68. Navrátilová P, Toegelová H, Tulpová Z, Kuo Y, Stein N, Doležel J, et al. Prospects of telomere-to-telomere assembly in barley: analysis of sequence gaps in the MorexV3 reference genome. Plant Biotechnol J. 2022;20:1373–86. https://doi.org/10.1111/pbi.13816
69. Li H, Durbin R. Genome assembly in the telomere-to-telomere era. Nat Rev Genet. 2024;25:658–70. https://doi.org/10.1038/s41576-024-00718-w
70. Mc Cartney AM, Shafin K, Alonge M, Bzikadze AV, Formenti G, Fungtammasan A, et al. Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies. Nat Methods. 2022;19:687–95. https://doi.org/10.1038/s41592-022-01440-3
71. Miga KH. The promises and challenges of genomic studies of human centromeres. In: Black BE, editor. Centromeres kinetochores. Volume 56. Cham: Springer International Publishing; 2017. pp. 285–304. https://doi.org/10.1007/978-3-319-58592-5_12
72. Sullivan LL, Sullivan BA. Genomic and functional variation of human centromeres. Exp Cell Res. 2020;389:111896. https://doi.org/10.1016/j.yexcr.2020.111896
73. Chen W, Chen H, Liao J, Tang M, Qin H, Zhao Z, et al. Chromosome-level genome assembly of a high-altitude-adapted frog (Rana kukunoris) from the Tibetan plateau provides insight into amphibian genome evolution and adaptation. Front Zool. 2023;20:1. https://doi.org/10.1186/s12983-022-00482-9
74. Lisachov A, Rumyantsev A, Prokopov D, Ferguson-Smith M, Trifonov V. Conservation of major satellite DNAs in snake heterochromatin. Animals. 2023;13:334. https://doi.org/10.3390/ani13030334
75. Da Silva M, Gazoni T, Haddad CFB, Parise-Maltempi PP. Analysis in Proceratophrys boiei genome illuminates the satellite DNA content in a frog from the Brazilian Atlantic forest. Front Genet. 2023;14:1101397. https://doi.org/10.3389/fgene.2023.1101397
76. Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with rapidly evolving DNA. Science. 2001;293:1098–102. https://doi.org/10.1126/science.1062939
77. Zhang H, Guo Q, Iliopoulos CS. Locating tandem repeats in weighted sequences in proteins. BMC Bioinformatics. 2013;14 Suppl 8:S2. https://doi.org/10.1186/1471-2105-14-S8-S2
78. Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby J, et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. 2013;14:R10. https://doi.org/10.1186/gb-2013-14-1-r10
79. Marracci S, Michelotti V, Guex G-D, Hotz H, Uzzell T, Ragghianti M. RrS1-like sequences of water frogs from central Europe and around the Aegean Sea: chromosomal organization, evolution, possible function. J Mol Evol. 2011;72:368–82. https://doi.org/10.1007/s00239-011-9436-5
80. Ragghianti M, Guerrini F, Bucci S, Mancino G, Hotz H, Uzzell T, et al. Molecular characterization of a centromeric satellite DNA in the hemiclonal hybrid Fro-gRana esculenta and its parental species. Chromosome Res. 1995;3:497–506. https://doi.org/10.1007/BF00713965
81. Edwards NS, Murray AW. Identification of *Xenopus* CENP-A and an associated centromeric DNA repeat. Mol Biol Cell. 2005;16:1800–10. https://doi.org/10.1091/mbc.e04-09-0788
82. Shang W-H, Hori T, Toyoda A, Kato J, Popendorf K, Sakakibara Y, et al. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Res. 2010;20:1219–28. https://doi.org/10.1101/gr.106245.110
83. Smith OK, Limouse C, Fryer KA, Teran NA, Sundararajan K, Heald R, et al. Identification and characterization of centromeric sequences in *Xenopus laevis*. Genome Res. 2021;31:958–67. https://doi.org/10.1101/gr.267781.120
84. Sun X, Wahlstrom J, Karpen G. Molecular structure of a functional Drosophila centromere. Cell. 1997;91:1007–19. https://doi.org/10.1016/s0092-8674(00)80491-2

85. Bredeson JV, Mudd AB, Medina-Ruiz S, Mitros T, Smith OK, Miller KE, et al. Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. Nat Commun. 2024;15:579. https://doi.org/10.1038/s41467-023-43012-9
86. Hassan AH, Mokhtar MM, El Allali A. Transposable elements: multifunctional players in the plant genome. Front Plant Sci. 2024;14:1330127. https://doi.org/10.3389/fpls.2023.1330127
87. Kretschmer R, Toma GA, Deon GA, Dos Santos N, Dos Santos RZ, Utsunomia R, et al. Satellitome analysis in the Southern lapwing (Vanellus chilensis) genome: implications for SatDNA evolution in charadriiform birds. Genes. 2024;15:258. https://doi.org/10.3390/genes15020258
88. Peona V, Kutschera VE, Blom MPK, Irestedt M, Suh A. Satellite DNA evolution in corvoidea inferred from short and long reads. Mol Ecol. 2023;32:1288–305. https://doi.org/10.1111/mec.16484
89. Pons J, Petitpierre E, Juan C. Characterization of the heterochromatin of the darkling beetle misolampus goudoti: cloning of two satellite DNA families and digestion of chromosomes with restriction enzymes. Hereditas. 2004;119:179–85. https://doi.org/10.1111/j.1601-5223.1993.00179.x
90. Sader M, Vaio M, Cauz-Santos LA, Dornelas MC, Vieira MLC, Melo N, et al. Large vs small genomes in passiflora: the influence of the mobilome and the satellitome. Planta. 2021;253:86. https://doi.org/10.1007/s00425-021-03598-0
91. Utsunomia R, Silva DMZDA, Ruiz-Ruano FJ, Goes CAG, Melo S, Ramos LP, et al. Satellitome landscape analysis of megaleporinus macrocephalus (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. Sci Rep. 2019;9:5856. https://doi.org/10.1038/s41598-019-42383-8
92. Guzmán K, Roco ÁS, Stöck M, Ruiz-García A, García-Muñoz E, Bullejos M. Identification and characterization of a new family of long satellite DNA, specific of true toads (Anura, amphibia, Bufonidae). Sci Rep. 2022;12:13960. https://doi.org/10.1038/s41598-022-18051-9
93. Kramerov DA, Vassetzky NS. Origin and evolution of sines in eukaryotic genomes. Heredity. 2011;107:487–95. https://doi.org/10.1038/hdy.2011.43
94. Kojima KK. A new class of sines with SnRNA gene-derived heads. Genome Biol Evol. 2015;7:1702–12. https://doi.org/10.1093/gbe/evv100
95. Ackerman EJ. Molecular cloning and sequencing of OAX DNA: an abundant gene family transcribed and activated in Xenopus oocytes. EMBO J. 1983;2:1417–22. https://doi.org/10.1002/j.1460-2075.1983.tb01600.x
96. Bucci S, Ragghianti M, Mancino G, Petroni G, Guerrini F, Giampaoli S, Rana/Pol III. A family of SINE-like sequences in the genomes of Western Palearctic water frogs. Genome. 1999;42:504–11. https://doi.org/10.1139/g98-149
97. Nagahashi S, Endoh H, Suzuki Y, Okada N. Characterization of a tandemly repeated DNA sequence family originally derived by retroposition of tRNAGlu in the Newt. J Mol Biol. 1991;222:391–404. https://doi.org/10.1016/0022-2836(91)90218-U
98. Zuo B, Nneji LM, Sun Y-B. Comparative genomics reveals insights into Anuran genome size evolution. BMC Genomics. 2023;24:379. https://doi.org/10.1186/s12864-023-09499-8
99. Vassetzky NS, Kosushkin SA, Ryskov AP. SINE-derived satellites in scaled reptiles. Mob DNA. 2023;14:21. https://doi.org/10.1186/s13100-023-00309-2
100. Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, et al. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nat Commun. 2018;9:2774. https://doi.org/10.1038/s41467-018-05279-1
101. Mao H, Wang H. SINE_scan: an efficient tool to discover short interspersed nuclear elements (SINEs) in large-scale genomic datasets. Bioinformatics. 2017;33:743–5. https://doi.org/10.1093/bioinformatics/btw718
102. Suvorova YM, Kamionskaya AM, Korotkov EV. Search for SINE repeats in the rice genome using correlation-based position weight matrices. BMC Bioinformatics. 2021;22:42. https://doi.org/10.1186/s12859-021-03977-0
103. Jo S-H, Koo D-H, Kim JF, Hur C-G, Lee S, Yang T, et al. Evolution of ribosomal DNA-derived satellite repeat in tomato genome. BMC Plant Biol. 2009;9:42. https://doi.org/10.1186/1471-2229-9-42
104. Macas J, Navratilova A, Meszaros T. Sequence subfamilies of satellite repeats related to rDNA intergenic spacer are differentially amplified on Vicia sativa chromosomes. Chromosoma. 2003;112:152–8. https://doi.org/10.1007/s00412-003-0255-3
105. De Lucchini S, Andronico F, Nardi I. Molecular structure of the rDNA intergenic spacer (IGS) in Triturus: implications for the hypervariability of rDNA loci. Chromosoma. 1997;106:315–26. https://doi.org/10.1007/s004120050253
106. Martins C, Ferreira IA, Oliveira C, Foresti F, Galetti PM. A tandemly repetitive centromeric DNA sequence of the fish Hoplias malabaricus (Characiformes: Erythrinidae) is derived from 5S rDNA. Genetica. 2006;127:133. https://doi.org/10.1007/s10709-005-2674-y
107. Vittorazzi SE, Lourenço LB, Del-Grande ML, Recco-Pimentel SM. Satellite DNA derived from 5S rDNA in *Physalaemus cuvieri*. (Anura, Leiuperidae) Cytogenet Genome Res. 2011;134:101–7. https://doi.org/10.1159/000325540
108. Gatto KP, Busin CS, Lourenço LB. Unraveling the sex chromosome heteromorphism of the paradoxical frog pseudis tocantins. PLoS ONE. 2016;11:e0156176. https://doi.org/10.1371/journal.pone.0156176
109. Gatto KP, Mattos JV, Seger KR, Lourenço LB. Sex chromosome differentiation in the frog genus pseudis involves satellite DNA and chromosome rearrangements. Front Genet. 2018;9:301. https://doi.org/10.3389/fgene.2018.00301
110. Garrido-Ramos M. Satellite DNA: an evolving topic. Genes. 2017;8:230. https://doi.org/10.3390/genes8090230
111. Kuhn GCS, Küttler H, Moreira-Filho O, Heslop-Harrison JS. The 1.688 repetitive DNA of drosophila: concerted evolution at different genomic scales and association with genes. Mol Biol Evol. 2012;29:7–11. https://doi.org/10.1093/molbev/msr173
112. Pavlek M, Gelfand Y, Plohl M, Meštrović N. Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. DNA Res. 2015;22:387–401. https://doi.org/10.1093/dnares/dsv021
113. Feliciello I, Akrap I, Ugarković Đ. Satellite DNA modulates gene expression in the beetle Tribolium castaneum after heat stress. PLOS Genet. 2015;11:e1005466. https://doi.org/10.1371/journal.pgen.1005466
114. Bureau TE, Wessler SR. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell. 1992;4:1283–94. https://doi.org/10.1105/tpc.4.10.1283
115. Venkatesh NB. Miniature inverted-repeat transposable elements (MITEs), derived insertional polymorphism as a tool of marker systems for molecular plant breeding. Mol Biol Rep. 2020;47:3155–67. https://doi.org/10.1007/s11033-020-05365-y
116. Scalvenzi T, Pollet N. Insights on genome size evolution from a miniature inverted repeat transposon driving a satellite DNA. Mol Phylogenet Evol. 2014;81:1–9. https://doi.org/10.1016/j.ympev.2014.08.014
117. Plohl M, Meštrović N, Mravinac B. Satellite DNA Evolution. In: Garrido-Ramos MA,Genome Dyn., vol. 7, Karger S. AG; 2012, pp. 126–52. https://doi.org/10.1159/000337122

Publisher's note