**Journal of Genetics** (2020)99:58
https://doi.org/10.1007/s12041-020-01222-w



**RESEARCH ARTICLE**

# Genomewide analysis of microsatellite markers based on sequenced database in two anuran species

LI WEI, WEIWEI SHAO, LI MA and ZHIHUA LIN*

*College of Ecology, Lishui University, Lishui 323000, Zhejiang, People's Republic of China*
*For correspondence. E-mail: zhlin1015@126.com.

Received 16 September 2019; revised 31 March 2020; accepted 27 April 2020

**Abstract.** Eukaryotic and prokaryotic cell genomes exhibit multiple microsatellites. In this study, we characterized microsatellites in genomes and genes of *Nanorana parkeri* and *Xenopus laevis.* This characterization was used for gene ontology (GO) analysis of coding sequences (CDS). Compared to the genome of *N. parkeri*, the genome of *X. laevis* is larger and contains more number of microsatellites, but the diversity of both species are similar. Trinucleotide repeats in the genome of *N. parkeri* and dinucleotide and tetranucleotide repeats in the genome of *X. laevis* were the most diverse. In both the species, diversity of microsatellites was highest in intergenic regions, followed by intron and exon regions, and lowest in coding regions. Microsatellites in CDS are thus subject to higher selective pressure. Many microsatellites are concentrated upstream and downstream of genes in both species, suggesting suppression of repeats in the middle of protein–CDS. Repeats are enriched in regions near gene termini purely due to the biophysical constraints of protein structure. In GO analysis, two and five unique GO terms, only found in *N. parkeri* and *X. laevis*, respectively, indicate advantageous mutations during species evolution. Biological process, cellular component and molecular function ontology reflected in the GO analysis predicted that the microsatellites located in CDS can alter protein function and may provide a molecular basis for species adaptation to new and changing environments.

**Keywords.** Anura; simple-sequence repeat; genomic microsatellite; diversity distribution; gene ontology analysis.

## Introduction

Microsatellites, also known as simple-sequence repeats (SSRs), are tandemly repeated sequences of DNA that are composed of one to six base pairs (bp) (Tóth *et al*. 2000; Hong *et al*. 2007; Ding *et al*. 2017). SSRs are found throughout the prokaryotic and eukaryotic genomes (Hancock 1996; Tóth *et al*. 2000) and in both noncoding and coding regions of DNA (Mayer *et al*. 2010). Moreover, SSRs may also be associated with retrotransposons (Tay *et al*. 2010). SSR instability is primarily due to the slipped-strand mispairing errors that occur during DNA replication (Huntley and Golding 2006). Strand slippage and unequal recombination leads to the insertions/deletions of one to several repeated units, and such high instability makes them attractive polymorphic molecular markers (Deback *et al*. 2009). High diversity, high polymorphism, and neutral and codominant inheritance make microsatellite valuable molecular markers for fingerprinting, parentage analysis, genetic mapping, and genome structure analysis (Xu *et al*. 2018).

To isolate SSR, traditional methods like magnetic bead enrichment, selective hybridization and EST-SSR are used that are effective, number-limited, time-consuming, labour-intensive and costly (Zane *et al*. 2002). However, next-generation sequencing technology makes SSR development easier and allows larger numbers of microsatellites to be recognized (Rogers and Gibbs 2014). Currently, these laboratory methods are being replaced by *in silico* mining of SSR loci from genome databases (Liu *et al*. 2017). The development of *in silico* mining tools (i.e. SciRoko, MIcroSAtellite (MISA), and MSDB) provide new opportunities for genomewide scans of large numbers of microsatellite (Behura *et al*. 2015; Xu *et al*. 2018; Song *et al*. 2019).

Genomewide microsatellite mining has been carried out in different species, including plants (Han *et al*. 2015), insects (Ding *et al*. 2017), reptiles (Nie *et al*. 2017; Song *et al*. 2019), birds (Cui and Yue 2018), mice (Tu *et al*. 2018a, b), primates (Xu *et al*. 2018), and human (Subramanian *et al*. 2003). These investigations indicate that SSRs are found less

Published online: 30 June 2020

frequently in protein-coding sequences than in intronic and intergenic regions (Nie *et al*. 2017). Overall, microsatellites are more diverse in coding regions than in noncoding regions due to a higher coding density of the former (Alam *et al*. 2013). Microsatellite length expansion may affect gene regulation, transcription, and protein function of coding sequences (CDS), especially for trinucleotide repeats, which are associated with human diseases (Xu *et al*. 2016). Microsatellite distribution characteristics and function may be different in unique genomes (Wang *et al*. 2016). Unlike most other species, trinucleotide SSRs are the most diverse types in the CDS region of *Blattella germanica* (Wang *et al*. 2015), *Boa constrictor* and *Protobothrops mucrosquamatus* (Nie *et al*. 2017), and *Centruroides exilicauda* and *Mesobuthus martensii* (Wang *et al*. 2016).

In this work, we investigated the anuran genomes of *Nanorana parkeri* and *Xenopus laevis*, which were reported to the open databases. *N. parkeri* is a species of Neobatrachia, while *X. laevis* is a representative species of Archaeobatrachia. This study aims (i) to identify similarities and difference in numbers and diversity of microsatellites at the genomic level; (ii) to compare the characteristics of microsatellites in different regions (intergenic, exon and intron); (iii) to compare the position of microsatellite diversity among gene regions; and (iv) to investigate functions of coding sequences containing microsatellites and analyse differences in coding sequences of genes containing microsatellites. This study contributes to the understanding of the anuran amphibian genome and facilitates subsequent screening and development of large numbers of high-quality microsatellite markers.

## Materials and methods

The *N. parkeri* genome assembly was downloaded directly from the National Center for Biotechnology Information (NCBI) under BioProject accession PRJNA243398, with annotation files downloaded from https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/935/625/GCF_000935625.1_ASM93562v1, including CDS sequences. The genome assembly of *X. laevis* was downloaded from NCBI under BioProject accession PRJNA313213, with annotation files downloaded from https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/663/975/GCF_001663975.1_Xenopus_laevis_v2/, including CDS sequences. Microsatellites in genome and CDS were identified using MISA identification tool software, which has been used for microsatellite analysis of many species, including *Hordeum vulgare* (Thiel *et al*. 2003), *Epinephelus awoara* (Gao *et al*. 2017), *B. constrictor* and *P. mucrosquamatus* (Nie *et al*. 2017). Def in the misa.ini file was set as 1–12, 2–7, 3–5, 4–4, 5–4, and 6–4, to restrict the detection criteria for perfect SSR of 1–6 bp with minimum repeat numbers of 12, 7, 5, 4, 4 and 4 for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide microsatellites, respectively (Demuth and Drury 2007; Nie *et al*. 2017; Song *et al*. 2019; Wang *et al*. 2016). Once the distance between two microsatellites is shorter than 100 bp, they were considered as a single-compound microsatellite. Repeat motifs consisting of different frames (e.g. AGC, GCA, CAG) were regarded as the same type of repeat (Fujimori *et al*. 2003; Wang *et al*. 2016; Nie *et al*. 2017).

To investigate the distribution of microsatellites, the relative position of exon, intron, gene and intergene regions were extracted from the annotation files via custom Python scripts. Also, the microsatellites that are located within 500 bp of gene termini were defined as upstream or downstream of genes (Nie *et al*. 2017). To calculate the microsatellite diversity, the microsatellites located in different regions of genes were divided into 13 elements: 500 bp upstream, first exon/intron, second exon/intron, middle left exon, middle intron, middle right exon, last second intron, last second exon, last intron, last exon and 500 bp downstream (Nie *et al*. 2017). To avoid the overlaps in measurements, only the genes consisting of more than six exons and five introns were considered (Fujimori *et al*. 2003). The overlap here means, e.g., an SSR would be calculated twice if it was located in the second exon of a gene containing three exons and also in the second last exon of this gene (Wang *et al*. 2016). The relative position of a microsatellite in a certain type of element is the distance from the microsatellite to the left end of the element divided by the distance between the length of the element and the length of the microsatellite. The diversity of microsatellites is calculated using the methods described by Fujimori *et al*. (2003) and Nie *et al*. (2017).

CDS with SSR were aligned against NCBI nonredundant and SWISS-PROT protein databases (http://www.uniprot.org), using BLASTx with a cut-off E-value of $1E^{-5}$ (Nie *et al*. 2017). Blast results were imported into Blast2Go (Conesa *et al*. 2005) for gene ontology (GO) classification.

## Results

### *Characterization of SSRs in genomes of N. parkeri and X. laevis*

A total of 183,113 SSRs were identified in the genome assembly of about 205.4 Mb for *N. parkeri* and a total of 207,990 SSRs were identified in the genome assembly of about 271.8 Mb for *X. laevis* (table 1). Most SSRs in both species were dinucleotide types, 72,799 in *N. parkeri* and 69,688 in *X. laevis*, respectively. The total microsatellite diversity between species was similar: 89.16 SSRs/Mbp in *N. parkeri* and 76.51 SSRs/Mbp in *X. laevis* (table 1). For *N. parkeri*, trinucleotide repeats were the most diversified category followed by mononucleotide and tetranucleotide repeats. In *X. laevis*, dinucleotide and tetranucleotide repeats were the most diversified category followed by mononucleotide and trinucleotide repeats (table 1).

**Table 1.** Distribution of microsatellites in the genomes of *N. parkeri* and *X. laevis.*

| | *N. parkeri* | | | | *X. laevis* | | | |
|---|---|---|---|---|---|---|---|---|
| Motif length | Number of microsatellites | Length (bp) | Abundance (SSRs/Mbp) | Frequency (%) | Number of microsatellites | Length (bp) | Abundance (SSRs /Mbp) | Frequency (%) |
| Mononucleotide | 46,478 | 866,012 | 22.63 | 25.38 | 45,804 | 880,444 | 16.85 | 22.02 |
| Dinucleotide | 72,799 | 1,697,396 | 35.44 | 39.76 | 69,688 | 1,698,166 | 25.64 | 33.51 |
| Trinucleotide | 18,209 | 401,148 | 8.87 | 9.94 | 20,796 | 478,479 | 7.65 | 10.00 |
| Tetranucleotide | 42,581 | 1,413,556 | 20.73 | 23.25 | 69,111 | 2,205,908 | 25.42 | 33.23 |
| Pentanucleotide | 2395 | 52,295 | 1.17 | 1.31 | 1587 | 33,525 | 0.58 | 0.76 |
| Hexanucleotide | 651 | 19,662 | 0.32 | 0.36 | 1004 | 30,876 | 0.37 | 0.48 |
| Total | 183,113 | 4,450,069 | 89.16 | 100 | 207,990 | 5,327,398 | 76.51 | 100.00 |
| Whole genome length/bp | 2,053,867,363 | | | | 2,718,433,805 | | | |
| SSR content of genome | 0.22% | | | | 0.20% | | | |

**Table 2.** The most frequent microsatellite motifs found in genomes of *N. parkeri* and *X. laevis.*

| | *N. parkeri* | | | *X. laevis* | | |
|---|---|---|---|---|---|---|
| Motif length | Repeat unit | Microsatellites | Frequency (%) | Repeat unit | Microsatellites | Frequency (%) |
| Mononucleotide | T | 29,821 | 64.16 | A | 44,067 | 96.21 |
| | C | 16,657 | 35.84 | G | 1737 | 3.79 |
| Dinucleotide | TA | 28,420 | 39.04 | TA | 32,282 | 46.32 |
| | GT | 24,986 | 34.32 | GT | 18,763 | 26.92 |
| | GA | 19,378 | 26.62 | GA | 18,638 | 26.74 |
| | GC | 15 | 0.02 | GC | 5 | 0.02 |
| Trinucleotide | ATT | 13,417 | 73.68 | ATA | 13,508 | 64.95 |
| | TCA | 1084 | 5.95 | TCA | 3036 | 14.60 |
| | ACT | 902 | 4.95 | TAG | 1497 | 7.20 |
| | CAC | 664 | 3.64 | GCA | 862 | 4.15 |
| Tetranucleotide | CTAT | 19,327 | 45.39 | GATA | 39,873 | 57.69 |
| | TGTA | 5610 | 13.17 | ATAA | 6600 | 9.55 |
| | TTTA | 5388 | 12.65 | TACA | 5547 | 8.02 |
| | TCTT | 3140 | 7.37 | ACAG | 4717 | 6.83 |
| Pentanucleotide | GACCA | 279 | 11.65 | TATTA | 188 | 11.85 |
| | AATAA | 170 | 7.10 | ATTTT | 148 | 9.33 |
| | TTTGT | 168 | 7.01 | AGTAT | 96 | 6.05 |
| | ATACA | 120 | 5.01 | TATTG | 95 | 6.00 |
| Hexanucleotide | TATATC | 87 | 13.36 | CTATAT | 288 | 28.69 |
| | TAACCC | 62 | 9.52 | TGTATA | 97 | 9.66 |
| | GTATGT | 51 | 7.83 | TGTGAG | 63 | 6.27 |
| | TATGTA | 49 | 7.53 | AGATAG | 57 | 5.68 |

The most diverse SSR types from mononucleotide to hexanucleotide in *N. parkeri* were (T)n, (TA)n, (ATT)n, (CTAT)n, (GACCA)n, and (TATATC)n, and in *X. laevis* were (A)n, (TA)n, (ATA)n, (GATA)n, (TATTA)n, and (CTATAT)n. Similarities between species were noted in dinucleotide (TA)n, (GT)n, (GA)n, and (GC)n. Differences were concentrated in mono-, tri-, tetra-, penta-, and hexanucleotide types (table 2).

The 15 most diverse microsatellite repeats in the *N. parkeri* genome were (T)n, (TA)n, (GT)n, (GA)n, (CTAT)n, (C)n, (ATT)n, (TGTA)n, (TTTA)n, (TCTT)n, (TTCC)n, (TCTG)n, (GATG)n, (TCA)n, and (ACT)n that comprised 94.09% of all microsatellites identified. For *X. laevis*, the 15 most diverse microsatellite motifs were (A)n, (GATA)n, (AT)n, (GT)n, (AG)n, (ATA)n, (ATAA)n, (TACA)n, (ACAG)n, (TTTC)n, (TCA)n, (G)n, (TAG)n,

**Table 3.** The number and diversity (microsatellites/Mbp)) of microsatellites in different genomic regions of *N. parkeri* and *X. laevis.*

| | Gene | | | | |
|---|---|---|---|---|---|
| Species | CDS | Untranslated | Exon | Intron | Intergenic |
| *N. parkeri* | 217 (6.76) | 370 (66.67) | 587 (15.59) | 80,187 (94.50) | 100,303 (87.43) |
| *X. laevis* | 443 (8.23) | 3856 (95.18) | 4299 (45.59) | 90,370 (84.11) | 109,569 (72.37) |

(AAAC)n, and (ATCC)n that comprised 94.39% of all microsatellites identified.

### *Distribution of SSRs in different genomic regions*

Intergenic regions have most numbers of microsatellites, and CDS exhibits a few in both the species (table 3). Numbers of microsatellites in intergenic and intron regions of *X. laevis* were greater than *N. parkeri*, but diversity of microsatellites in intergenic and intron regions of *X. laevis* was less in *N. parkeri*. Numbers and diversity of microsatellites in exons, CDS and untranslated region in *X. laevis* were all larger than in *N. parkeri*. SSRs in CDS were found to be less diverse than in noncoding regions. The frequency of different microsatellite types in different genome regions is shown in figure 1. In both species, trinucleotides were the most diverse microsatellite type in CDS, with 88.02% and 88.94% in *N. parkeri* and *X. laevis*, respectively (figure 1b). The numbers of mono-, di-, tri-, tetra-, penta-, and hexanucleotide in CDS and exons of *X. laevis* were larger than that of *N. parkeri* (figures 1, b&c). The distribution of SSRs in intergenic regions (figure 1d) was similar to the distribution in whole genomes (figure 1a), with the most diversity among mono-, di-, tri-, and tetra-nucleotides.

All SSRs in exons or introns were compared to 12,055 and 16,789 genes with more than six exons and five introns in *N. parkeri* and *X. laevis*, respectively (figure 2). Microsatellite-enriched regions were upstream and downstream of genes in *N. parkeri* genomes with numbers of microsatellites in exons gradually decreasing from the first exon toward the last second exon and increasing toward the last exon. This pattern was also detected in introns (figure 2). However, for *X. laevis*, the greatest SSR diversity was found in the last second exon. SSR diversity in upstream and downstream regions was similar. SSR diversity in exons was bimodal with a gradual decrease from the first exon toward the right middle exon, an increase toward the last second exon, and a decrease again toward the last exon. SSR diversity in various introns was similar (figure 2).

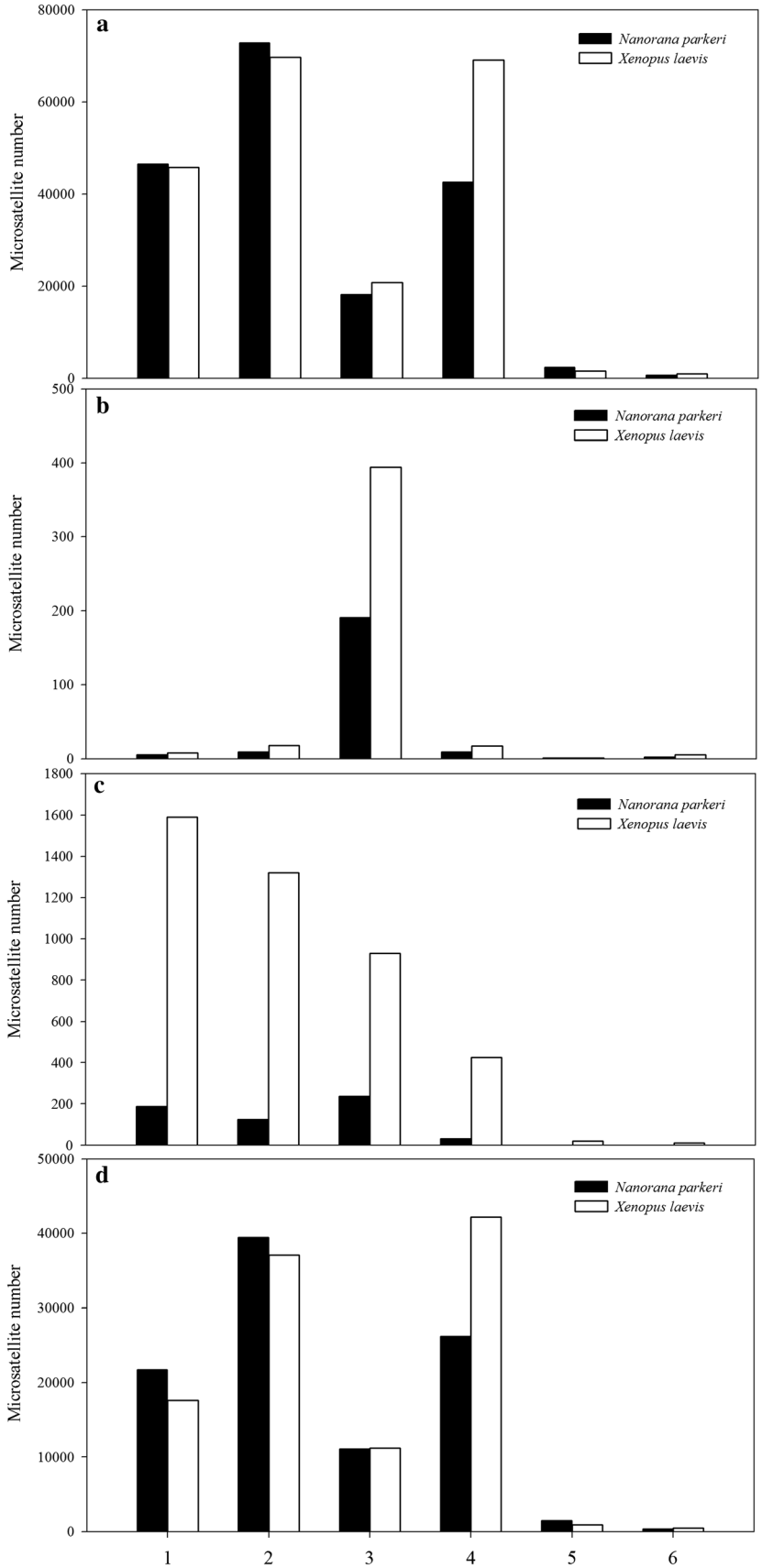
### *GO analysis of CDS with microsatellites for two species*

In genomes of *N. parkeri* and *X. laevis*, 200 and 415 CDS with SSR, respectively, were used in GO analysis based on sequence alignment. All these CDS were assigned to 5082 (*N. parkeri*) and 7973 (*X. laevis*) GO terms of known functions. Numbers of CDS with SSRs assigned to each subcategory are shown in figure 3. Of these GO functional classifications, 54 pairs were represented in both species. Nitrogen utilization (GO:0019740) in biological process ontology and chemorepellent activity (GO:0045499) in molecular function ontology were only present in *N. parkeri*. Nucleoid (GO:0009295) and virion (GO:0019012) in cellular component ontology, antioxidant activity (GO:0016209), translation regulator activity (GO:0045182), and molecular function regulator (GO:0098772) in molecular function ontology were present only in *X. laevis*. Comparing the function distribution between these two species, cellular processes in biological process ontology (GO:0009987) were most frequent. Binding (GO:0005488) was prominent in molecular function ontology. In the cellular component ontology, cell and cell parts (GO:0005623 and GO:0044464) were the top two terms.

## Discussion

In the present study, we characterized the microsatellites in genomes and genes of *N. parkeri* and *X. laevis* and analysed their frequency and distribution in different genomic regions. We found total numbers of SSRs identified in *X. laevis* was larger than in *N. parkeri*. Differences in the number of SSRs may be related to genome size, assembly quality, and specificity of species (Nie *et al*. 2017). These patterns are in agreement with other species, such as scorpions (*C. exilicauda* and *M. martensii*) (Wang *et al*. 2016), snakes (*B. constrictor* and *P. mucrosquamatus*) (Nie *et al*. 2017), and mites (*Tetranychus urticae* and *Ixodes scapularis*) (Wang *et al*. 2013). Microsatellite content in genomes of *N. parkeri* and *X. laevis* is similar, accounting for 0.22% and 0.2%, respectively. This result is consistent with previous studies in mammals, such as *Bos mutus* (5.80%), *Bubalus bubalis* (5.69%) (Qi *et al*. 2015), *Capra hircus* (5.56%), and *Pantholops hodgsonii* (5.39%) (Qi *et al*. 2016), and snakes, such as *B. constrictor* (0.59%) and *P. mucrosquamatus* (0.73%)

**Figure 1.** The distribution of microsatellite types in different genomic regions (a) genome, (b) CDS, (c) exon and (d) intergenetic of *N. parkeri* and *X. laevis*. 1, Mononucleotide; 2, dinucleotide; 3, trinucleotide; 4, tetranucleotide; 5, pentanucleotide; 6, hexanucleotide unit length. ▶


a
b
c
d
Nanorana parkeri
Xenopus laevis
Microsatellite number
80000
60000
40000
20000
0
500
400
300
200
100
1800
1600
1400
1200
1000
800
600
400
50000
30000
10000
1
2
3
4
5
6

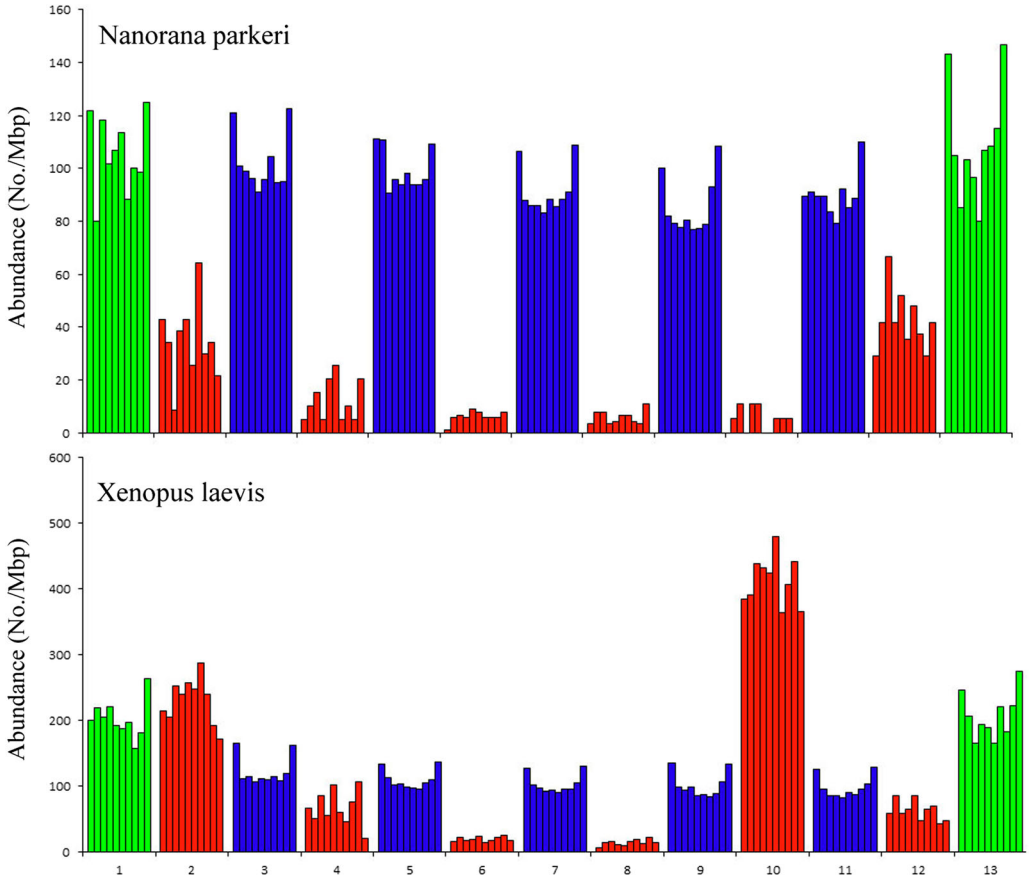



**Figure 2.** The microsatellite diversity in gene regions and their upstream and downstream regions of *N. parkeri* and *X. laevis*. Green represents upstream and downstream, red represents exons and blue represents introns. 1–13 Represents gene upstream 500 bp, the first exon, the first intron, the second exon, the second intron, middle left exon, middle intron, middle right exon, the last second intron, the last second exon, the last intron, the last exon and gene downstream 500 bp, respectively.

(Nie *et al*. 2017). The total SSR diversity in genomes of *N. parkeri* and *X. laevis* are 89.16 SSRs/Mbp and 76.51 SSRs/Mbp, respectively, which were lower compared to the diversity of mosquito (*Anopheles sinensis*) with 365.5 SSRs/Mb; scorpions, such as *C. exilicauda* with 123.08 SSRs/Mb and *M. martensii* with 228.91 SSRs/Mb (Wang *et al*. 2016); snakes, such as *B. constrictor* with 275.46 SSRs/Mb and *P. mucrosquamatus* with 252.33 SSRs/Mb (Nie *et al*. 2017); fish, such as *Scleropages formosus* with 469.65–576.35 SSRs/Mb (Duan *et al*. 2019); bird, such as *Lophophorus lhuysii* with 290.47 SSRs/Mb (Cui and Yue 2018); primates, such as *Macaca fascicularis* with 451.3 SSRs/Mb (Tu *et al*. 2018b); and human with 315.93 SSRs/Mb (Nie *et al*. 2017). Thus, the genomic size and quality of sequencing have a great influence on the identification of microsatellites (Nie *et al*. 2017).

The sequence proportions of six SSR types in *N. parkeri* and *X. laevis* genomes are different, as are the five most diverse microsatellite types (table 1). This result is also found in patterns of genomic SSRs of *B. constrictor* and *P. mucrosquamatus* (Nie *et al*. 2017). Unlike other species, genomes of *Ailuropoda melanoleuca* and *Ursus maritimus* show similarity in the five most diverse SSRs (Li *et al*. 2014). These species are in the same family, suggesting that the differences in microsatellite composition in genome can reflect the relationship among species to some extent (Ding *et al*. 2017). The greatest numbers of SSRs in *N. parkeri* and *X. laevis* genomes are dinucleotide repeat types, which is in agreement with previous study of *Fenneropenaeus chinensis* (Gao *et al*. 2004). However, in genomes of *B. constrictor* and *P. mucrosquamatus* (Nie *et al*. 2017), and *C. exilicauda* and *M. martensii* (Wang *et al*. 2016), the greatest numbers of mononucleotide SSRs were found, and in *Saccharomyces cerevisiae* (Katti *et al*. 2001), trinucleotide were present in greatest numbers.

Microsatellite diversity in untranslated regions was larger than those in CDS regions, indicating that microsatellites aggregate in untranslated regions, presumably affecting gene transcriptional activity (Ellegren 2004). In comparisons with *N. parkeri* and *X. laevis*, differences in both frequency and diversity of SSRs in CDS were small, while those in exon, intron, and intergenic regions differed greatly (table 3). Coding regions are conservative among different species and are subject to high selective pressure (Lin and Kussell 2012).

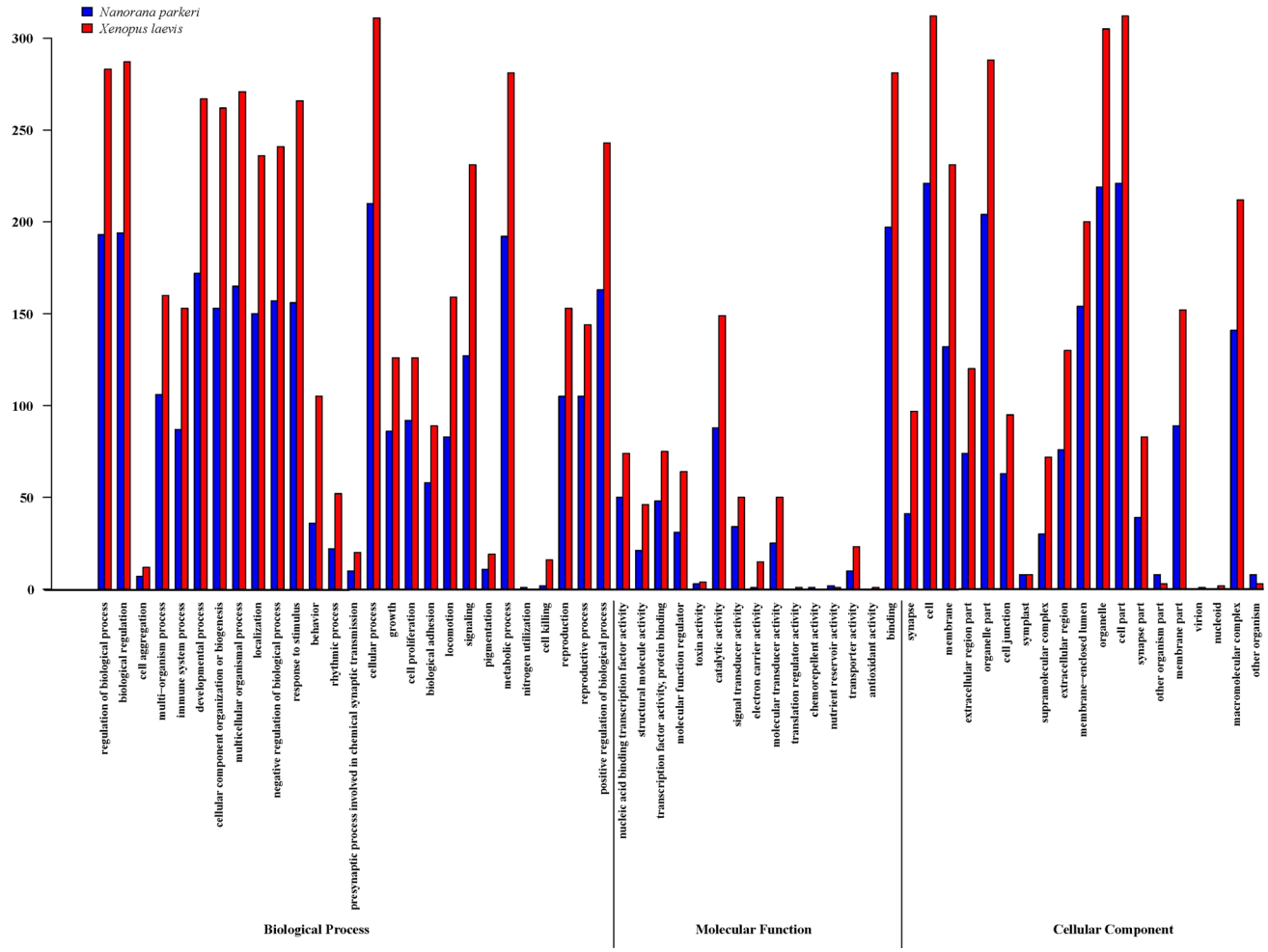



**Figure 3.** GO classifications of CDS with microsatellites in the genomes of *N. parkeri* and *X. laevis*.

Therefore, genomic characteristics of different species could be reflected in their different dominant microsatellite repeats.

Microsatellite expansion and contraction are very sensitive under selective pressure (Tay *et al*. 2010). The insertion and deletion of mono-, di-, tetra-, and pentanucleotide repeat units in coding regions could lead to code shifts (Loire *et al*. 2013). In our study, trinucleotide SSRs in CDS were the most diverse SSR types in both anuran species (figure 1b). Also, the diversity of trinucleotide SSRs in CDS of the *X. laevis* genome is greater than those in the *N. parkeri*, perhaps because *X. laevis* is the more ancient species. Similar results are also found among *Daphnia pulex* (Mayer *et al*. 2010), *B. constrictor* and *P. mucrosquamatus* (Nie *et al*. 2017). This phenomenon may be explained by an increase in trinucleotide repetitions in coding regions that can increase traits diversity, facilitate adaptive changes to environment alterations, and consequently remain fit in their habitats (Trifonov 2004; Loire *et al*. 2013).

The location of SSR in genes were different between *N. parkeri* and *X. laevis* (figure 2). In both species, SSR diversity in upstream and downstream regions was similar. For *N. parkeri*, SSR diversity in exons gradually decreased from the first exon toward the last second exon and then increased toward the last exon. A similar pattern was found in *C. exilicauda* and *M. martensii* which were reported by Wang *et al*. (2016). SSR diversity in introns was similar to exons; however, trends were less prominent. For *X. laevis*, the greatest SSR diversity was found in the last second exon, representing a significant difference between species. The underlying reason may be the larger genome of *X. laevis* (2.72 Gb). SSR diversity in exons shows a bimodal trend with a gradual decrease from the first exon toward the right middle exon, an increase toward the last second exon, and then a decreased again toward the last exon. SSR diversity in various introns was similar.

SSR diversity comparisons in gene regions between the two species indicated that different numbers and diversity of SSR in genes could facilitate adaptation to changing life environments. *N. parkeri* usually lives at high elevation (up to 5000 m) characterized by hypoxia, high levels of UV radiation, and dramatic daily changes in temperature, while *X. laevis* mainly lives in less extreme and variable conditions at lower elevation (Sun *et al*. 2015).

Microsatellites were once regarded as junk, but recent evidence for molecular and phenotypic effects associated with variation in SSR repeat numbers supports the hypothesis that SSRs may modify genes and provide a positive role in adaptive evolution (Kashi and King 2006; Wang *et al*. 2016). Moreover, SSR mutations caused by large expansions and contractions can increase mutation rates and may result in disease (Ellegren 2004; Mirkin 2007). For example, Huntington and Machado–Joseph diseases occurred as a result of a large expansion of trinucleotide coding repeats (Mirkin 2007).

The results of GO analysis found two (GO:0019740 and GO:0045499) and five (GO:0009295, GO:0019012, GO:0016209, GO:0045182 and GO:0098772) unique GO terms for *N. parkeri* and *X. laevis*, respectively, indicating a significant difference in the genomes between species. Many CDS with SSRs are associated with environmental interactions, such as response to stimulus (GO:0050896), cellular processes (GO:0009987), signalling (GO:0023052), metabolic processes (GO:0008152), and catalytic activity (GO:0003824). We also found some obvious enrichment in both species, such as reproductive processes (GO:0000003), developmental processes (GO:0032502), positive (GO:0048518), and negative (GO:0048519) regulation of biological processes. Similar results are also reported in previous study of *C. exilicauda* and *M. martensii* (Wang *et al*. 2016). Based on the previous GO analysis of CDS with SSRs, the SSRs located in CDS can alter the functions of proteins and may provide a molecular basis for species adaptation to new and changing environments (Li *et al*. 2002; Verstrepen *et al*. 2005; Kashi and King 2006; Loire *et al*. 2013). Further, classification and function of genes with SSRs in these two anuran species should continue; results will contribute to further understand their evolutionary history of other anuran species.

#### Acknowledgements

This study was supported by the Natural Science Foundation of Zhejiang, China (LY19C040001) and the Key Research Projects of Lishui City (20151206, SH2017001). The anonymous referees provided helpful insights and comments on the paper.

## Reference

Alam C. M., Singh A. K., Sharfuddin C. and Ali S. 2013 In-silico analysis of simple and imperfect microsatellites in diverse tobamovirus genomes. *Gene* **530**, 193–200.

Behura S. K. and Severson D. W. 2015 Motif mismatches in microsatellites: insights from genome-wide investigation among 20 insect species. *DNA Res.* **22**, 29–38.

Conesa A., Götz S., García-Gómez J. M., Terol J., Talón M. and Robles M. 2005 Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676.

Cui K. and Yue B. S. 2018 Distribution patterns of microsatellites in the genome of *Lophophorus lhuysii*. *Sichuan J. Zool.* **37**, 533–540.

Deback C., Boutolleau D., Depienne C., Luyt C. E., Bonnafous P., Gautheret-Dejean A. *et al*. 2009 Utilization of microsatellite polymorphism for differentiating herpes simplex virus type 1 strains. *J. Clin. Microbiol*. **47**, 533–540.

Demuth J. P. and Drury D. W. 2007 Genome-wide survey of *Tribolium castaneum* microsatellites and description of 509 polymorphic markers. *Mol. Ecol. Notes* **7**, 1189–1195.

Ding S. M., Wang S. P., He K., Jiang M. X. and Li F. 2017 Large-scale analysis reveals that the genome features of simple sequence repeats are generally conserved at the family level in insects. *BMC Genomics* **18**, 848.

Duan Y. N., Liu Y., Hu Y. C., Liu C., Song H. M., Wang X. J. *et al*. 2019 Distribution regularity of microsatellites in *Scleropages formosus* genome. *Chin. Agricul. Sci. Bull*. **35**, 152–158.

Ellegren H. 2004 Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* **5**, 435–445.

Fujimori S., Washio T., Higo K., Ohtomo Y., Murakami K., Matsubara K. *et al.* 2003 A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. *FEBS Lett*. **554**, 17–22.

Gao H., Liu P., Meng X. H., Wang W. J. and Kong J. 2004 Analysis of microsatellite sequences in Chinese shrimp *Fenneropenaeus chinensis* genome. *Oceanol. Limnol. Sin.* **35**, 424–431.

Gao F. T., Shao C. W., Cui Z. K., Wang S. P., Wei M., Chen S. L. and Yang G. P. 2017 Development and population genetic diversity analysis of microsatellite markers in *Epinephelus awoara*. *Period. Ocean Univ. China* **47**, 52–57 (in Chinese with English abstract).

Han B., Wang C. B., Tang Z. H., Ren Y. K., Li Y. L., Zhang D. Y. *et al*. 2015 Genome-wide analysis of microsatellite markers based on sequenced database in Chinese spring wheat (*Triticum aestivum* L.). *PLoS One* **10**, e0141540.

Hancock J. M. 1996 Simple sequences in a 'minimal' genome. *Nat. Genet*. **14**, 14–15.

Hong C. P., Piao Z. Y., Kang T. W., Batley J., Yang T. J., Hur Y. K. *et al*. 2007 Genomic distribution of simple sequence repeats in *Brassica rapa*. *Mol. Cells* **23**, 349–356.

Huntley M. A. and Golding G. B. 2006 Selection and slippage creating serine homopolymers. *Mol. Biol. Evol*. **23**, 2017–2025.

Kashi Y. and King D. G. 2006 Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* **22**, 253–259.

Katti M. V., Ranjekar P. K. and Gupta V. S. 2001 Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol*. **18**, 1161–1167.

Li W. J., Li Y. Z., Du L. M., Huang J., Shen Y. M., Zhang X. Y *et al.* 2014 Comparative analysis of microsatellite sequences distribution in the genome of giant panda and polar bear. *Sichuan J. Zool*. **33**, 874–878.

Li Y. C., Korol A. B., Fahima T. and Nevo E. 2002 Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol*. **11**, 2453–2465.

Lin W. H. and Kussel E. 2012 Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. *Nucleic Acid Res.* **40**, 2399–2413.

Liu L., Qin M. Z., Yang L., Song Z. Z., Luo L., Bao H. Y. *et al*. 2017 A genome-wide analysis of simple sequence repeats in *Apis cerana* and its development as polymorphism markers. *Gene* **599**, 53–59.

Loire E., Higuet D., Netter P. and Achaz G. 2013 Evolution of coding microsatellites in primate genomes. *Genom. Biol. Evol*. **5**, 283–295.

Mayer C., Leese F. and Tollrian R. 2010 Genome-wide analysis of tandem repeats in *Daphnia Pulex* - a comparative approach. *BMC Genomics* **11**, 277.

Mirkin S. M. 2007 Expandable DNA repeats and human disease. *Nature* **447**, 932–940.

Nie H., Cao S. S., Zhao M. L. and Du L. F. 2017 Comparative analysis of microsatellite distributions in genomes of *Boa constrictor* and *Protobothrops mucrosquamatus*. *Sichuan J. Zool.* **36**, 639–648.

Qi W. H., Jiang X. M., Du L. M. and Xiao G. S. 2015 Distribution regularities and comparative analysis of microsatellite in the whole genomes of Yak and water buffalo. *Genom. Appl. Biol.* **34**, 1406–1412.

Qi W. H., Yan C. C., Xiao G. S., Zhang W. Q. and Yue B. S. 2016 Distribution regularities and bioinformatics analysis of microsatellite in the whole genomes of goat and Tibetan antelope. *J. Sichuan Univ.* **53**, 937–944.

Rogers J. and Gibbs R. A. 2014 Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat. Rev. Genet.* **15**, 347–359.

Song Q., Liu J. L. and Guo X. G. 2019 Characterization of microsatellites in *Phrynocephalus axillaris* genome using Roche 454 GS FLX. *Sichuan J. Zool*. **38**, 62–67.

Subramanian S., Mishra R. K. and Singh L. 2003 Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genom. Biol.* **4**, R13.

Sun Y. B., Xiong Z. J., Xiang X. Y., Liu S. P., Zhou W. W., Tu X. L. *et al.* 2015 Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc. Natl. Acad. Sci. USA* **112**, 1257–1262.

Tay W. T., Behere G. T., Batterham P. and Heckel D. G. 2010 Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. *BMC Evol. Biol*. **10**, 144.

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

Tóth G., Gáspóri Z. and Jurka J. 2000 Microsatellites in different eukaryotic genomes: survey and analysis. *Genom. Res*. **10**, 967–981.

Trifonov E. N. 2004 Tuning function of tandemly repeating sequences: molecular device for fast Adaptation. In *Evolutionary theory and processes: modern horizons* (ed. S. P. Wasser), pp. 115–138. Kluwer Academic Publishers, Berlin.

Tu F. Y., Li P., Han W. J. and Huang X. F. 2018a Analysis of microsatellites distribution characteristics in rat (*Rattus norvegicus*) genome. *Genom. Appl. Biol*. **37**, 2374–2382.

Tu Y. F., Liu J., Han W. J., Huang T. and Huang X. F. 2018b Analysis of microsatellite distribution characteristics in the entire genome of *Macaca fascicularis*. *Chin. J. Wildl*. **39**, 400–404.

Verstrepen K. J., Jansen A., Lewitter F. and Fink G. R. 2005 Intragenic tandem repeats generate functional variability. *Nat. Genet*. **37**, 986–990.

Wang C., Du L. M., Li P., Yang M. Y., Li W. J., Shen Y. M. *et al*. 2015 Distribution patterns of microsatellites in the genome of the German cockroach (*Blattella germanica*). *Acta Entomol. Sin*. **58**, 1037–1045.

Wang C., Kubiak L. J., Du L. M., Li W. J., Jian Z. Y., Tang C. *et al*. 2016 Comparison of microsatellite distribution in genomes of *Centruroides exilicauda* and *Mesobuthus martensii*. *Gene* **594**, 41–46.

Wang Z., Huang J., Du L. M., Li W. J., Yue B. S. and Zhang X. Y. 2013 Comparison of microsatellites between the genomes of *Tetranychus urticae* and *Ixodes scapularis*. *Sichuan J. Zool*. **32**, 481–486.

Xu Y. T., Hu Z. X., Wang C., Zhang X. Y., Li J. and Yue B. S. 2016 Characterization of perfect microsatellite based on genome-wide and chromosome level in Rhesus monkey (*Macaca mulatta*). *Gene* **592**, 269–275.

Xu Y. T., Li W. J., Hu Z. X., Zeng T., Shen Y. M., Liu S. X. *et al*. 2018 Genome-wide mining of perfect microsatellites and tetranucleotide orthologous microsatellites estimates in six primate species. *Gene* **643**,124–132.

Zane L., Bargelloni L. and Patarnello T. 2002 Strategies for microsatellite isolation: a review. *Mol. Ecol*. **11**, 1–16.

Corresponding editor: H. A. Ranganath