

RESEARCH ARTICLE

Open Access



Microsatellite development from genome skimming and transcriptome sequencing: comparison of strategies and lessons from frog species

Yun Xia¹, Wei Luo^{1,2}, Siqi Yuan^{1,3}, Yuchi Zheng¹ and Xiaomao Zeng^{1*} 

Abstract

Background: Even though microsatellite loci frequently have been isolated using recently developed next-generation sequencing (NGS) techniques, this task is still difficult because of the subsequent polymorphism screening requires a substantial amount of time. Selecting appropriate polymorphic microsatellites is a critical issue for ecological and evolutionary studies. However, the extent to which assembly strategy, read length, sequencing depth, and library layout produce a measurable effect on microsatellite marker development remains unclear. Here, we use six frog species for genome skimming and two frog species for transcriptome sequencing to develop microsatellite markers, and investigate the effect of different isolation strategies on the yield of microsatellites.

Results: The results revealed that the number of isolated microsatellites increases with increased data quantity and read length. Assembly strategy could influence the yield and the polymorphism of microsatellite development. Larger k-mer sizes produced fewer total number of microsatellite loci, but these loci had a longer repeat length, suggesting greater polymorphism. However, the proportion of each type of nucleotide repeats was not affected; dinucleotide repeats were always the dominant type. Finally, the transcriptomic microsatellites displayed lower levels of polymorphisms and were less abundant than genomic microsatellites, but more likely to be functionally linked loci.

Conclusions: These observations provide deep insight into the evolution and distribution of microsatellites and how different isolation strategies affect microsatellite development using NGS.

Keywords: Transcriptome, Genome assembly, MiSeq, Amphibians, Population genetics, Next-generation sequencing

Background

Microsatellites, also known as simple sequence repeats (SSRs), are highly polymorphic and generally neutral molecular markers that are still extensively used as genetic markers in evolution and ecology [1–3]. However, the traditional protocols to produce SSRs using microsatellite-enriched libraries, cloning and Sanger sequencing [4] are time consuming and resource intensive. Especially for nonmodel species, microsatellites often have been developed de novo using this process, due to a low rate of cross-species microsatellite amplification success [5, 6].

To overcome these obstacles, next-generation sequencing (NGS) has achieved faster and higher throughput nucleotide sequence data, which facilitate the identification and characterization of microsatellites in nonmodel species. Indeed, hundreds of microsatellite loci were isolated without any screening steps using this high-throughput sequencing technology [7, 8]. Microsatellite loci have been isolated using many NGS platforms, such as Roche 454 [9–11], Ion Torrent PGM [12–14], Illumina [15, 16], and PacBio [17]. The popularity of using NGS not only varies by platforms but also by the sources of the sequences. Both genome skimming and transcriptome sequencing have been used to mine microsatellite markers [18–20].

* Correspondence: zengxm@cib.ac.cn

¹Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China

Full list of author information is available at the end of the article



Although the recent development of NGS techniques has facilitated the de novo development of SSRs, this task is still quite expensive, and the subsequent polymorphism screening requires a substantial amount of time [21, 22]. Obtaining good polymorphic loci usually requires further polymerase chain reaction (PCR) optimization. The mutational dynamics of microsatellites have diversely evolved in different species and in variant repeat types of SSRs [1, 23], and they are important in understanding how to quickly obtain SSRs with high successful amplification rate and non-monomorphic loci for microsatellite isolation. One strategy can efficiently differentiate polymorphic loci from monomorphic loci, by screening multiple genotypes and selecting only those loci that already show differences in repeat length within the set of sequence reads for marker development [21, 24, 25]. Yet, this strategy requires the pooling of multiple genotypes for NGS library construction. Actually, most studies do not make use of such mixed multiple genotypes for library layout [12, 15–17]. Instead, microsatellite data may already or soon be available as a byproduct of NGS. For such cases, we can also use the properties of microsatellites to determine the selection range. A well-known relationship exists between microsatellite length and polymorphism level, due to mutation rate increases with an increasing number of repeat units in an appropriate range [1, 26]. Microsatellites with longer repeat lengths are likely to be highly variable. Intuitively, this seems understandable; more repeat units generally show higher mutation rates because replication slippage may increase in proportion to the number of repeats [1, 27]. In addition, a negative correlation between microsatellite repeat length and substitution rate in nearby flanking sequence was found in mammalian genomes [28]. A low substitution rate in nearby flanking sequences would increase the rate of successful amplification. Getting a high number of repeats may help improve the polymorphism of microsatellites and their allelic richness, especially important for species with low genetic variability.

Researchers often tend to select the most polymorphic markers for subsequent studies and to discard limited polymorphic or monomorphic loci. However, empirical studies have shown that choosing the most polymorphic markers leads to an ascertainment bias and generally overestimates genetic diversity [26, 29]. Furthermore, researchers also tend to select tetranucleotide repeat markers, because they are easier to distinguish, as they cause fewer stutter bands than dinucleotides when screening the polymorphism [1, 23]. However, mutation rates of tetranucleotide loci are varied among species. Some studies reported that tetranucleotide loci have higher mutation rates [13, 30], but not supported in *Drosophila* [31, 32]. Hence, researchers should primarily take this information into account when selecting the appropriate microsatellite loci for ecological and evolutionary studies [29, 33].

The polymorphism of microsatellites is not only affected by the properties of microsatellite repeat units, but also related to the sources of omics data and the assembly of NGS [19, 34]. Genome skimming is by far one of the simplest methodologies for NGS, involving random sampling of a small percentage of total genomic DNA, which can obtain sequences containing plenty of microsatellites [35]. Transcriptome sequencing, or RNA-seq, is also used to isolate SSRs. Microsatellites present in protein coding regions, however, have low allele polymorphism and short repeat lengths, because SSRs could lead to gain or loss of gene function via frame shift mutations [36]. SSR expansions or contractions within genes can eventually lead to phenotypic changes. Yet, microsatellite size homoplasy, characterizing DNA fragments identical in state but not by descent due to convergent evolution, would be more likely occur in genomic SSRs, due to higher mutation rate of genomic SSRs compared to transcriptomic SSRs [37, 38]. Additionally, the assembler chosen for NGS would significantly affect the quality of genome or transcriptome reconstructions, which require special attention during downstream analysis [39, 40]. The microsatellites isolated from NGS could differ depending on sequencing method, read length, library layout, sequencing depth and genome composition. The genome size could also affect the frequency of microsatellites [23, 41]. However, such comprehensive evaluations are only fragmentarily reported in a limited number of species [11, 17]. It remains unclear the extent to which assembly strategy imposes a measurable effect on microsatellite marker development. A deeper understanding of the isolation of microsatellites from NGS is therefore needed, not only to understand evolutionary and mutational properties of microsatellites but also to appropriately use microsatellite markers in ecological and evolutionary studies.

Here, we chose six frog species for genome skimming in two runs of Illumina MiSeq and two frog species for transcriptome sequencing by using Illumina HiSeq to isolate microsatellites for these species. The efficiency of microsatellite development from different sequencing methods and genome composition was compared. Additionally, we investigated how the quantity or sequencing depth of NGS data, read length, and assembly strategy affect the number and polymorphism of the isolated microsatellites.

Methods

Six frog species were collected for genome skimming by Illumina MiSeq sequencing: *Amolops chunganensis*, *A. mantzorum*, *Hyla chinensis*, *Hylarana latouchii*, *Quasipaa boulengeri*, and *Rhacophorus chenfui*. A single individual for each species was chosen for genomic sequencing. A single individual from two species, *A. mantzorum* and *Q.*

boulengeri, were also used for transcriptome sequencing. All the samples were collected in the field of China, and the location and other information for each sample is shown in Additional file 1: Table S1. The frogs in this study were anesthetized and euthanized in 0.2% MS222 (ethyl-3-aminobenzoate methanesulfonate) to collect target tissues. Genomic DNA was extracted from muscle tissue using SDS-proteinase K/phenol-chloroform protocol [42]. RNA was extracted individually from six types of tissues (cerebrum, heart, liver, skeletal muscle, skin, and testicle) using TRIzol[®] reagent according to the manufacturer's protocol (Life Technologies, Shanghai, China) and mixed with approximately the same quantity. All work with animals was conducted according to relevant national and international guidelines on the Protection of Wildlife (Chengdu Institute of Biology issued permit number CIB#2014–36 and CIB#2014–110). All animal care and experimental procedures were approved by the Animal Care and Use Committee of Chengdu Institute of Biology, Chinese Academy of Sciences (Permit Number: CIB-20121220A). All voucher specimens were deposited in the Herpetological Museum of Chengdu Institute of Biology, Chinese Academy of Sciences.

Genomic DNA and RNA were shipped to Novogene Bioinformatics Technology (Beijing, China) for library construction and sequencing with Illumina MiSeq (PE300) and Illumina HiSeq 2000 (PE125), respectively. Briefly, 3–32 μ g of DNA or RNA for each sample is used for library preparation. The concentrations and volumes of DNA and RNA are shown in Additional file 1: Table S1. The paired-end library for genome skimming was prepared using the TruSeq DNA PCR-Free Sample Preparation Kit (Illumina, USA) (insert size approximately 500 bp), following the protocols for Illumina DNA sample preparation. For transcriptome sequencing, libraries were generated using TruSeq SR Cluster Kit v3-cBot-HS (Illumina, USA) following the manufacturer's recommendations. These libraries were pooled for sequencing on the Illumina platforms (MiSeq for genomic libraries, HiSeq 2000 for transcriptomic libraries) at Novogene Bioinformatics Technology (Beijing, China).

Contig assembly and microsatellite detection

The raw sequence reads were inspected with FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and adaptors were trimmed using Trimmomatic [43]. In this step, all raw reads with more than 10% missequenced nucleotides (poly-N) were discarded, and reads with more than 50% of the bases having a Q-value ≤ 20 were filtered out. De novo assembly of the cleaned reads for genome sequencing for each dataset was performed using SOAPdenovo2 [44] and Trinity [45]. Read preprocessing and assembly parameters followed the individual programs' guidelines. For SOAPdenovo2, we used 6 k-mer sizes between 55 and 80, with a step size of 5 for the manual

specification of assembly parameter values. For each assembly, we monitored the change in total number of contigs and N50 size over the assessed parameter range. For transcriptome sequencing, Trinity and SOAPdenovo2 (8 k-mer sizes between 25 and 60, with a step size of 5) were used for the assembly of the cleaned reads.

The identification of microsatellite loci from the assembly data was performed with QDD3 v3.1.2 [46]. We searched for di-, tri-, tetra-, penta- and hexanucleotide pure repeated units, with at least 7 tandem repeats for dinucleotides, and 6 repeats for the others. To eliminate redundancy, the similarity of sequences containing microsatellite regions was detected by all-against-all BLAST searching. Primer pairs were designed using Primer 3.0 software [47] to meet the following restrictions: the target amplification size range was 120–500 bp, the primer annealing temperature was restricted to 55–65 °C and the primer length was 20–27 bp. In order to avoid the possible bias from software, MISA [48] was also used to identify microsatellite markers, following the default parameter values.

Length and quantity simulation

We conducted a simulation experiment to investigate the effects that different read length and quantity distributions may have on the discovery of microsatellites and on their primer design. For the read length simulation, we selected the genome sequencing data from *Amolops chunganensis*, and chose five size distributions of reads (100, 125, 150, 175 and 200) for detection. The reduced dataset for each size were performed in ART [49]. The genomic datasets of *Amolops mantzorum* and *Quasipaa boulengeri* were used for data quantity simulation. Five datasets (full, 1/2, 1/4, 1/8, and 1/16 of the total reads) were selected randomly for the quantity simulation. Both the length and the quantity simulation datasets were assembled using SOAPdenovo2 and Trinity. Microsatellite discovery and primer design were performed using QDD3 v3.1.2 and Primers 3.0 software.

PCR amplification and genotyping

To test the polymorphism and stability of the detected SSRs, we chose *Amolops mantzorum* and *Quasipaa boulengeri* for PCR amplification. The sample information for each species is provided in Additional file 2: Table S2. We synthesized primers for the loci with a target product size of approximately 120–400 bp for amplification. A total of 100 loci for *A. mantzorum*, and 62 loci for *Q. boulengeri* were used for amplification. Eight samples for each species were used for preliminary testing of the primer amplification success rate. Then, electrophoresis was performed with an 8% nondenaturing polyacrylamide gel to evaluate polymorphism. The loci,

which could not be successfully amplified, and gave multiple bands, faint bands or bands of unexpected sizes, were discarded from subsequent analyses. Eventually, the forward primers of those loci that showed variation were labeled with one of the fluorescent dyes (FAM, HEX or TAMRA) for a further polymorphism test. For the polymorphism test, we used 64 individuals of *A. mantzorum* and 96 individuals of *Q. bouleengeri* to analyses. PCR amplifications were performed in a volume of 15 μ l containing 7.5 μ l 2 \times EasyTaq PCR SuperMix (TransGen Biotech, Beijing, China), 0.5 μ l each primer (10 pm/ μ L), 5.5 μ l ddH₂O and approximately 1 μ l diluted genomic DNA (50–200 ng/ μ l). Reaction conditions included an initial 5 min denaturing step at 95 °C, followed by 32 cycles of 95 °C for 30 s, 55–65 °C annealing for 30 s, and 72 °C extension for 30 s, with a final extension of 7 min at 72 °C. The primer sequences, specific annealing temperatures and other information for all the primers used, are shown in Additional file 3: Table S3. The PCR products were run on an ABI-3730xl sequencer and genotyped using GENEMAPPER 4.0 (Applied Biosystems).

The null alleles and genotyping errors for each locus were checked using Microchecker v2.2.3 [50]. The number of alleles, allelic size range, observed and expected heterozygosity were determined using GENEPOP v4.2 [51]. Tests for deviation from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were also calculated using GENEPOP v4.2.

Results

Sequence data from Illumina NGS

Genomic sequencing was performed using two runs of Illumina MiSeq, and each run included three frog species (Additional file 1: Table S1). Each run of Illumina MiSeq (PE300) had 22 million raw PE reads. The statistics for sequencing data from each species are shown in Table 1. Each species had almost 4 Gb raw data. For transcriptome sequencing, a total of 125.53 million raw PE reads were obtained from Illumina HiSeq (5.04 Gb and 7.32 Gb, for *Amolops mantzorum* and *Quasipaa bouleengeri*, respectively). The raw reads have been deposited in the NCBI SRA database (Bioproject accession number: PRJNA329172, PRJNA347498 and PRJNA374468). After removal of adaptors and low-quality reads, the clean reads were used for the subsequent assembly. The data utilization ratio for MiSeq (43.98–85.09%) was much smaller than that for HiSeq (96.27–97.37%) (Additional file 1: Table S1).

NGS assembly and microsatellite detection

The total number of contigs and the average length of contigs are shown in Fig. 1. For the SOAPdenovo2 assembly for each species, the smaller k-mer sizes produced more contigs in total, but these contigs had a shorter average length. All the k-mer sizes used for

Table 1 Overview of the sequencing data from Illumina next-generation sequencing for each species of frog

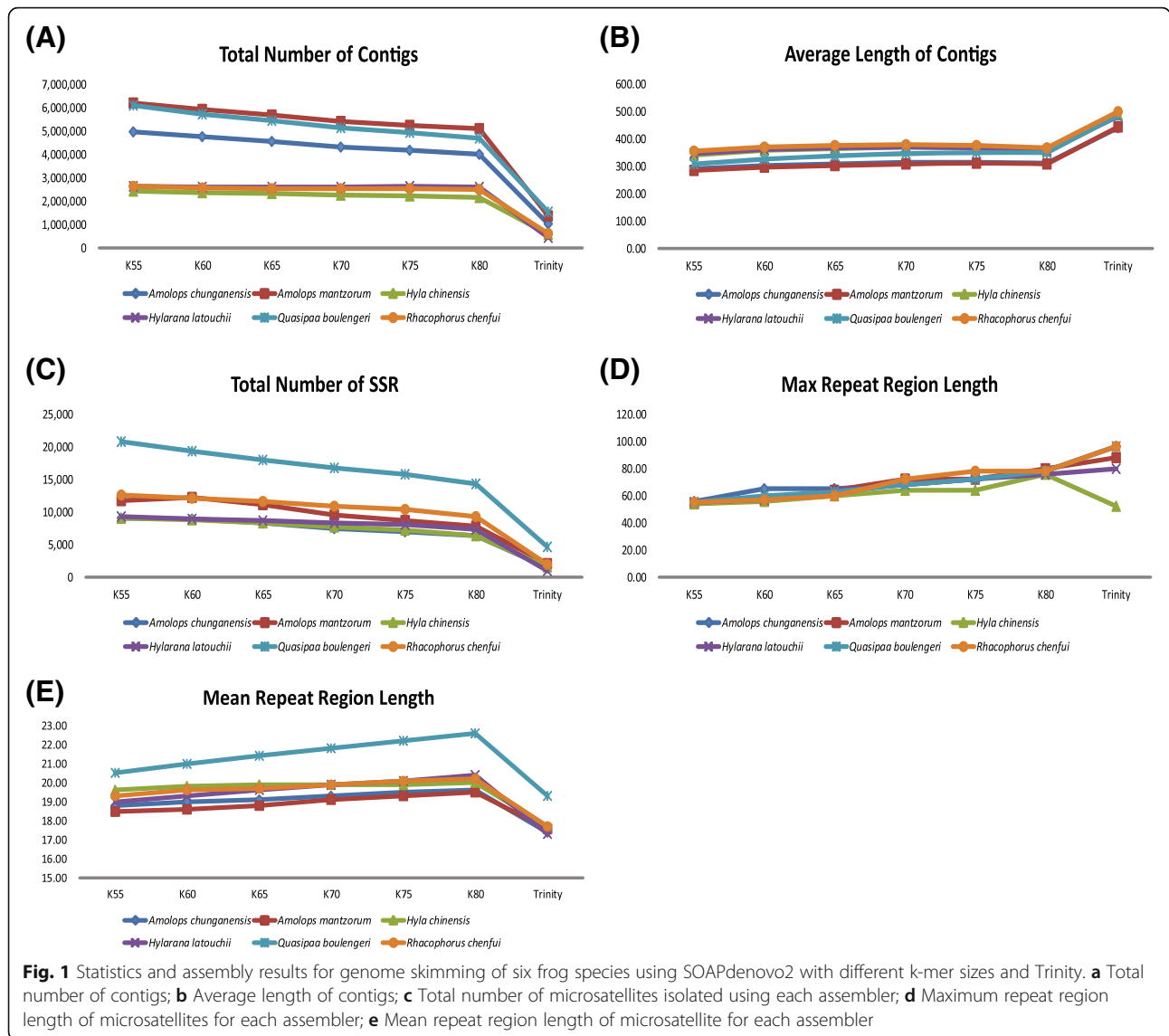
NGS Source	Species	Total sequence reads (Million)	Total No. of bases (Gigabase)
Genomic ^a	<i>Amolops chunganensis</i>	7.01	4.21
	<i>Amolops mantzorum</i>	10.19	6.11
	<i>Hyla chinensis</i>	6.93	4.16
	<i>Hylarana latouchii</i>	6.63	3.98
	<i>Quasipaa bouleengeri</i>	8.28	4.97
	<i>Rhacophorus chenfui</i>	7.28	4.37
Transcriptomic ^b	<i>Amolops mantzorum</i>	52.28	5.04
	<i>Quasipaa bouleengeri</i>	73.25	7.32

^a Genomic samples were sequenced using the Illumina MiSeq (PE300) system

^b Transcriptomic samples were sequenced using the Illumina HiSeq 2000 (PE125) system

SOAPdenovo2 produced over 2 million contigs for each species, but Trinity produced only a quarter as many contigs. Although fewer contigs were assembled from Trinity than SOAPdenovo2, the Trinity assembler collected contigs with significantly longer average lengths (Fig. 1a, b).

For microsatellite loci detection, each assembly strategy also revealed impressive differences based on QDD results. The smaller k-mer sizes led to the detection of more microsatellite loci with flanking sequences that could be used for primer design (Fig. 1c). Nevertheless, the smaller k-mer size led to the detection of microsatellite loci with lower maximum and mean repeat region length (RRL) (Fig. 1d, e). In contrast, Trinity identified the fewest microsatellite loci and the lowest mean RRL, but with the longest maximum RRL, except in one species (*Hyla chinensis*; Fig. 1). All the assembly strategies produced sufficient numbers of microsatellite loci for further analysis, even the lowest number produced by Trinity was greater than 1000. For the entire dataset, dinucleotide repeats predominate, followed by tri- and tetranucleotide repeats, and penta- and hexanucleotides repeats are least abundant. From the results of MISA, the number of SSRs decreased with increasing k-mer sizes, while the maximum RRL increased with increasing k-mer sizes (Additional file 4: Table S4). Take *Amolops chunganensis* for example, the largest number of SSRs is when K = 60 (196,837), and the smallest number of SSRs is when K = 80 (179885). For the maximum RRL, the smallest is 62 (K = 60), and the largest is 84 (K = 80). The tendency was not affected by the adopted software. Notably, the numbers of identified SSRs were different between the two softwares, since MISA includes all the



number of identified SSRs, while QDD only counts the number of identified SSRs with primers.

For RNA-Seq, Trinity is better than SOAPdenovo to obtain average length of contigs and total number of microsatellite loci at any k-mer size (Additional file 5: Table S5). But the maximum and mean RRL in SOAPdenovo were larger than those in Trinity. A total of 121,416 and 278,803 contigs were obtained from Trinity program for *Amolops mantzorum* and *Quasipaa boulengeri*, respectively (Table 2). The total number of contigs for RNA-Seq was much smaller than that obtained for genomic sequencing, but the average length of contigs was longer with RNA-Seq for both species. Both the total number of microsatellite loci and the maximum RRL in RNA-Seq were obviously lower than that obtained from genomic sequencing ($p < 0.01$). Dinucleotide repeats were

Table 2 Assembly and microsatellite loci detection statistics for transcriptome sequencing using Trinity

	<i>Amolops mantzorum</i>	<i>Quasipaa boulengeri</i>
Total number of contigs	121,416	278,803
Average length of contigs	958	919
N50	1953	2080
Total number of SSRs	958	1554
Max RRL ^a	25	32
Mean RRL	16	15.7
Dinucleotide (%)	716 (74.74)	1174 (75.55)
Trinucleotide (%)	226 (23.59)	364 (23.42)
Tetranucleotide (%)	15 (1.57)	14 (0.91)
Pentanucleotide (%)	1 (0.10)	1 (0.06)
Hexanucleotide (%)	0	1 (0.06)

^a RRL repeat region length, is the length of repeat units

also the most abundant category, followed by tri- and tetranucleotide repeats, similar to the results from genome sequencing.

Data quantity and microsatellite discovery

Amolops mantzorum and *Quasipaa boulengeri* were chosen for data quantity simulation. The statistics of SSR discovery for each dataset are shown in Table 3 and Fig. 2. The total number of identified microsatellite loci was reduced by the decline in the quantity of data. As expected, the number of microsatellite loci was associated with the size of dataset. For instance, a twofold increase in data quantity, such as from half to full in *Amolops mantzorum*, resulted in a nearly twofold increase in the number of microsatellite loci detected. Furthermore, the number of SSRs for each type of nucleotide repeat also followed this tendency. In other words, the smaller the quantity of NGS data, the fewer the microsatellite loci identified. However, the results showed very little difference in maximum and mean RRL for each dataset (Table 3). The relation of the k-mer sizes to the data quantity reflected that the total number of contigs and microsatellites decreased with increasing k-mer sizes (Fig. 2a-d). There are overlaps of contigs and microsatellites when different k-mers with varied data quantity.

Read length simulation

To test how the length of the NGS data affects the final quality and the number of SSR loci, we selected the genome sequencing data from *Amolops chunganensis* for length simulation. The results showed a positive relationship between read length and effectiveness of microsatellite isolation (Fig. 3). The shorter the read length, the fewer microsatellite loci were identified. Interestingly, the number of microsatellite loci identified was

nearly equal to the read lengths of 125 and 150, implying that for short sequence read assembly, when read length was not increased dramatically, the read length could not directly influence the number of microsatellite loci isolated. The proportion of each type of nucleotide repeat was not affected by read length (Fig. 3b). With increasing k-mer sizes, the total number of contigs and microsatellites would decrease (Fig. 2 e, f).

Microsatellite development and polymorphism

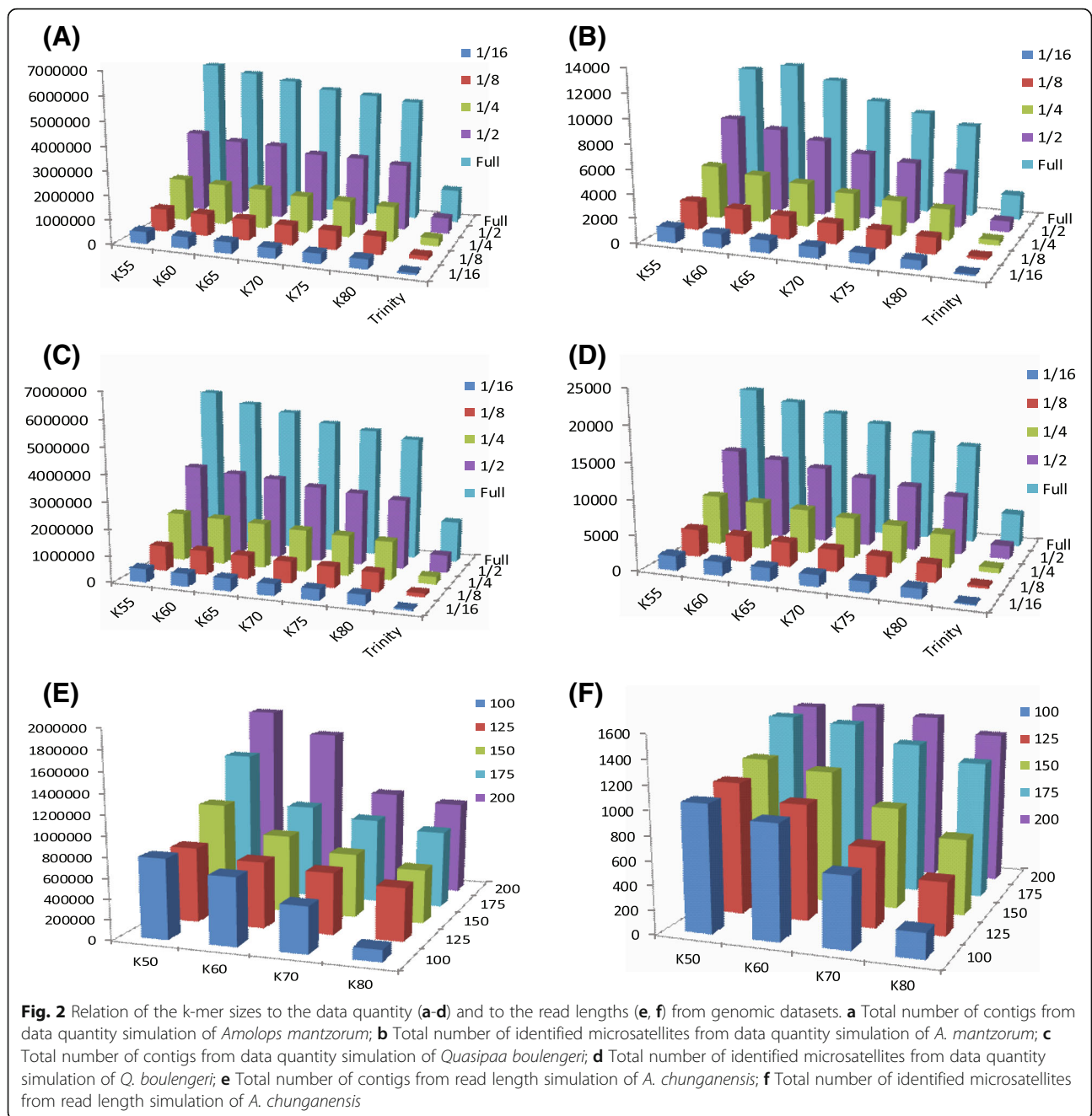
The characteristics of these newly discovered microsatellites of *Amolops mantzorum* and *Quasipaa boulengeri* are summarized in Additional file 3: Table S3. For *A. mantzorum*, 30 loci could be easily genotyped and displayed polymorphism in 64 individuals from western Sichuan province, China (Additional file 2: Table S2). This population is possibly in Hardy Weinberg equilibrium with loci MT88, MT10, MT84, MT39, MT1, showing signs of a null allele. The number of alleles per locus ranged from 2 to 21 with an average of 9. The observed and expected heterozygosity per locus ranged from 0.078 to 0.984 and from 0.076 to 0.918, respectively (Additional file 3: Table S3). For *Q. boulengeri*, 29 loci showed a clear band of the expected size and displayed polymorphisms in 96 individuals from two populations (Additional file 2: Table S2). The loci QB02, QB18, QB34, QB16, showed signs of a null allele. The number of alleles ranged from 2 to 23, with an average of 7.23. The observed and expected heterozygosity ranged from 0.300 to 0.922 and from 0.237 to 0.937, respectively. A few markers in both species were found to deviate from Hardy-Weinberg equilibrium, suggesting an association with the presence of null alleles (Additional file 2: Table S2).

Combining the genomic library microsatellites and transcriptome derived microsatellites from previous studies for

Table 3 Effects of data quantity on microsatellite discovery based on genomic data (k-mer size = 65)

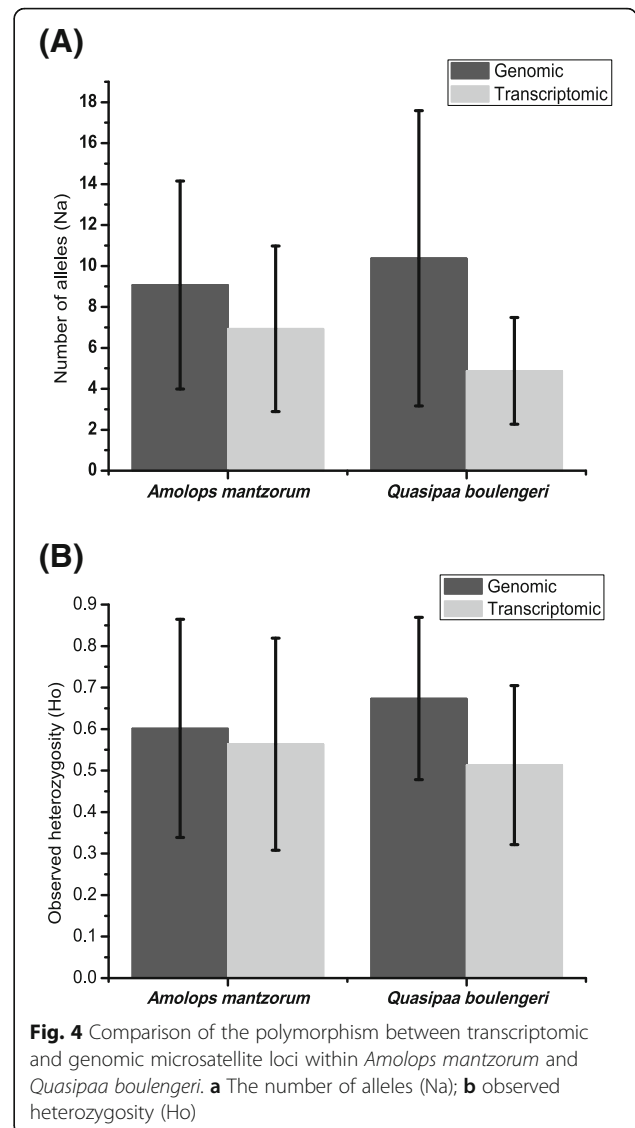
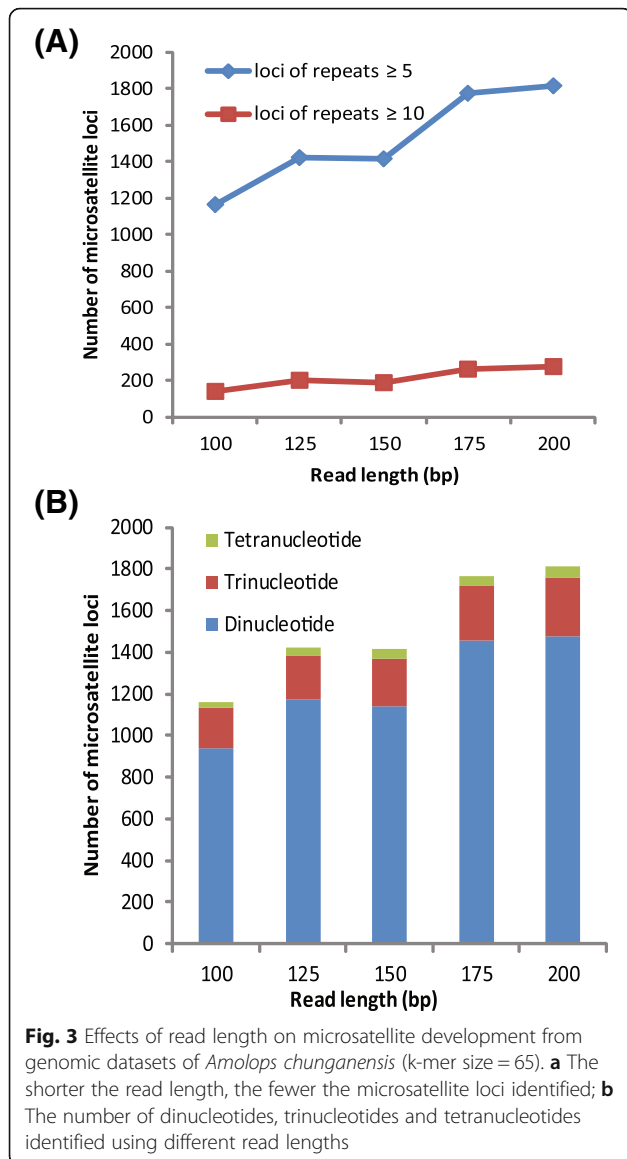
	<i>Amolops mantzorum</i>					<i>Quasipaa boulengeri</i>				
	Full	1/2	1/4	1/8	1/16	Full	1/2	1/4	1/8	1/16
Data in megabases (Mb)	5200	2600	1300	650	327	4700	2400	1200	580	294
Microsatellite loci	11,116	6511	3665	1929	1043	17,942	10,907	6317	3448	1883
Max RRL ^a (bp)	64	64	60	60	60	66	66	65	65	63
Loci with RRL > 40 bp	241	138	74	36	23	940	561	345	196	118
Mean RRL	18.8	18.8	18.7	18.7	18.6	21.4	21.3	21.3	21.3	21.2
Dinucleotides	8794	5171	2871	1517	836	11,527	7123	4215	2272	1289
Trinucleotides	1640	943	550	300	145	4891	2832	1575	889	447
Tetranucleotides	609	358	219	98	55	1361	854	460	266	140
Pentanucleotides	55	29	18	11	5	127	76	54	18	6
Hexannucleotides	18	10	7	3	2	36	22	13	3	1

^a RRL repeat region length, is the length of repeat units



Amolops mantzorum and *Quasipaa boulengeri* [52–54], we compared the properties of microsatellites from different sources. In total, there were 30 genomic and 15 transcriptomic loci for *A. mantzorum*; and 40 genomic and 32 transcriptomic loci for *Q. boulengeri* (Additional file 3: Table S3). Both the number of alleles (N_a) and the observed heterozygosity (H_o) were affected by the sources of omics data (Fig. 4). The microsatellites derived from transcriptome have significantly lower N_a and H_o values than those from genomic loci in *Quasipaa boulengeri* (independent-samples T test, $p < 0.01$), but not significantly lower in *Amolops*

mantzorum ($p > 0.05$). In addition, transcriptomic microsatellites possessed fewer repeats and a narrower size distribution than genomic loci (Fig. 5; Additional file 3: Table S3). The number of repeats positively influenced the N_a for genomic data (for *A. mantzorum*, Pearson’s correlation coefficient $r = 0.61$, two-tailed $p < 0.01$; for *Q. boulengeri*, Pearson’s correlation coefficient $r = 0.481$, two-tailed $p < 0.01$). However, the number of repeats and N_a were not significantly correlated in transcriptomic SSRs (for *A. mantzorum*, Pearson’s correlation coefficient $r = 0.329$, $p = 0.231$; for *Q.*

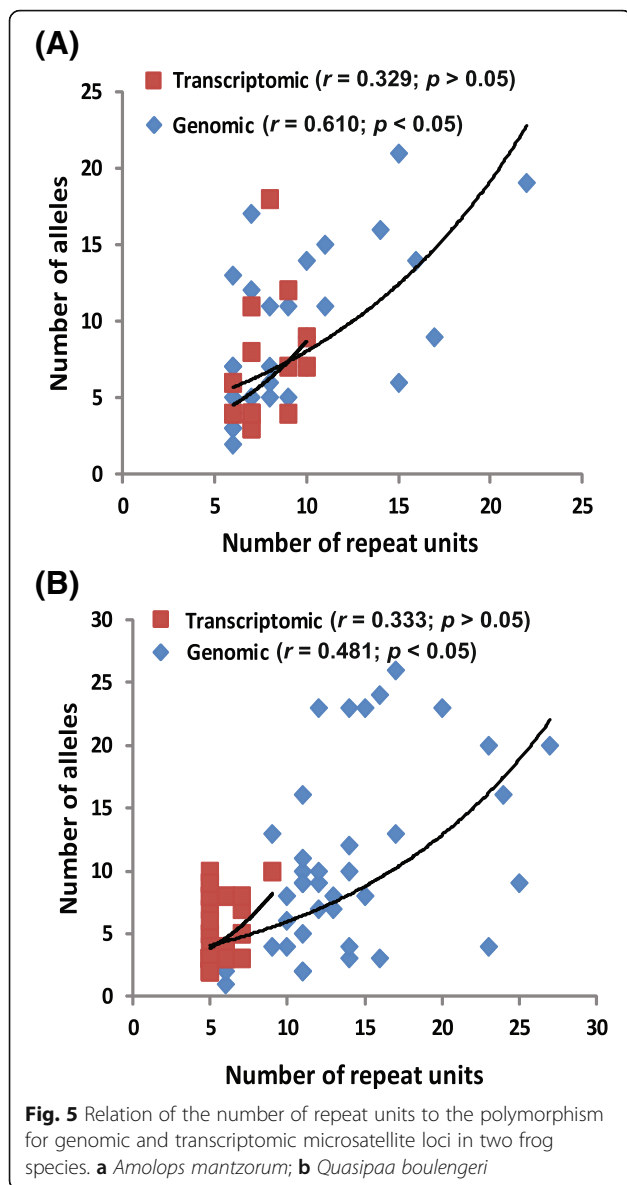


boulengeri, Pearson's correlation coefficient $r = 0.333$, two-tailed $p = 0.06$).

Discussion

In the era of NGS, SSRs are still widely used in ecology and evolution [2], because they have their own advantages for population genetics compared to other molecular markers (e.g., single nucleotide polymorphisms, SNPs) [55, 56]. For example, SSRs can be successfully amplified from low quality of DNA (e.g. ancient DNA or museum specimens); the success rate of cross-amplification for SSRs in closely related species is typically higher than that for SNPs [55, 57]. Microsatellites have remained the markers of choice for various population and conservation genetics studies for parentage analysis and for haplotyping in non-model

organisms [58, 59]. Specifically, they can provide new insights into genome and chromosome evolution [60, 61]. Their applications in genetics are extensive due to their ceaseless mutation, widespread length variations, co-dominance and reproducibility [1, 23]. Using conventional methods, obtaining nucleotide sequence data for microsatellites and their flanking regions in a target species requires cost and time, particularly for screening, cloning and sequencing microsatellite regions. Utilization of NGS to identify microsatellites is becoming more accessible and feasible, but it still lacks a comprehensive evaluation of how read length, library layout, sequencing depth and assembly strategy influence microsatellite isolation. In this study, we find large variations in the quality of results depending on the different datasets and assemblers used.



Data quantity, read length, assembly and SSR abundance
 Microsatellites are present in the genome of most taxa and are abundant in the genome. Although the abundance of microsatellites differs between taxa and even closely related species, millions of microsatellite loci have been detected in human and mouse genomes [1]. The results from our data quantity simulation revealed that the number of isolated microsatellites increases following an increase in the data quantity when the data quantity is smaller than the entire genome (Table 3, Fig. 2), which can be attributed to the abundance of microsatellites in frog species. In addition, we have isolated these microsatellites randomly or almost randomly from the genome, because there are no dramatic differences in nucleotide repeats and maximum and mean RRL between each dataset. At least two explanations might account for the

observation of such a distribution. Firstly, microsatellites may be randomly distributed over the euchromatic genome [62]. However, empirical studies using genome sequencing and fluorescence in situ hybridization (FISH) have shown uneven and nonrandom SSR distribution within and between genomes, where microsatellite density is higher near the centromeres and telomeres [60, 63]. Secondly, the random construction of library layout with genome skimming also resulted in such random isolation. Genome skimming is random sampling of a small percentage of total genomic DNA, which can randomly accept sequences containing microsatellites [35].

The yield of microsatellites through genome skimming could be affected not only by sequencing depth, but also by genome size (C-value). Generally, the frequency of microsatellites correlates positively with genome size [23, 41]. For these six frogs, unfortunately, the genome sizes are unclear, except *Quasipaa boulengeri* [64]. The C-value of *Q. boulengeri* is 3.06–3.11 [65, 66]. The coverage for *Q. boulengeri* is approximately 1.6X (4.97 Gigabase). For other species, the coverage could be estimated to 1–2X according to the C-value of the closely related species for these frogs (C-value: 2–7). Due to low coverage sequencing for genome skimming, the data quantity simulation shows the yield of microsatellites for each species increased with increasing sequencing depth (Table 3). However, the highest yield of microsatellites among these species was *Q. boulengeri*, which didn't have the largest sequencing depth (Fig. 1; Table 1). The number of isolated microsatellites don't increase following an increase in the data quantity between species. The genome size could affect the frequency of microsatellites. Due to lack of C-value of these species, this assumption need to be further verified. Notably, genome size has a negative effect on amplification success rate [67]. Species with a large genome size, such as salamanders, have a low amplification success rate for microsatellite development [68], due to the relative amounts of target DNA and of available primers for amplifying target DNA could be reduced.

Read length is considered to be one of the most important factors for microsatellite development using NGS [12, 17]. The shorter read lengths can influence SSR development by: 1) reducing the likelihood of discovering longer repeat length microsatellites and 2) reducing the potential success of designing primers [12, 17]. The read length for Illumina is generally shorter than that for the Ion Torrent and the GS-FLX (454) platforms. MiSeq has the longest read lengths (2 × 300 bp) of the Illumina platforms, but its output consists of relatively few reads. HiSeq has shorter reads than MiSeq; the new HiSeq v2 reagents allow 2 × 250 bp in rapid run mode [2]. Despite the shortcoming of read length using Illumina platforms,

microsatellite isolation can largely overcome this shortcoming with upgrades in the assembly strategy. After short read assembly, many microsatellite markers have been developed (Fig. 1, Additional file 3: Table S3). Our results (Figs. 2, 3) are in accordance with prior evidence that the shorter the read length, the fewer the microsatellite loci identified [17]. This indicates that microsatellite development from short sequence read assembly would also be influenced by both the length of the flanking regions and the repeat length of microsatellites.

Assembly strategy can also influence the processes involved in microsatellite development. Generally, de Bruijn graph (DBG) based assemblers, such as SOAPdenovo and Trinity, are suited for short read lengths and high-coverage data with less run time and memory complexity [69]. However, the choice of an appropriate k-mer size is one of the major difficulties in using DBG assemblers [39]. Our results show that the quality of both genome reconstructions and microsatellite development depends essentially on k parameter values (Figs. 1, 2). The difference of the yields of microsatellites can be attributed to the low coverage of genome skimming and the properties of DBG. When the coverage is low, using a short k-mer for DBG assemblers can be critical for successful assembly, but has a limited ability to resolve repeat sequences, such as microsatellite. In contrast, a longer k-mer size can better resolve repeats but may result in fragmented assembly due to insufficient coverage [70]. Therefore, a longer k-mer size led to the detection of fewer total microsatellite loci and longer repeat length, which have a higher polymorphism level for downstream analysis [26, 27]. For Trinity, because of eliminating both low-complexity and singleton k-mers as initial seeds for contig extensions in Inchworm's step, it causes the low number of contigs and microsatellites [71]. The longer repeat lengths are not an error of assembly, because they have the same amplification success as shorter repeats (Additional file 2: Table S2, Additional file 3: S3). For transcriptome sequencing, Trinity is better than SOAPdenovo at any k-mer size with regard to average length of contigs and total number of microsatellite loci (Additional file 5: Table S5). It accounts for the high coverage of transcriptome sequencing and algorithmic advantages of Trinity. This observation is consistent with empirical studies that Trinity is suited for de novo transcriptome assembly [72].

Microsatellite polymorphism

NGS may provide lots of potential microsatellites, but selecting appropriate polymorphic microsatellites is still a critical issue for downstream analysis. Initial polymorphism detection can be performed by pooling multiple individuals

when constructing sequencing libraries to investigate whether candidate loci are polymorphic [21, 73]. However, such prior mixing of samples for library construction has not been used in many NGS studies [12, 15–17]. Whatever the methods used, finding polymorphic microsatellite loci is the critical issue. Fortunately, these microsatellites provide repeat type and length, which is indirect information to infer polymorphism. The mutation process seems to be heterogeneous with respect to loci, repeat types and organisms, but it is widely accepted that microsatellites with a greater number of repeats are more mutable [1, 23, 74]. Alleles with a higher number of repeats often mutate at a higher rate, and the relationship between length and rate has been reported to be exponential rather than linear [23]. The results from genomic SSRs revealed that the number of repeats or repeat lengths positively influenced the Na ($p < 0.01$) (Fig. 4), supporting this hypothesis. But these correlations are very weakly for transcriptomic SSRs ($p > 0.05$), and it could be attributed to the functional constraint on transcriptomic SSRs [36]. In addition, those microsatellite loci with long repeat lengths have another advantage in that their nearby flanking sequences have a low substitution rate, which would increase the rate of successful amplification [28]. Of course, a higher mutation rate SSRs with longer repeat length is not necessarily better. Due to ascertainment bias and overestimation of genetic diversity caused by the most polymorphic markers [26, 29], the random selection of SSRs should depend upon the subsequent downstream analyses.

Transcriptome versus genome microsatellite markers

Both transcriptomic and genomic sequencing identified many potential microsatellites, but the number and properties of these microsatellites were quite different. Transcriptomic SSRs are expected to display lower levels of polymorphism than genomic SSRs, as they are associated with conserved regions of the genome [19, 75]. In our study, the polymorphic screening in *Quasipaa boulengeri* showed that both Na and Ho were lower in transcriptomic than genomic microsatellites (Figs. 4, 5). This low polymorphism could be explained by transcriptomic SSRs are often non-neutral loci, which might bias the estimates of population divergence under the assumption of a neutral model of drift, mutation and migration [19, 34]. In addition, a few study found that the level of SSR variability is likely to differ between coding and untranslated regions [48, 76]. That could be the reason why microsatellites derived from transcriptome were different between *Quasipaa boulengeri* and *Amolops mantzorum*. But it needs to be further investigated. The expansions or contractions of microsatellites within genes can eventually lead to phenotypic changes or loss of function [36]. Though low polymorphism for SSRs from transcriptomic sequencing, some studies

would still chose this method to develop functionally linked loci. Because of the functional constraint on transcriptome SSRs, it is more easily to find functional linked loci from transcriptome than genome. The transcriptomic sequence is more conserved than noncoding genome sequence, and more likely to Blast on its sibling species with reference sequences of genome or transcriptome. It is very useful for non-model species to study loci location and function. For *Quasipaa boulengeri*, we employed both genomic and transcriptomic SSRs to isolate sex-linked microsatellite markers. And one locus from transcriptome sequencing data was found [77]. Overall, these results forge a framework for our deep understanding of the evolution and distribution of microsatellites and how different isolation strategies affect microsatellite development using NGS.

Conclusions

In the era of NGS, microsatellite data may already or soon be available as a byproduct of NGS, and it is still widely used in ecology and evolution. By using genome skimming and transcriptomic sequencing of non-model frog species, we investigated how the assembly strategy, read length, sequencing depth, and library layout affect the number and polymorphism of the isolated microsatellites. These results forge a framework for our deep understanding of the evolution and isolation strategies of microsatellites: 1) the number of isolated microsatellites increases with increased data quantity and read length; 2) assembly strategy could also influence the processes involved in microsatellite development; 3) the transcriptomic microsatellites displayed lower levels of polymorphisms and were less abundant than genomic microsatellites; 4) the proportion of each type of nucleotide repeats was not affected.

Additional files

Additional file 1: Table S1. Summary of sample information for next-generation sequencing and microsatellite discovery. (XLSX 12 kb)

Additional file 2: Table S2. Microsatellite screening data of *Amolops mantzorum* and *Quasipaa boulengeri* for each sample. (XLS 94 kb)

Additional file 3: Table S3. Characteristics of the microsatellite markers for *Amolops mantzorum* and *Quasipaa boulengeri*. (XLSX 38 kb)

Additional file 4: Table S4. Number of identified SSRs by MISA from genomic data of *Amolops chunganensis* and *Amolops mantzorum*. (XLSX 12 kb)

Additional file 5: Table S5. Statistics and assembly results for transcriptome sequencing of *Amolops mantzorum* and *Quasipaa boulengeri* using SOAPdenovo2 with different k-mer sizes and Trinity. (XLSX 12 kb)

Abbreviations

DBG: de Bruijn graph; FISH: Fluorescence in situ hybridization; Ho: observed heterozygosity; Na: Number of alleles; NGS: Next-generation sequencing; PCR: Polymerase chain reaction; RRL: Repeat region length; SNPs: Single nucleotide polymorphisms; SSRs: Simple sequence repeats

Acknowledgments

We thank Han Yang, Weizhao Yang and Xiang Tao for the NGS bioinformatic analysis. We gratefully acknowledge technical support from the laboratory of X.Z.

Funding

This research was supported by the National Key Programme of Research and Development, Ministry of Science and Technology (2017YFC0505202), and National Natural Science Foundation of China (NSFC-31401960, 31572241, 31572243), and the China Postdoctoral Science Foundation (2014 M552386).

Availability of data and materials

Genome skimming sequence data deposited in the NCBI SRA: Accession no. SRR3929656, SRR5248586, SRR5248585, SRR5248584, SRR3929655, SRR5248583; Transcriptome sequencing data, NCBI SRA Accession no. SRR5248938, SRR5248937.

Author's contributions

YX and XZ conceived and designed the study. YX and YZ performed the NGS bioinformatic analysis. SY and WL performed the laboratory experiments, microsatellite genotyping and population genetic analyses. SY, WL and YZ collected samples. YX, YZ and XZ wrote the manuscript. All authors reviewed the manuscript and approved the final version.

Ethics approval and consent to participate

All work with animals was conducted according to relevant national and international guidelines on the Protection of Wildlife (Chengdu Institute of Biology issued permit number CIB#2014-36 and CIB#2014-110). All animal care and experimental procedures were approved by the Animal Care and Use Committee of Chengdu Institute of Biology, Chinese Academy of Sciences (Permit Number: CIB-20121220A). All voucher specimens were deposited in the Herpetological Museum of Chengdu Institute of Biology, Chinese Academy of Sciences.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China. ²University of Chinese Academy of Sciences, Beijing 100049, China. ³College of Bioengineering, Sichuan University of Science & Engineering, Zigong 643000, China.

Received: 10 November 2017 Accepted: 28 November 2018

Published online: 07 December 2018

References

- Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet.* 2004;5(6):435–45.
- Hodel RGJ, Segovia-Salcedo MC, Landis JB, Crowl AA, Sun M, Liu X, Gitzendanner MA, Douglas NA, Germain-Aubrey CC, Chen S, et al. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *Appl Plant Sci.* 2016;4(6):1600025.
- Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett.* 2006;9(5):615–29.
- Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation: a review. *Mol Ecol.* 2002; 11(1):1–16.
- Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C. Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Mol Ecol.* 2007;16(18):3759–67.
- Primmer CR, Merilä J. A low rate of cross-species microsatellite amplification success in Ranid frogs. *Conserv Genet.* 2002;3(4):445–9.

7. Abdelkrim J, Robertson BC, Stanton J-AL, Gemmel NJ. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*. 2009;46(3):185–92.
8. Csencsics D, Brodbeck S, Holderegger R. Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. *J Hered*. 2010; 101(6):789–93.
9. Fernandez-Silva I, Whitney J, Wainwright B, Andrews KR, Ylitalo-Ward H, Bowen BW, Toonen RJ, Goetze E, Karl SA. Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. *PLoS One*. 2013;8(2):e55990.
10. Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ. Rise of the machines – recommendations for ecologists when using next generation sequencing for microsatellite development. *Mol Ecol Resour*. 2011;11(6):1093–101.
11. Schoebel C, Brodbeck S, Buehler D, Cornejo C, Gajurel J, Hartikainen H, Keller D, Leys M, Řičánová Š, Segelbacher G. Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. *J Evol Biol*. 2013;26(3):600–11.
12. Elliott CP, Enright NJ, Allcock RJN, Gardner MG, Megléc E, Anthony J, Krauss SL. Microsatellite markers from the ion torrent: a multi-species contrast to 454 shotgun sequencing. *Mol Ecol Resour*. 2014;14(3):554–68.
13. Igawa T, Nozawa M, Nagaoka M, Komaki S, Oumi S, Fujii T, Sumida M. Microsatellite marker development by multiplex ion torrent PGM sequencing: a case study of the endangered *Odorrana narina* complex of frogs. *J Hered*. 2015;106(1):131–7.
14. Sultana N, Igawa T, Nozawa M, Islam MM, Hasan M, Alam MS, Khan MMR, Sumida M. Development and characterization of 27 new microsatellite markers for the Indian bullfrog *Hoplobatrachus tigerinus* and its congeneric species. *Genes Genet Syst*. 2014;89(3):137–41.
15. Castoe TA, Poole AW, de Koning APJ, Jones KL, Tomback DF, Oylar-McCance SJ, Fike JA, Lance SL, Streicher JW, Smith EN, et al. Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS One*. 2012;7(2):e30953.
16. Lance SL, Love CN, Nunziata SO, O'Bryhim JR, Scott DE, Flynn RW, Jones KL. 32 species validation of a new Illumina paired-end approach for the development of microsatellites. *PLoS One*. 2013;8(11):e81853.
17. Wei N, Bemmels JB, Dick CW. The effects of read length, quality and quantity on microsatellite discovery and primer development: from Illumina to PacBio. *Mol Ecol Resour*. 2014;14(5):953–65.
18. Drechsler A, Geller D, Freund K, Schmeller DS, Künzel S, Rupp O, Loyau A, Denoël M, Valbuena-Ureña E, Steinfartz S. What remains from a 454 run: estimation of success rates of microsatellite loci development in selected new species (*Calotriton asper*, *Lisotriton helveticus*, and *Triturus cristatus*) and comparison with Illumina-based approaches. *Ecol Evol*. 2013;3(11): 3947–57.
19. Dufresnes C, Brelsford A, Béziers P, Perrin N. Stronger transferability but lower variability in transcriptomic- than in anonymous microsatellites: evidence from Hylid frogs. *Mol Ecol Resour*. 2014;14(4):716–25.
20. Rico C, Normandeau E, Dion-Côté A-M, Rico MI, Côté G, Bernatchez L. Combining next-generation sequencing and online databases for microsatellite development in non-model organisms. *Sci Rep*. 2013;3:3376.
21. Vukosavljev M, Esselink GD, van 't Westende WPC, Cox P, Visser RGF, Arens P, Smulders MJM. Efficient development of highly polymorphic microsatellite markers based on polymorphic repeats in transcriptome sequences of multiple individuals. *Mol Ecol Resour*. 2015;15(1):17–27.
22. Lopez L, Barreiro R, Fischer M, Koch M. Mining microsatellite markers from public expressed sequence tags databases for the study of threatened plants. *BMC Genomics*. 2015;16(1):781.
23. Bhargava A, Fuentes FF. Mutational dynamics of microsatellites. *Mol Biotechnol*. 2010;44(3):250–66.
24. Cardoso SD, Gonçalves D, Robalo JI, Almada VC, Canário AVM, Oliveira RF. Efficient isolation of polymorphic microsatellites from high-throughput sequence data based on number of repeats. *Mar Genom* 2013; 11(0):11–16.
25. Eschbach E, Schöning S. Identification of high-resolution microsatellites without a priori knowledge of genotypes using a simple scoring approach. *Methods Ecol Evol*. 2013;4(11):1076–82.
26. Brandström M, Ellegren H. Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. *Genome Res*. 2008;18(6):881–7.
27. Petit RJ, Deguilloux MF, Chat J, Grivet D, Garnier-Géré P, Vendramin GG. Standardizing for microsatellite length in comparisons of genetic diversity. *Mol Ecol*. 2005;14(3):885–90.
28. Santibáñez-Koref MF, Gangeswaran R, Hancock JM. A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. *Mol Biol Evol*. 2001;18(11):2119–23.
29. Queirós J, Godinho R, Lopes S, Gortazar C, de la Fuente J, Alves PC. Effect of microsatellite selection on individual and population genetic inferences: an empirical study using cross-specific and species-specific amplifications. *Mol Ecol Resour*. 2015;15(4):747–60.
30. Weber JL, Wong C. Mutation of human short tandem repeats. *Hum Mol Genet*. 1993;2(8):1123–8.
31. Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc Natl Acad Sci U S A*. 1997;94(3):1041–6.
32. Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol Biol Evol*. 1998;15(12):1751–60.
33. Cao L-J, Li Z-M, Wang Z-H, Zhu L, Gong Y-J, Chen M, Wei S-J. Bulk development and stringent selection of microsatellite markers in the western flower thrips *Frankliniella occidentalis*. *Sci Rep*. 2016;6:26512.
34. Postolache D, Leonarduzzi C, Piotti A, Spanu I, Roig A, Fady B, Roschanski A, Liepelt S, Vendramin GG. Transcriptome versus genomic microsatellite markers: highly informative multiplexes for genotyping *Abies alba* mill. And congeneric species. *Plant Mol Biol Rep*. 2014;32(3): 750–60.
35. Dodswoth S. Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci*. 2015;20(9):525–7.
36. Li Y-C, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and Evolution *Mol Biol Evol*. 2004;21(6):991–1007.
37. Estoup A, Jarne P, Cornuet JM. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol*. 2002;11(9):1591–604.
38. Khimoun A, Ollivier A, Faivre B, Garnier S. Level of genetic differentiation affects relative performances of expressed sequence tag and genomic SSRs. *Mol Ecol Resour*. 2017;17(5):893–903.
39. Greshake B, Zehr S, Dal Grande F, Meiser A, Schmitt I, Ebersberger I. Potential and pitfalls of eukaryotic metagenome skimming: a test case for lichens. *Mol Ecol Resour*. 2016;16(2):511–23.
40. Hunt M, Newbold C, Berriman M, Otto TD. A comprehensive evaluation of assembly scaffolding tools. *Genome Biol*. 2014;15(3):42.
41. Hancock JM. Genome size and the accumulation of simple sequence repeats: implications of new data from genome sequencing projects. *Genetica*. 2002;115(1):93–103.
42. Sambrook J, Russell DW. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Press; 2001.
43. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
44. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience*. 2012;1(1):18.
45. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M. *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494–512.
46. Megléc E, Pech N, Gilles A, Dubut V, Hingamp P, Trilles A, Grenier R, Martin JF. QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. *Mol Ecol Resour*. 2014;14(6):1302–13.
47. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. 2000;132(3):365–86.
48. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*. 2003;106(3):411–22.
49. Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read simulator. *Bioinformatics*. 2012;28(4):593–4.
50. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes*. 2004;4(3):535–8.
51. Rousset F. genepop'007: a complete re-implementation of the genepop software for windows and Linux. *Mol Ecol Resour*. 2008;8(1):103–6.
52. Luo W, Xia Y, Wang Q, Zeng X. Isolation and characterization of fifteen microsatellite markers for Sichuan torrent frog *Amolops mantzorum* (Anura: Ranidae). *Conserv Genet Resour*. 2015;7(2):489–91.

53. Xia Y, Hu LJ, Shan X, Zheng YC, Zeng XM. Isolation and characterization of eleven polymorphic tetranucleotide microsatellite loci for *Quasipaa boulengeri* (Anura: Dicroglossidae). *Conserv Genet Resour.* 2013;5(1):5–7.
54. Yuan S, Xia Y, Zheng Y, Zeng X. Development of microsatellite markers for the spiny-bellied frog *Quasipaa boulengeri* (Anura: Dicroglossidae) through transcriptome sequencing. *Conserv Genet Resour.* 2015;7(1):229–31.
55. Guichoux E, Lagache L, Wagner S, Chaumeil P, LÉGer P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, et al. Current trends in microsatellite genotyping. *Mol Ecol Resour.* 2011;11(4):591–611.
56. Ljungqvist M, Åkesson M, Hansson B. Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli et al. (2008). *Mol Ecol.* 2010;19(5):851–5.
57. Sharma PC, Grover A, Kahl G. Mining microsatellites in eukaryotic genomes. *Trends Biotechnol.* 2007;25(11):490–8.
58. Huang J, Li YZ, Du LM, Yang B, Shen FJ, Zhang HM, Zhang ZH, Zhang XY, 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(1):61.
59. Jan C, Fumagalli L. Polymorphic DNA microsatellite markers for forensic individual identification and parentage analyses of seven threatened species of parrots (family Psittacidae). *PeerJ.* 2016;4:e2416.
60. Ruiz-Ruano F, Cuadrado Á, Montiel E, Camacho J, López-León M. Next generation sequencing and FISH reveal uneven and nonrandom microsatellite distribution in two grasshopper genomes. *Chromosoma.* 2015; 124(2):221–34.
61. Stolle E, Kidner JH, Moritz RFA. Patterns of evolutionary conservation of microsatellites (SSRs) suggest a faster rate of genome evolution in Hymenoptera than in Diptera. *Genome Biol Evol.* 2013;5(1):151–62.
62. Schlötterer C. Evolutionary dynamics of microsatellite DNA. *Chromosoma.* 2000;109(6):365–71.
63. Li YC, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol.* 2002;11(12):2453–65.
64. Gregory TR. 2005. Animal genome size Database. <http://www.genomesize.com>. Accessed 10 Oct 2017.
65. Murphy RW, Lowcock LA, Smith C, Darevsky IS, Orlov N, MacCulloch RD, Upton DE. Flow cytometry in biodiversity surveys: methods, utility, and constraints. *Amphibia-Reptilia.* 1997;18(1):1–13.
66. Vinogradov AE. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry.* 1998;31(2):100–9.
67. Garner TW. Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome.* 2002;45(1):212–5.
68. Van De Vliet MS, Diekmann OE, Serrao EA, Beja P. Isolation of highly polymorphic microsatellite loci for a species with a large genome size: sharp-ribbed salamander (*Pleurodeles waltl*). *Mol Ecol Resour.* 2009;9(1):425–8.
69. Compeau PEC, Pevzner PA, Tesler G. How to apply de Bruijn graphs to genome assembly. *Nat Biotechnol.* 2011;29(11):987–91.
70. Chen K, Chen L, Fan X, Wallis J, Ding L, Weinstock G. TIGRA: a targeted iterative graph routing assembler for breakpoint assembly. *Genome Res.* 2014;24(2):310–7.
71. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–52.
72. Huang X, Chen XG, Armbruster PA. Comparative performance of transcriptome assembly methods for non-model organisms. *BMC Genomics.* 2016;17(1):1–14.
73. Andersen JC, Mills NJ. iMSAT: a novel approach to the development of microsatellite loci using barcoded Illumina libraries. *BMC Genomics.* 2014;15(1):1.
74. Bachtrog D, Agis M, Imhof M, Schlötterer C. Microsatellite variability differs between dinucleotide repeat motifs—evidence from *Drosophila melanogaster*. *Mol Biol Evol.* 2000;17(9):1277–85.
75. Ellis J, Burke J. EST-SSRs as a resource for population genetic analyses. *Heredity.* 2007;99(2):125–32.
76. Scott KD, Egger P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ. Analysis of SSRs derived from grape ESTs. *Theor Appl Genet.* 2000; 100(5):723–6.
77. Yuan S, Xia Y, Zeng XA. Sex-linked microsatellite marker reveals male Heterogamety in *Quasipaa boulengeri* (Anura: Dicroglossidae). *Asian Herpetol Res.* 2017;8(3):184–9.

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

