

RESEARCH

Open Access



# Comparison of microsatellite distribution in the genomes of *Pteropus vampyrus* and *Miniopterus natalensis* (Chiroptera)

Weiwei Shao, Wei Cai, Fen Qiao, Zhihua Lin and Li Wei\*

## Abstract

**Background** Microsatellites are a ubiquitous occurrence in prokaryotic and eukaryotic genomes. Microsatellites have become one of the most popular classes of genetic markers due to their high reproducibility, multi-allelic nature, co-dominant mode of inheritance, abundance and wide genome coverage. We characterised microsatellites in the genomes and genes of two bat species, *Pteropus vampyrus* and *Miniopterus natalensis*. This characterisation was used for gene ontology analysis and the Kyoto Encyclopedia of Genes and Genomes pathway enrichment of coding sequences (CDS).

**Results** Compared to *M. natalensis*, the genome size of *P. vampyrus* is larger and contains more microsatellites, but the total diversity of both species is similar. Mononucleotide and dinucleotide repeats were the most diverse in the genome of the two species. In each bat species, the microsatellite bias was obvious. The microsatellites with the largest number of repeat motifs in *P. vampyrus* from mononucleotide to hexanucleotide were (A)<sub>n</sub>, (AC)<sub>n</sub>, (CAA)<sub>n</sub>, (AAAC)<sub>n</sub>, (AACAA)<sub>n</sub> and (AAACAA)<sub>n</sub>, with frequencies of 97.94%, 58.75%, 30.53%, 22.82%, 54.68% and 22.87%, respectively, while in *M. natalensis* were (A)<sub>n</sub>, (AC)<sub>n</sub>, (TAT)<sub>n</sub>, (TTTA)<sub>n</sub>, (AACAA)<sub>n</sub> and (GAGAGG)<sub>n</sub>, with of 92.00%, 34.08%, 40.36%, 21.83%, 25.42% and 12.79%, respectively. In both species, the diversity of microsatellites was highest in intergenic regions, followed by intronic, untranslated and exonic regions and lowest in coding regions. Location analysis indicated that microsatellites were mainly concentrated at both ends of the genes. Microsatellites in the CDS are thus subject to higher selective pressure. In the GO analysis, two unique GO terms were found only in *P. vampyrus* and *M. natalensis*, respectively. In KEGG enriched pathway, the biosynthesis of other secondary metabolites and metabolism of other amino acids in metabolism pathways were present only in *M. natalensis*. The combined biological process, cellular components and molecular function ontology are reflected in the GO analysis and six functional enrichments in KEGG annotation, suggesting advantageous mutations during species evolution.

**Conclusions** Our study gives a comparative characterization of the genomes of microsatellites composition in the two bat species. And also allow further study on the effect of microsatellites on gene function as well as provide an insight into the molecular basis for species adaptation to new and changing environments.

**Keywords** Genome-wide identification, Microsatellite, Diversity, GO analysis, KEGG enrichment, Chiroptera

## Background

Microsatellites or Simple-Sequence Repeats (SSRs) are tandemly repeated DNA sequences composed of mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide units located throughout the prokaryotic [1] and eukaryotic genomes

\*Correspondence:

Li Wei  
weiliweili2007@163.com  
College of Ecology, Lishui University, Lishui 323000, Zhejiang, People's Republic of China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

[2–4], in both non-coding and coding regions of DNA [5]. Moreover, retrotransposons may also be associated with microsatellites [6]. Furthermore, microsatellites have become one of the most popular classes of genetic markers due to their high reproducibility, multi-allelic nature, co-dominant mode of inheritance, abundance and wide genome coverage [3]. Despite their ubiquitous occurrence, microsatellite density and distribution vary significantly across genomes [7]. Moreover, high mutability at microsatellite loci contributes to genome evolution by creating genetic variation within a gene pool [8, 9]. Slipped-strand mispairing and subsequent error(s) during DNA replication, repair or recombination are the primary cause of this genetic variation [10, 11]. Strand slippage and unequal recombination results in the insertion or deletion of one to several repeated units. This high instability makes them attractive polymorphic molecular markers [12].

In recent years, in silico mining of microsatellite sequences from DNA-sequence databases has rapidly replaced the conventional methods for generating microsatellite markers from genomic libraries [13, 14]. Subsequently, several search tools are available for mining microsatellite repeats in assembled genome sequences, including Tandem Repeats Finder, Simple-Sequence Repeat Identification Tool, Tandem Repeats Occurrence Locator, SciRoko, MSDB and MIncroSATellite (MISA) [3]. MISA is sophisticated and user-friendly microsatellite mining software [15]. Furthermore, MISA was performed for microsatellite mining in the genomes of *Anopheles sinensis* [16], *Epinephelus awoara* [17], *Boa constrictor* and *Protobothrops mucrosquamatus* [18], *Nanorana parkeri* and *Xenopus laevis* [19]. These investigations indicate that microsatellites are found less frequently in protein-coding sequences than in intronic and intergenic regions [18]. Microsatellites in coding regions are more diverse than those in non-coding regions due to higher coding density [20]. The microsatellite length expansion may affect gene regulation, transcription and protein function of coding sequences (CDS), particularly for trinucleotide repeats, which are associated with human diseases [21], such as Huntington and Machado-Joseph disease [22], neurological disease [23] and colorectal cancer [24]. Microsatellite distribution characteristics and functions may vary among genomes [25]. Therefore, whole genome sequencing encourages the development of microsatellite markers derived from the database [3, 26].

In the present study, we investigated the Chiroptera genomes of the large flying fox (*Pteropus vampyrus*) and Natal long-fingered bat (*Miniopterus natalensis*) that have been reported in the open databases. *P. vampyrus* is the largest of any bat species belonging to Yinpterochiroptera

that cannot vocalise echolocation calls [27], whereas *M. natalensis* is a representative species of Yangochiroptera that can produce modulated frequency (FM) echolocation calls [28]. Furthermore, we analysed the characteristics and functional annotation of microsatellites at the genomic level of the two bat species. These findings should contribute to our understanding of the bat genome and facilitates subsequent screening and development of large numbers of high-quality microsatellite markers.

## Methods

The *P. vampyrus* genome assembly was downloaded from the National Center for Biotechnology Information (NCBI) under BioProject accession PRJNA20325, with annotation files downloaded from [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/845/GCF\\_000151845.1\\_Pvam\\_2.0/](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/845/GCF_000151845.1_Pvam_2.0/), including CDS sequences. Similarly, the genome assembly of *M. natalensis* was downloaded from NCBI under BioProject accession PRJNA283550, with annotation files downloaded from [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/595/765/GCF\\_001595765.1\\_Mnat.v1/](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/595/765/GCF_001595765.1_Mnat.v1/), including CDS sequences. Microsatellites in the genome and CDS were identified using MISA identification tool software, which has been used for microsatellite analysis of several species, including *Nanorana parkeri* (high Himalaya frog), *Xenopus laevis* (African clawed frog) [19], *Boa constrictor* (red-tailed boa) and *Protobothrops mucrosquamatus* (brown-spotted pit viper) [18]. Def in the misa.ini file was set as 1–12, 2–6, 3–5, 4–5, 5–4 and 6–4 to restrict the detection criteria for perfect SSR of 1–6 bp with minimum repeat numbers of 12, 6, 5, 5, 4 and 4 for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide microsatellites, respectively [29, 30]. Further, when the distance between two microsatellites was shorter than 100 bp, they were considered single-compound microsatellites [31]. Moreover, repeats with unit patterns being circular permutations and/or reverse complements were considered as one type [32, 33], such as the AAG contains CTT, AGA, TCT, GAA, and TTC or GCGT contains ACGC, CGTG, CACG, GTGC, GCAC, TGCG, and CGCA in different reading frames or on the complementary strand.

Furthermore, the frequency and diversity of SSRs in each bat genome were calculated. The frequency was determined as the percentage of the total number of SSRs per megabase (Mb) of the genome sequence. The diversity of microsatellites, which is the SSR number per Mb of the sequence analysed, was calculated using the methods reported in the literature by Fujimori et al. [31], Qian et al. [34], Nie et al. [18] and Wei et al. [19]. The relative positions of the exon, intron, gene and intergene

regions were extracted from the annotation files via custom Python scripts to explore the distribution of microsatellites in the genomes of *P. vampyrus* and *M. natalensis* [16]. The microsatellites on different regions of the genes were then located. The genes were divided into 13 elements containing 500 bp upstream, the 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 [18, 19]. Further, to avoid overlap in measurements, only genes with more than six exons and five introns were considered [31]. The relative position (from P0.1 to P1.0) 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 [19].

CDS with microsatellites were aligned against NCBI non-redundant and SWISS-PROT protein databases (<http://www.uniprot.org>) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg>), using BLASTx with an E-value threshold of  $1e^{-5}$  [35]. Protein functional annotations were then obtained according to the best alignment results. The Blast2GO software was used to analyse the gene ontology (GO) annotation of genes [36], and WEGO software was employed to investigate the functional classification of genes such as biological processes, cellular components and molecular function [37].

**Results**

**Microsatellite frequency and distribution in the genomes of the two species**

Table 1 shows the results of the microsatellite analysis. A total of 512,647 SSRs were found in the genome assembly of approximately 2.20 Gb for *P. vampyrus*,

and a total of 448,674 SSRs were found in the genome assembly of approximately 1.80 Gb for *M. natalensis*. The SSR content of the genome between species was similar, with 0.46% in *P. vampyrus* and 0.47% in *M. natalensis*. Additionally, the total microsatellite diversity between species was similar, i.e., 233.20 SSRs/Mb in *P. vampyrus* and 248.83 SSRs/Mb in *M. natalensis*. The mononucleotide motifs were the most abundant category, followed by dinucleotide and tetranucleotide motifs for *P. vampyrus*. Whereas in *M. natalensis*, dinucleotide repeats were the most diversified category, followed by mononucleotide and tetranucleotide repeats (Table 1). The most diverse SSR types from mononucleotide to hexanucleotide motifs in the *P. vampyrus* genome were (A)<sub>n</sub>, (AC)<sub>n</sub>, (CAA)<sub>n</sub>, (AAAC)<sub>n</sub>, (AACAA)<sub>n</sub> and (AAACAA)<sub>n</sub> and in *M. natalensis* were (A)<sub>n</sub>, (AC)<sub>n</sub>, (TAT)<sub>n</sub>, (TTTA)<sub>n</sub>, (AACAA)<sub>n</sub> and (GAGAGG)<sub>n</sub>. Moreover, similarities between species were noted in dinucleotide (TA)<sub>n</sub>, (GT)<sub>n</sub>, (GA)<sub>n</sub> and (GC)<sub>n</sub>, trinucleotides (CAT)<sub>n</sub>, tetranucleotides (ATAG)<sub>n</sub> and (CATT)<sub>n</sub>, in pentanucleotide (AACAA)<sub>n</sub>, (TTATT)<sub>n</sub> and (TTTCT)<sub>n</sub> and in hexanucleotide (CTGTCT)<sub>n</sub>. Table 2 shows the concentration of differences in trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide types (Table 2).

The 15 most diverse microsatellite repeats in the *P. vampyrus* genome were (A)<sub>n</sub>, (AC)<sub>n</sub>, (CT)<sub>n</sub>, (TA)<sub>n</sub>, (CAA)<sub>n</sub>, (AAAC)<sub>n</sub>, (TAT)<sub>n</sub>, (AACAA)<sub>n</sub>, (ATAG)<sub>n</sub>, (CATT)<sub>n</sub>, (G)<sub>n</sub>, (TTTA)<sub>n</sub>, (CCTT)<sub>n</sub>, (CAT)<sub>n</sub> and (GAG)<sub>n</sub> comprising of 92.84% of all microsatellites identified. Similarly, the 15 most diverse microsatellite motifs in *M. natalensis* were (A)<sub>n</sub>, (CT)<sub>n</sub>, (AC)<sub>n</sub>, (TA)<sub>n</sub>, (G)<sub>n</sub>, (TAT)<sub>n</sub>, (TTTA)<sub>n</sub>, (ATAG)<sub>n</sub>, (CATT)<sub>n</sub>, (CCTT)<sub>n</sub>, (CAA)<sub>n</sub>, (TGGA)<sub>n</sub>, (AACAA)<sub>n</sub>, (AAAC)<sub>n</sub> and (TTATT)<sub>n</sub> comprising of 94.10% of all microsatellites identified.

**Table 1** Distribution of microsatellites in the genomes of *Pteropus vampyrus* and *Miniopterus natalensis*

Motif length	<i>Pteropus vampyrus</i>				<i>Miniopterus natalensis</i>			
	Numbers of microsatellites	Length(bp)	Abundance (SSRs/Mb)	Frequency (%)	Numbers of microsatellites	Length(bp)	Abundance (SSRs/Mb)	Frequency (%)
Mononucleotide	246,947	3,647,964	112.34	48.17	144,835	2,174,691	80.33	32.28
Dinucleotide	163,249	3,649,342	74.26	31.84	235,344	4,030,076	130.52	52.45
Trinucleotide	36,521	750,138	16.61	7.12	20,959	386,283	11.62	4.67
Tetranucleotide	43,966	1,409,268	20.00	8.58	32,493	1,259,172	18.02	7.24
Pentanucleotide	15,137	382,635	6.89	2.95	10,320	367,900	5.72	2.30
Hexanucleotide	6827	199,332	3.11	1.33	4723	188,970	2.62	1.05
Total	512,647	10,038,679	233.20	100.00	448,674	8,407,092	248.83	100.00
Whole genome length/bp	2,198,284,804				1,803,099,001			
SSR content in the genome	0.46%				0.47%			

**Table 2** The most frequent microsatellite motifs found in the genomes of *Pteropus vampyrus* and *Miniopterus natalensis*

Motif length	<i>Pteropus vampyrus</i>			<i>Miniopterus natalensis</i>		
	Repeat unit	Microsatellites	Frequency (%)	Repeat unit	Microsatellites	Frequency (%)
Mononucleotide	A	241,850	97.94	A	133,249	92.00
	G	5097	2.06	G	11,586	8.00
Dinucleotide	AC	95,909	58.75	AC	80,208	34.08
	CT	37,060	22.70	CT	126,869	53.91
	GC	1394	0.85	GC	467	0.20
Trinucleotide	TA	28,886	17.69	TA	27,800	11.81
	CAA	11,151	30.53	TAT	8458	40.36
	TAT	9997	27.37	CAA	3341	15.94
	CAT	4202	11.51	CAT	2508	11.97
Tetranucleotide	GAG	2974	8.14	ACC	2380	11.36
	AAAC	10,035	22.82	TTTA	7092	21.83
	ATAG	6429	14.62	ATAG	5488	16.89
	CATT	5268	11.98	CATT	3802	11.70
	TTTA	4488	10.21	CCTT	3721	11.45
Pentanucleotide	AACAA	8277	54.68	AACAA	2623	25.42
	TTATT	2174	14.36	TTATT	2515	24.37
	TTTCT	851	5.62	TTTCT	621	6.02
	CCACC	295	1.95	AGGGA	606	5.87
Hexanucleotide	AAACAA	1561	22.87	GAGAGG	604	12.79
	GGGTTA	1282	18.78	TATCTA	271	5.74
	CTGTCT	442	6.47	CTGTCT	261	5.53
	TATCTA	414	6.06	GGGTTA	215	4.55

Table 3 displays the distributions of microsatellites in the genomes of *P. vampyrus* and *M. natalensis*. Intergenic regions had the most numbers of microsatellites, and CDS exhibited a few in both species. The number of microsatellites in the intergenic, intron, exon and untranslated regions of *P. vampyrus* was greater than that in *M. natalensis*; however, the diversity of microsatellites in intron regions of *P. vampyrus* was less than that in *M. natalensis*. The numbers and diversity of microsatellites in CDS in *M. natalensis* were larger than those in *P. vampyrus*. Further, microsatellites in the CDS were found to be less diverse than those in other regions. Figure 1 illustrates the frequency of different microsatellite types in different genomic regions. In both species, trinucleotides were the most diverse microsatellite type

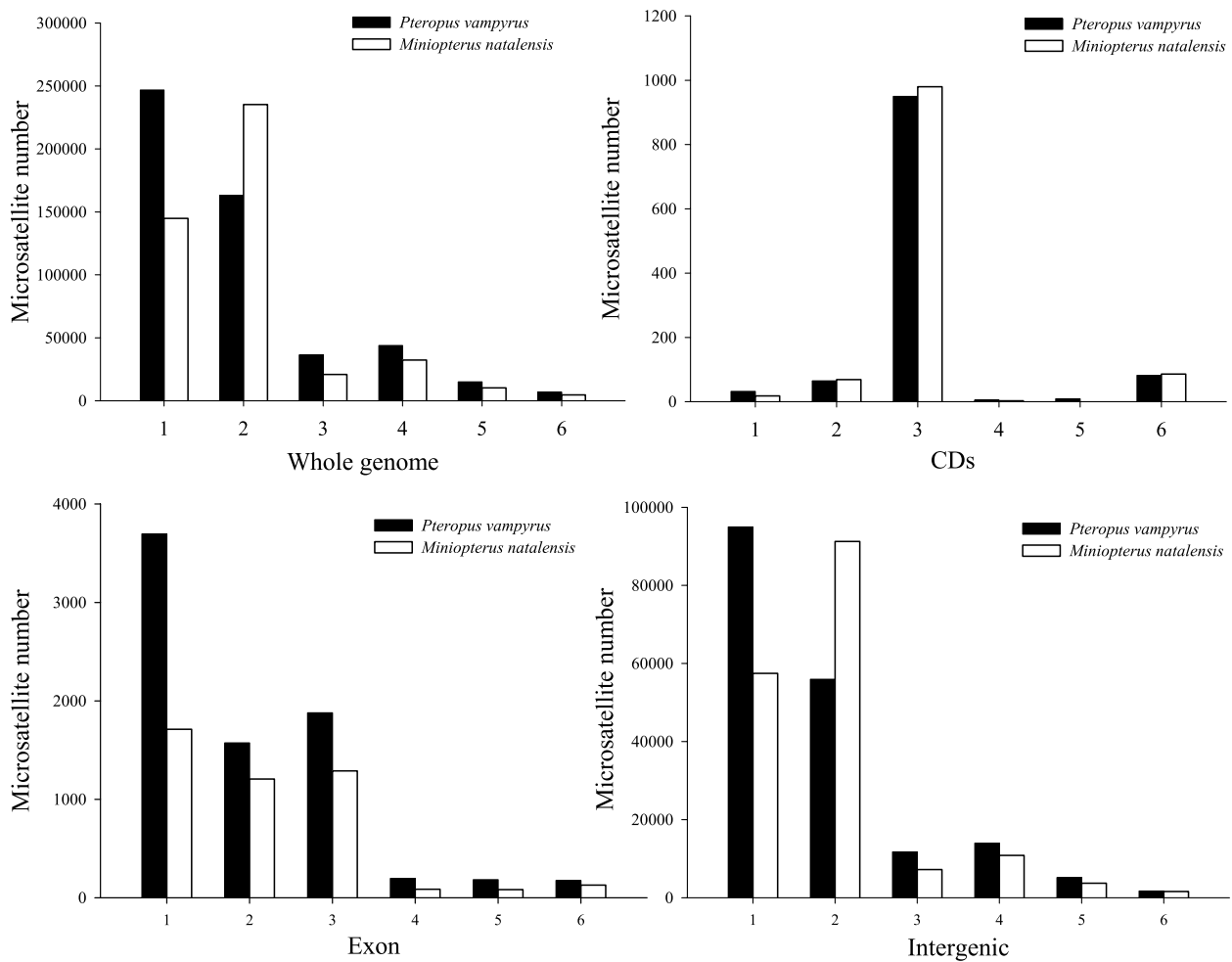
in CDS, with 83.11% and 84.70% in *P. vampyrus* and *M. natalensis*, respectively. The numbers of mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide in the exons of *P. vampyrus* were much greater than that of *M. natalensis*. The distribution of SSRs in intergenic regions was similar to the distribution in whole genomes, with the most diversity among mononucleotides and dinucleotides.

**Location analysis of microsatellites in genes**

All microsatellites in exons or introns were compared with 979 and 1010 genes, with more than six exons and five introns in *P. vampyrus* and *M. natalensis*, respectively. Microsatellite-enriched regions were upstream and downstream of genes in both *P. vampyrus* and *M.*

**Table 3** The number and diversity (microsatellites/Mb) of microsatellites in different genomic regions of *Pteropus vampyrus* and *Miniopterus natalensis*

Species	Gene				Intergenic
	CDS	Untranslated	Exon	Intron	
<i>Pteropus vampyrus</i>	1143(355.55)	4710(1537.15)	7702(1226.66)	183,514(2244.51)	402,059(3050.79)
<i>Miniopterus natalensis</i>	1157(371.89)	2953(1323.03)	4503(842.76)	171,977(2436.11)	292,798(2805.33)



**Fig. 1** Distribution of microsatellite types in different genomic regions of *Pteropus vampyrus* and *Miniopterus natalensis*. 1–6 indicated mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide unit length, respectively

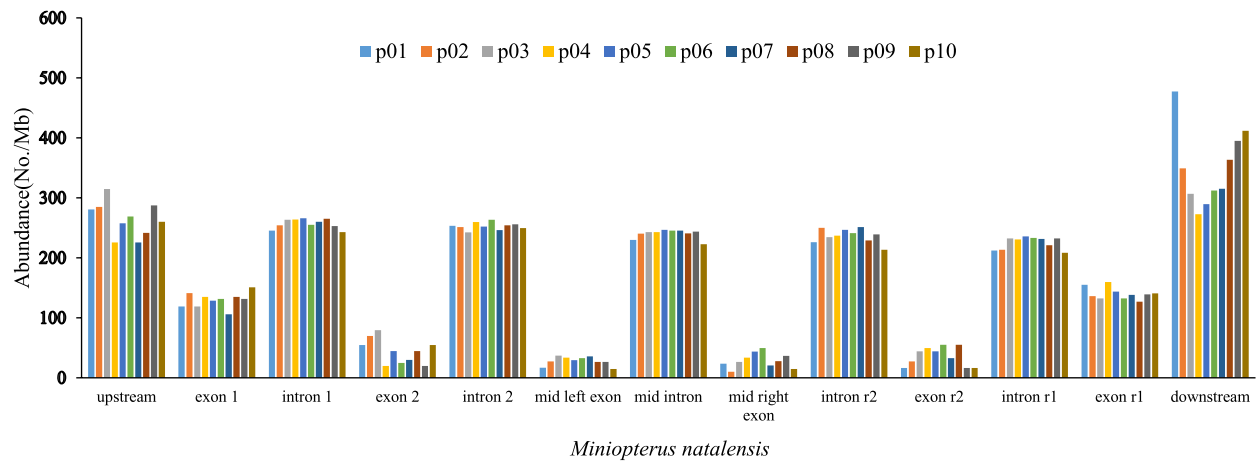
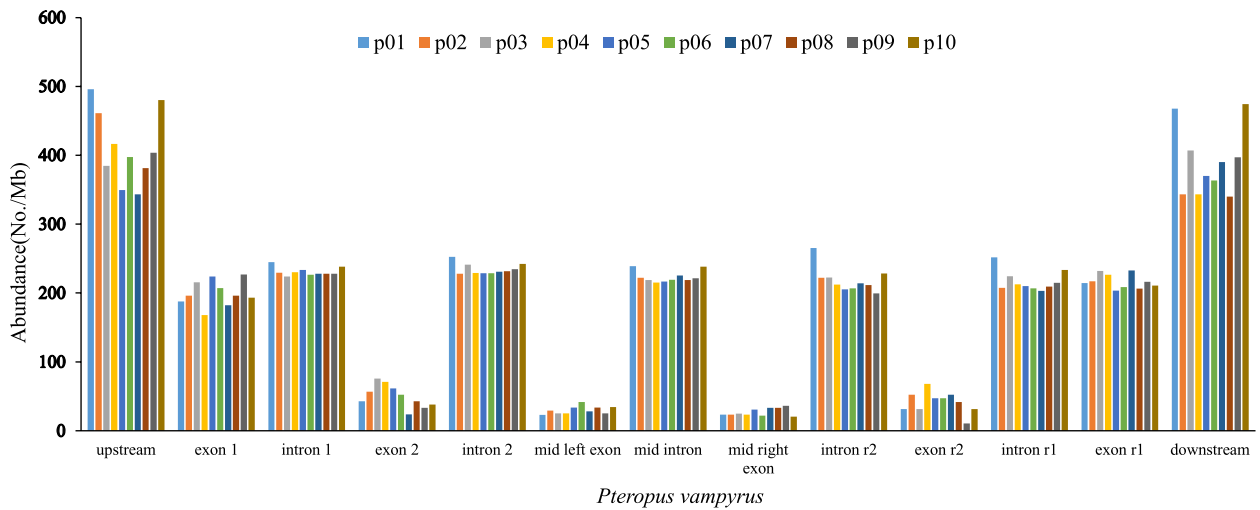
*natalensis* genomes, with the numbers of microsatellites in exons, gradually decreasing from the first exon toward the last second exon and increasing toward the last exon (Fig. 2). In each bat species, microsatellite diversity in upstream and downstream regions was similar. Likewise, microsatellite diversity in various introns was also similar (Fig. 2).

#### Functional analysis of CDS with microsatellites for two species

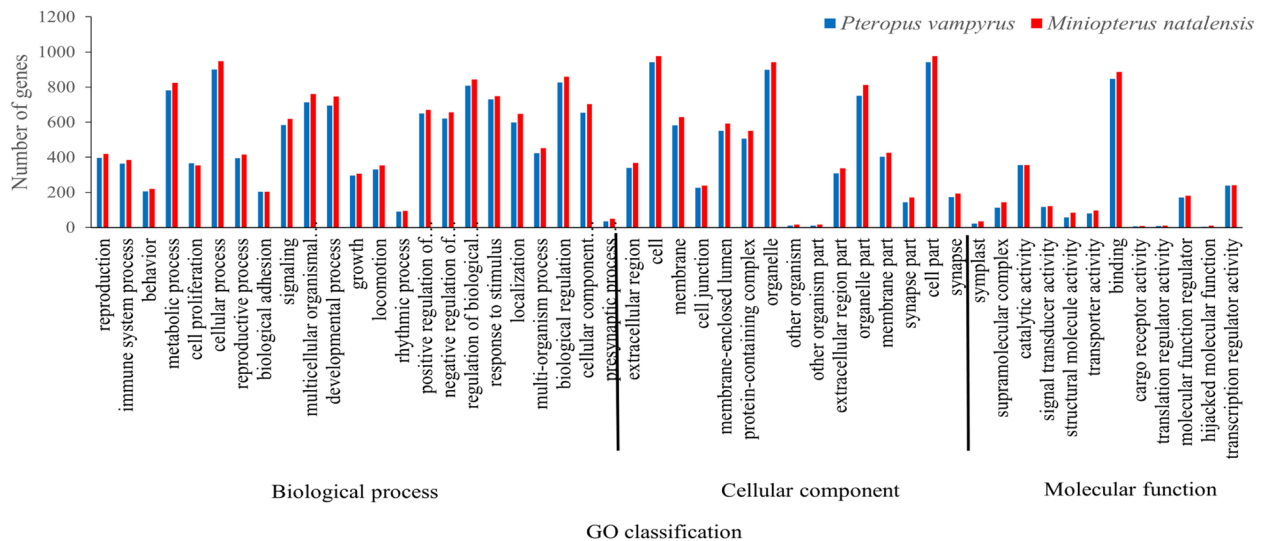
In genomes of *P. vampyrus* and *M. natalensis*, 1019 and 1043 CDS with SSR, respectively, were imported into GO analysis based on sequence alignment. All these CDS were assigned to 20572 (*P. vampyrus*) and 21816 (*M. natalensis*) GO in terms of their known functions. Figure 3 shows the number of CDS with SSRs assigned to each subcategory. Further, 50 pairs were represented in both species of these GO functional classifications. Carbon utilisation (GO: 0015976) and biological phase (GO:

0044848) in the biological process ontology were only present in *P. vampyrus*, while the virion (GO: 0019012) and virion part (GO: 0044423) in cellular component ontology were present only in *M. natalensis*. Furthermore, comparing the function distribution between the two species, cellular process (GO: 0009987) in biological process ontology was most frequent. Cell (GO: 0005623) and cell part (GO: 0044464) were the top two terms in the cellular component ontology. In the molecular function ontology, binding (GO: 0005488) was prominent.

CDS were assigned to 828 for *P. vampyrus* and 847 for *M. natalensis* in terms of known functions for KEGG annotation. Figure 4 shows these KO functional classifications indicating that 41 and 43 pathways were enriched in *P. vampyrus* and *M. natalensis*, respectively. All the enrichment pathways were divided into six functional classification categories, i.e., metabolism, environmental information processing, genetic information processing, cell process, organismal

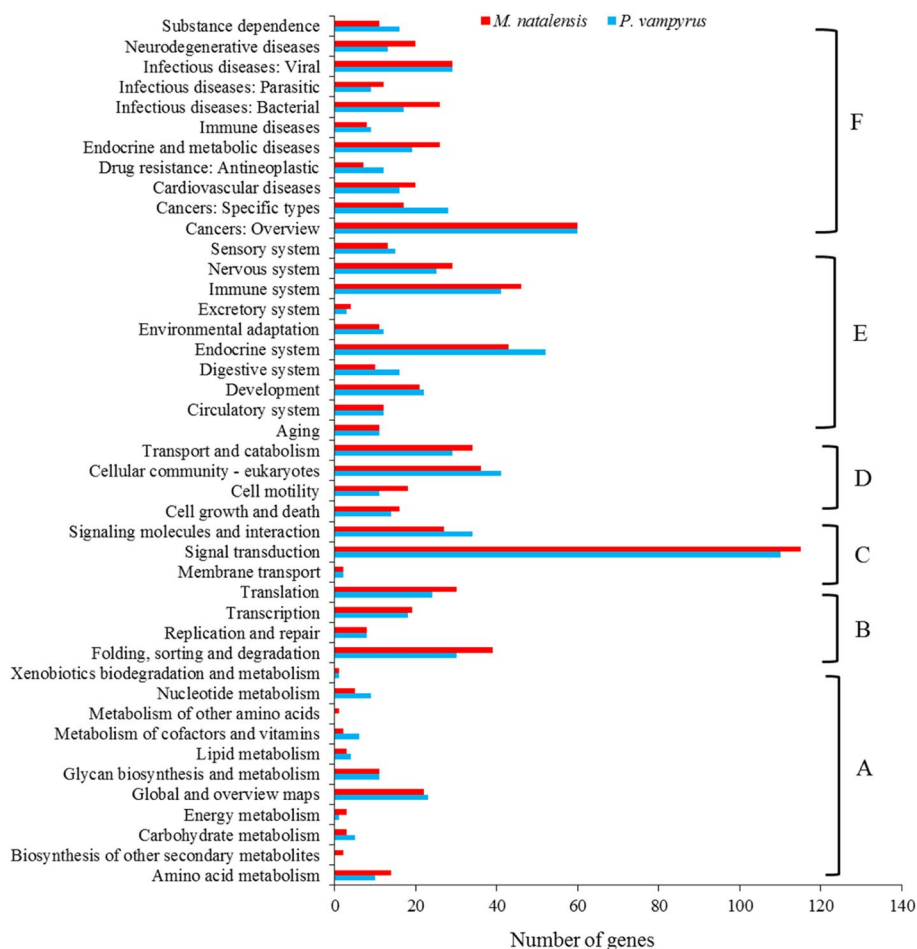


**Fig. 2** Microsatellite abundance in gene regions and their upstream and downstream regions of *Pteropus vampyrus* and *Miniiopteris natalensis*



**Fig. 3** GO classifications of coding sequencing (CDS) with microsatellites in the genomes of *Pteropus vampyrus* and *Miniiopteris natalensis*





**Fig. 4** KEGG enrichment of microsatellites with CDS in *Pteropus vampyrus* and *Miniopterus natalensis*: (A) Metabolism, (B) Environmental information processing, (C) Genetic information processing, (D) Cell process, (E) Organismal systems and (F) Human diseases and drug development

systems and human diseases and drug development (Fig. 4). The biosynthesis of other secondary metabolites and metabolism of other amino acids in metabolism pathways were present only in *M. natalensis*. Among these pathways, the signal transduction pathway was the most enriched, with 110 genes in *P. vampyrus* and 115 genes for *M. natalensis*.

**Discussion**

Genome-wide identification of SSR markers have been successfully performed in various animals [38]. To our best knowledge, the present study is the comprehensive report on the characterization of microsatellites in bat species of *P. vampyrus* and *M. natalensis*. Genome size, total number of SSR and total length of SSR identified in *P. vampyrus* were all larger than those in *M. natalensis* (Table 1). These differences in genomes of the two species may be caused by their genome size, assembly quality, the number of positions of the unknown base and specificity of species [3, 39]. This phenomenon has been reported

in other species, such as *B. constrictor* and *P. mucrosquamatus* [18], *Tetranychus urticae* and *Ixodes scapularis* [40] and *Phytophthora* [41]. However, microsatellite content in the genomes of *P. vampyrus* and *M. natalensis* was similar, accounting for 0.46% and 0.47%, respectively. This result is consistent with other bat species *Rhinolophus ferrumequinum* (0.58%, unpublished data) and *Hipposideros armiger* (0.50%, unpublished data), as well as previous studies in other mammals, such as giant panda (*Ailuropoda melanoleuca*, 0.64%), the polar bear (*Ursus maritimus*, 0.79%) [42] and forest musk deer (*Moschus berezovskii*, 0.42%) [43]. Total SSR diversity in the genomes of *P. vampyrus* and *M. natalensis* are 233.20 SSRs/Mb and 248.83 SSRs/Mb, respectively, which were lower in comparison to the diversity of *R. ferrumequinum* with 263.65 SSRs/Mb (unpublished data) but higher compared to the diversity of *H. armiger* (222.61 SSRs/Mb (unpublished data). This indicates that the genomic size and quality of sequencing have a great influence on the identification of microsatellites [18].

The sequence proportions of six SSR types in *P. vampyrus* and *M. natalensis* genomes are different, as are the four most diverse microsatellite types (Table 2). This result has also been reported in patterns of genomic SSRs of *N. parkeri* and *X. laevis* [19], *B. constrictor* and *P. mucrosquamatus* [18], *C. exilicauda* and *M. martensii* [44]. However, genomes of *Eucryptorrhynchus brandti* and *E. scrobiculatus* exhibit similarities in the six SSR types [45] suggesting that the differences and similarities in microsatellite composition in the genome can reflect the relationship among species to some extent [46]. Frequency and abundance analysis of various motif repeats in *P. vampyrus* genome revealed that mononucleotide repeats were the dominant type of SSRs (Table 1). These results are in agreement with previous studies in other eukaryotic organisms. For example, mononucleotide was the dominant SSR types in *Lophophorus lhuysii* [47], *M. berezovskii* [43] and *Macaca fascicularis* [48]. On the contrary, dinucleotide was the dominant SSR types in the genome of *M. natalensis*, which is in agreement with other species of *N. parkeri* and *X. laevis* [19], *Rhodeus sinensis* [49] and *Eriocheir sinensis* [50]. Dinucleotides were the dominant types because of their higher mutation rates [37]. For example, dinucleotides in human non-pathogenic SSR loci have mutation rates of 1.5–2 times higher than tetranucleotides [51].

In comparisons with *P. vampyrus* and *M. natalensis*, differences in both frequency and diversity of SSRs in CDS were minor, whereas those in exon, intron, untranslated and intergenic regions were significant (Table 3). Furthermore, the diversity of microsatellites in untranslated regions was greater than those in CDS regions, indicating that microsatellites aggregate in untranslated regions, presumably influencing gene transcriptional activity [52]. Coding regions are generally conservative among different species and are subject to high-selective pressure [53]. In this study, trinucleotide SSRs in the CDS were the most diverse SSR types in both bat species. Further, the diversity of trinucleotide SSRs in the CDS of the *M. natalensis* genome is greater than that in the *P. vampyrus*, possibly due to the faster rate of evolution of *M. natalensis*. This phenomenon could be explained by an increase in trinucleotide repetitions in coding regions, which can increase trait diversity and facilitate adaptive changes in response to environmental alterations [54]. Therefore, the characteristics of microsatellite repeats in the genomes of various species could be reflected in their different dominants [3].

*P. vampyrus* and *M. natalensis* had different SSR locations in genes (Fig. 2). SSRs in the upstream and downstream regions of both species were similar, with the highest diversity. Instead, SSR diversity in upstream and downstream regions of *P. vampyrus* was greater than

in *M. natalensis*, predicting the underlying reason for the larger genome size of *P. vampyrus*. In each species, SSR diversity in exons showed a “U” shape that gradually decreased from the first exon toward the last second exon and then increased toward the last exon. This phenomenon is consistent with *C. exilicauda* and *M. martensii* reported by Wang et al. [44], and *B. constrictor* and *P. mucrosquamatus* reported by Nie et al. [18], respectively. SSR diversity in various introns was similar in each of the two species. Therefore, comparisons of SSR diversity in gene regions between the two species revealed that different numbers and diversity of SSR in genes may facilitate adaptation to evolutionary history. *P. vampyrus* is a fruit-eating bat that usually roosts in trees and has non-echolocation calls, whereas *M. natalensis* is an insectivorous bat with echolocation calls that primarily live in caves and mines that are used for hibernation and reproduction [27].

For functional annotation of coding genes, GO analysis found two (GO: 0015976 and GO: 0044848) for *P. vampyrus* and two (GO: 0019012; GO: 0044423) unique GO terms for *M. natalensis*, respectively, indicating a significant difference in the genomes between species. Moreover, many CDS with SSRs are associated with environmental interactions, such as metabolic processes (GO: 0008152), cellular processes (GO: 0009987), signalling (GO: 0023052) and response to stimulus (GO:0050896), which may be related to the different adaptability to the environment of the two bats. This pattern is also reported in a study of *N. parkeri* and *X. laevis* [19]. In KEGG annotation, 41 and 43 pathways were enriched in *P. vampyrus* and *M. natalensis*, respectively. We found that two (Biosynthesis of other secondary metabolites and metabolism of other amino acids) unique metabolism pathways were presented only in *M. natalensis*, which may further indicate some significantly different functions in the genes between species. In both species, genetic information processing has the fewest pathways, with only 3 pathways containing 146 genes in *P. vampyrus* and 144 genes in *M. natalensis*. Human diseases and drug development have the most pathways, with 11 pathways containing 228 genes in *P. vampyrus* and with 9 pathways containing 236 genes in *M. natalensis*, respectively, suggesting that bats are one of the most important natural hosts of mammalian viruses [55]. There are 28 families of viruses found in bats [56]. A recent study showed that the homology of the outbreak of the new coronavirus (Covid-19) in late 2019 is 79% compared to SARS-CoV at the genome-wide level and up to 89% compared to SARRr ZC45 sampled from a *Rhinolophus* bat in Zhejiang, China [57]. As different coronaviruses recombine to produce new viruses, SSRs in the genes of bats may evolve in adaptive changes to internal alterations and, consequently, remain fit in zoonosis [58–60].



## Conclusions

As summarised above, characteristics of microsatellites at the genomic level of *P. vampyrus* and *M. natalensis* were analysed and compared in this study. Further, the classification and functional evolution of genes with SSRs in these two bat species should continue; results will contribute to a further understanding of the evolutionary history of other Chiroptera species.

## Acknowledgements

The authors thank Ming Lei for his data analysis assistance and the anonymous referees provided helpful insights and comments on the paper.

## Authors' contributions

WWS, FQ and LW were involved in the design of the study, bioinformatics analysis and manuscript writing, WC and ZHL contributed to the bioinformatics work and helped to draft the manuscript. All authors read and approved the final manuscript.

## Funding

This study was supported by the Key Research Projects of Lishui City (2021ZDYF05; 2020ZDYF07) that is provided by Li Wei and National Natural Science Foundation of China (31901860) that is provided by Fen Qiao.

## Availability of data and materials

The datasets generated and/or analysed during the current study are available in the National Center for Biotechnology Information (NCBI) repository. The *Pteropus vampyrus* genome assembly was downloaded from BioProject accession PRJNA20325, with annotation files downloaded from [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/845/GCF\\_000151845.1\\_Pvam\\_2.0/](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/151/845/GCF_000151845.1_Pvam_2.0/), including CDS sequences. Similarly, the genome assembly of *Miniopterus natalensis* was downloaded from BioProject accession PRJNA283550, with annotation files downloaded from [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/595/765/GCF\\_001595765.1\\_Mnat.v1/](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/595/765/GCF_001595765.1_Mnat.v1/), including CDS sequences.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 3 October 2022 Accepted: 23 January 2023

Published online: 14 February 2023

## References

- Sreenu VB, Vishwanath A, Javaregowda N, Nagarajaram HA. MICdb: database of prokaryotic microsatellites. *Nucleic Acids Res.* 2003;31:106–8.
- Tóth G, Góspóri Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 2000;10:967–81.
- Sharma PC, Grover A, Kahl G. Mining microsatellites in eukaryotic genomes. *Trends Biotechnol.* 2007;25:490–8.
- Labiros DA, Catalig A, Ymbong R, Sakuntabhai A, Llusma AO, Edillo FE. Novel and broadly applicable microsatellite markers in identified chromosomes of the philippine dengue mosquitoes, *Aedes aegypti* (diptera: culicidae). *J Med Entomol.* 2022;59:545–53.
- Beckman JS, Weber JL. Survey of human and rat microsatellites. *Genomics.* 1992;12:627–31.
- Tay WT, Behere GT, Batterham P, Heckel DG. Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. *BMC Evol Biol.* 2010;10:144.
- Dieringer D, Schlotterer C. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res.* 2003;13:2242–51.
- Gow JL, Noble LR, Rollinson D, Jones CS. A high incidence of clustered microsatellite mutations revealed by parent-offspring analysis in the African freshwater snail, *Bulinus forskalii* (Gastropoda, Pulmonata). *Genetica.* 2005;124:77–83.
- Bae JH, Zhang DY. Predicting stability of DNA bulge at mononucleotide microsatellite. *Nucleic Acid Res.* 2021;49:7901–8.
- Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol.* 1987;4:203–21.
- Huntley MA, Golding GB. Selection and slippage creating serine homopolymers. *Mol Biol Evol.* 2006;23:2017–25.
- Deback C, Boutolleau D, Depienne C, Luyt CE, Bonnafous P, Gautheret-Dejean A, Garrigue I, Agut H. Utilization of microsatellite polymorphism for differentiating herpes simplex virus type 1 strains. *J Clin Microbiol.* 2009;47:533–40.
- Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33:2583–5.
- Jo E, Lee SJ, Choi E, Kim J, Lee SG, Lee JH, Kim JH, Park H. Whole genome survey and microsatellite motif identification of *Artemia franciscana*. *Biosci Rep.* 2021;41:BSR20203868.
- Sarika AV, Iqbal MA, Rai A, Kumar D. Pipemicrodb: microsatellite database and primer generation tool for pigeonpea genome. *Database.* 2013;3:bas054.
- Wang XT, Zhang YJ, He X, Mei T, Chen B. Identification, characteristics and distribution of microsatellites in the whole genome of *Anopheles sinensis* (Diptera: Culicidae). *Acta Entomol Sin.* 2016;59:1058–68.
- Gao FT, Shao CW, Cui ZK, Wang SP, Wei M, Chen SL, Yang GP. Development and population genetic diversity analysis of microsatellite markers in *Epinephelus awoara*. *Periodi Ocean Uni Chin.* 2017;47:52–7.
- Nie H, Cao SS, Zhao ML, Du LF. Comparative analysis of microsatellite distributions in genomes of *Boa constrictor* and *Protobothrops mucrosquamatus*. *Sichuan J Zool.* 2017;36:639–48.
- Wei L, Shao WW, Ma L, Lin ZH. Genomewide analysis of microsatellite markers based on sequenced database in two anuran species. *J Genet.* 2020;99:58.
- Alam CM, Singh AK, Sharfuddin C, Ali S. In-silico analysis of simple and imperfect microsatellites in diverse tobamovirus genomes. *Gene.* 2013;530:193–200.
- Collaborative R. Impact of microsatellite status in early-onset colonic cancer. *British J Surg.* 2022;109:632–6.
- Mirkin SM. Expandable DNA repeats and human disease. *Nature.* 2007;447:932–40.
- Brouwer JR, Willemsen R, Oostra BA. Microsatellite repeat instability and neurological disease. *BioEssays.* 2009;31:71–83.
- Yang Q, Huang G, Li L, Li E, Xu L. Potential mechanism of immune evasion associated with the master regulator ascl2 in microsatellite stability in colorectal cancer. *J Immunol Res.* 2021;2021:5964752.
- Shi J, Huang S, Fu D, Yu J, Wang X, Wei H, Liu S, Liu G, Wang H, Alexander VB. Evolutionary dynamics of microsatellite distribution in plants: insight from the comparison of sequenced brassica, arabidopsis and other angiosperm species. *PLoS One.* 2013;8:e59988.
- Oreshkova NV, Putintseva YA, Sharov VV, Kuzmin DA, Krutovsky KV. Development of microsatellite genetic markers in siberian larch (*Larix sibirica* Ledeb.) based on the de novo whole genome sequencing. *Russian J Genet.* 2017;53:1194–9.
- Taylor M. Bats: an illustrated guide to all species. Brighton: Ivy Press; 2019.
- Miller-Butterworth CM, Geeta E, Jacobs DS, Corrie SM, Harley EH. Genetic and phenotypic differences between south African long-fingered bats, with a global minioptera phylogeny. *J Mammal.* 2005;6:1121–35.
- Demuth JP, Drury DW. Genome-wide survey of *Tribolium castaneum* microsatellites and description of 509 polymorphic markers. *Mol Ecol Notes.* 2007;7:1189–95.

30. Song Q, Liu JL, Guo XG. Characterization of microsatellites in *Phrynocephalus axillaris* genome using Roche 454 GS FLX. *Sichuan J Zool.* 2019;38:62–7.
31. Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S, Tomita M. A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. *FEBS Lett.* 2003;554:17–22.
32. Li RQ, Fan W, Tian G, Zhu H, He L, Cai J. The sequence and de novo assembly of the giant panda genome. *Nature.* 2009;463:311–7.
33. Huang J, Li YZ, Du LM, Yang B, Shen FJ, Zhang HM, Zhang ZH, Zhang XJ, Yue BS. Genome-wide survey and analysis of microsatellites in giant panda (*Ailuropoda melanoleuca*), with a focus on the applications of a novel microsatellite marker system. *BMC Genomics.* 2015;16:61.
34. Qian J, Xu HB, Song JY, Xu J, Zhu YJ, Chen L. Genome-wide analysis of simple sequence repeats in the model medicinal mushroom *Ganoderma lucidum*. *Gene.* 2013;512:331–6.
35. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28:27–30.
36. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21:3674–6.
37. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 2006;34:293–7.
38. Fan SG, Huang H, Liu Y, Wang PF, Zhao C, Yan LL, Qiao XT, Qiu LH. Genome-wide identification of microsatellite and development of polymorphic SSR markers for spotted sea bass (*Lateolabrax maculatus*). *Aquacult Rep.* 2021;20:100677.
39. Neafsey DE. Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes. *Genome Res.* 2003;13:821–30.
40. Wang Z, Huang J, Du LM, Li WJ, Yue BS, Zhang XY. Comparison of microsatellites between the genomes of *Tetranychus urticae* and *Ixodes scapularis*. *Sichuan J Zool.* 2013;32:481–6.
41. Garnica DP, Pinzón AM, Quesada-Ocampo LM, Bernal AJ, Barreto, Grünwald NJ, Restrepo S. Survey and analysis of microsatellites from transcript sequences in *phytophthora* species: frequency, distribution, and potential as markers for the genus. *BMC Genomics.* 2006;7:245.
42. Li WJ, Li YZ, Du LM, Huang J, Shen YM, Zhang XY, Yue BS. Comparative analysis of microsatellite sequences distribution in the genome of giant panda and polar bear. *Sichuan J Zool.* 2014;33:874–8.
43. Lu T, Wang C, Du C, Liu, Shen YM, Zhang XY, Yue BS. Distribution regularity of microsatellites in *Moschus berezovskii* genome. *Sichuan J Zool.* 2017;36:420–4.
44. Wang C, Kubiak LJ, Du LM, Li WJ, Jian ZY, Tang C, Fnan ZX, Zhang XY, Yue BS. Comparison of microsatellite distribution in genomes of *Centruroides exilicauda* and *Mesobuthus martensii*. *Gene.* 2016;594:41–6.
45. Zhang YJ, Song W, Chen JC, Cao LJ, Wen JB, Wei SJ. Genome-wide characterization of microsatellites and development of polymorphic markers shared between two weevils of *Eucryptorrhynchus* (Coleoptera: Curculionidae). *Zool System.* 2021;46:273–80.
46. Ding SM, Wang SP, He K, Jiang MX, Li F. Large-scale analysis reveals that the genome features of simple sequence repeats are generally conserved at the family level in insects. *BMC Genomics.* 2017;18:848.
47. Cui K, Yue BS. Distribution patterns of microsatellites in the genome of *Lophophorus lhuysii*. *Sichuan J Zool.* 2018;37(5):533–40.
48. Tu FY, Liu J, Han WJ, Huang T, Huang XF. Analysis of microsatellite distribution characteristics in the entire genome of *Macaca fascicularis*. *Chin J Wildl.* 2018;39:400–4.
49. Xiong LW, Wang SB, Feng Q, Wang JG, Yue J, Zhang J, Wu YF, Wang Q. Characterization and development of microsatellite in the genome of *Rhodeus sinensis* based on high throughput sequencing. *Jiangsu Agricul Sci.* 2018;46:164–8.
50. Xiong LW, Wang Q, Qiu GF. Large-scale isolation of microsatellites from Chinese mitten crab *Eriocheir sinensis* via a solexa genomic survey. *Inter J Mol Sci.* 2012;13:16333–45.
51. Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc Natl Acad Sci.* 1997;94:1041–6.
52. Ellegren H. Microsatellite: simple sequences with complex evolution. *Nat Rev Genet.* 2004;5:435–45.
53. Lin WH, Kussel E. Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. *Nucleic Acids Res.* 2012;40:2399–413.
54. Loire E, Higuete D, Netter P, Achaz G. Evolution of coding microsatellites in primate genomes. *Genom Biol Evol.* 2013;5:283–95.
55. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. Global trends in emerging infectious diseases. *Nature.* 2008;451:990–3.
56. Moratelli R, Calisher CH. Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? *Mem Inst Oswal do Cruz.* 2015;110:1–22.
57. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. A new coronavirus associated with human respiratory disease in china. *Nature.* 2020;579:1–8.
58. Jiang TL, Zhao HB, He B, Zhang LB, Luo JH, Liu Y, Sun KP, Yu WH, Wu Y, Feng J. Research progress of bat biology and conservation strategies in China. *Acta Theriol Sin.* 2020;40:539–59.
59. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 2019;28:1947–51.
60. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 2021;49:D545–51.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

