MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 686–693 doi: 10.1111/1755-0998.12492

Mitogenome assembly from genomic multiplex libraries: comparison of strategies and novel mitogenomes for five species of frogs

D. J. MACHADO,* M. L. LYRA† and T. GRANT*

*Department of Zoology, Institute of Biosciences, University of São Paulo, R. do Matão 101, São Paulo SP, CEP 05508-090, Brazil, †Department of Zoology, Institute of Biosciences, São Paulo State University, Campus Rio Claro, Av. 24-A 1515, Rio Claro SP, CEP 13506-900, Brazil

Abstract

Next-generation sequencing continues to revolutionize biodiversity studies by generating unprecedented amounts of DNA sequence data for comparative genomic analysis. However, these data are produced as millions or billions of short reads of variable quality that cannot be directly applied in comparative analyses, creating a demand for methods to facilitate assembly. We optimized an *in silico* strategy to efficiently reconstruct high-quality mitochondrial genomes directly from genomic reads. We tested this strategy using sequences from five species of frogs: *Hylodes meridionalis* (Hylodidae), *Hyloxalus yasuni* (Dendrobatidae), *Pristimantis fenestratus* (Craugastoridae), and *Melanophryniscus simplex* and *Rhinella* sp. (Bufonidae). These are the first mitogenomes published for these species, the genera *Hylodes*, *Hyloxalus*, *Pristimantis*, *Melanophryniscus* and *Rhinella*, and the families Craugastoridae and Hylodidae. Sequences were generated using only half of one lane of a standard Illumina HiqSeq 2000 flow cell, resulting in fewer than eight million reads. We analysed the reads of *Hylodes meridionalis* using three different assembly strategies: (1) reference-based (using BOWTIE2); (2) *de novo* (using ABYSS, SOAPDENOVO2 and VELVET); and (3) baiting and iterative mapping (using MIRA and MITOBIM). Mitogenomes were assembled exclusively with strategy 3, which we employed to assemble the remaining mitogenomes. Annotations were performed with MITOS and confirmed by comparison with published amphibian mitochondria. In most cases, we recovered all 13 coding genes, 22 tRNAs, and two ribosomal subunit genes, with minor gene rearrangements. Our results show that few raw reads can be sufficient to generate high-quality scaffolds, making any Illumina machine run using genomic multiplex libraries a potential source of data for organelle assemblies as by-catch.

Keywords: Amphibia, Anura, gene order, Illumina HiSeq, mitochondria, second-generation sequencing

Received 28 July 2015; revision received 9 November 2015; accepted 17 November 2015

Introduction

Most vertebrate mitochondrial genomes (mitogenomes) are about 15–22 kbp, double-stranded, circular DNAs that encode a set of 37 genes (two rRNAs, 13 proteins and 22 tRNAs), as well as a major non-coding region (control region, CR) that accounts for much of the mitogenome size variation (Gissi *et al.* 2008). Mitochondrial DNA (mtDNA) sequences are used in a wide range of studies, from population genetics to phylogenetics, ultimately improving our knowledge on the evolution of both genomes and organisms (e.g. Hancock-Hanser *et al.* 2013; Hulsey *et al.* 2013; Bertrand *et al.* 2015). Until the 2000s, the molecular biology of mitochondrial systems had been studied for only a few model organisms (Boore 1999). More recently, next-generation sequencing (NGS) and advances in bioinformatics tools have enabled the analysis of mitogenomes to extend to non-model organisms on an unprecedented scale (Mardis 2008).

Numerous methods for rapidly assembling mitogenomes directly from shotgun sequencing have been proposed (e.g. Cameron 2014; Gan *et al.* 2014; Lounsberry *et al.* 2015). These methods are intended for fast recovery, assembly and annotation of mitogenomes as the primary research objective. For example, Gan *et al.* (2014) provide a detailed protocol for the fastest recovery, assembly, and annotation of mitogenome using the MITOBIM software (Hahn *et al.* 2013), the MITOS (Bernt *et al.* 2013) annotation web service and data from the Illumina MiSeq platform. However, in addition to studies designed

Correspondence: Denis Jacob Machado, Fax: +55-11-3091-7802; E-mail: denisjacobmachado@gmail.com

specifically to capture mitogenomic sequences, whole-genome sequencing, targeted amplicon sequencing and hybrid enrichment approaches also capture mitogenomic reads as by-catch, albeit with significantly lower coverage and quality. To make use of these reads, an efficient bioinformatics pipeline is required to extract and assemble mitogenomes from limited data.

Here, we add to the methods for harvesting complete mitogenomes from whole-genome multiplex libraries sequenced using the Illumina HiSeq platform and compare the performance of different assembly strategies when read number and quality are limited. As test data, we present novel, near-complete mitogenomes from five South American frog species of the families Bufonidae, Craugastoridae, Dendrobatidae and Hylodidae.

Material and methods

Taxon selection and data archiving

We sequenced five South American species of frogs from four families: the torrent frog *Hylodes meridionalis* (Mertens, 1927) (Hylodidae), the rocket frog *Hyloxalus yasuni* Páez-Vacas, Coloma, & Santos, 2010 (Dendrobatidae), the rain frog *Pristimantis fenestratus* (Steindachner, 1864) (Craugastoridae) and the red-belly toad *Melanophryniscus simplex* Caramaschi and Cruz, 2002 and an undescribed species of beaked toad (*Rhinella acrolopha* group *sensu* Grant & Bolívar-G 2014) that we refer to as *Rhinella* sp. C. (Bufonidae).

Total DNA extraction and sequencing

Muscle tissue samples were stored in 70% ethanol at −20 °C for several months or years. Materials were separately pooled for DNA extraction using the AGENCOURT® DNAdvance™ Genomic DNA Isolation Kit. Total genomic libraries were prepared using a NEBNext® DNA Library Prep Master Mix (NEb #E6040S) and sequenced using an Illumina HiSeq 2000™ at the multiuser high-throughput sequencing facility of the University of São Paulo Luiz de Queiroz College of Agriculture. Libraries were distributed in two lanes of a standard Illumina HiSeq 2000 flow cell and sequenced using the high-throughput module. However, each lane also received an unknown number of additional libraries, reducing the expected total number of reads from approx. 250 000 000 to 38 903 325 (approx. 75% fewer) paired-end reads of 100 bp.

We chose the Illumina platform because it produces high-quality data for various scales of analysis at costs that have decreased substantially relative to other second-generation sequencing instruments (e.g. 454/Roche and SOLiD; see Mardis 2013). Among Illumina platforms, the Genome Analyzer IIx (GA IIx) is less automatable and produces fewer data than the MiSeq and HiSeq systems, which are therefore preferred by most researchers interested in large-scale analysis. Gan *et al.* (2014) selected the Illumina MiSeq over the HiSeq platform due to its reduced run time and more tractable data. Nevertheless, the HiSeq platform is the system of preference in numerous research projects targeting elements of nuclear DNA (nuDNA), such as microsatellite analysis (Castoe *et al.* 2012) and whole-genome sequencing (Sun *et al.* 2015).

Computational resources

All *in silico* procedures were executed using 'ACE', an SGI rackable computer cluster housed in the Museum of Zoology of the University of São Paulo. Selected servers had four 2.3 GHz Operon CPUs with 16 cores each and 256 or 516 GB of memory. After optimization, we were able to reconstruct genomes using a single core and ca. 20 GB of memory. The software environment in ACE consists of a SUSE Linux Enterprise Server with SGI Performance Suite, SGI Management Center and PBS Pro Job Scheduler.

Quality control

As stated by Yang *et al.* (2013: 14), 'to get reliable result [s] in downstream analysis, it is necessary to remove low-quality reads, avoiding mismatches in read mapping and false paths during genome assembly'. Due to its function versatility and run-time efficiency, we selected the HTQC toolkit (Yang *et al.* 2013) to perform read quality assessment and filtration. The complete quality control protocol is described below and the step-by-step procedures are given in Appendix S1 (Supporting information).

Raw reads from each pair were pre-processed using a series of UNIX commands and a package of home-made PYTHON scripts (PATO-FU). The programs ht-stat, ht-filter and ht-trim are components of the HTQC toolkit and were employed as follows: the summary of the sequencing read quality was generated with ht-stat. In order for tile selection to be automated and repeatable, we post-processed the ht-stat results using a homemade Python script (selectTiles.py). Tile removal followed criteria derived from the HTQC guidelines: (i) more than 50% of the reads have quality score below 10; (ii) <10% of the reads have quality >30; and (iii) more than 50% of the reads have quality below 20. Selected tiles were removed with ht-filter. Remaining reads were trimmed with ht-trim, removing low-quality bases from reads' heads or tails. Finally, short reads were removed with ht-filter and the quality of filtered reads was evaluated using

FASTQC (Andrew 2010). Only paired-end filtered reads were used for assembly.

Mitogenome assembly

We analysed the filtered reads of *Hylodes meridionalis* using three assembly strategies: (1) mapping against a reference mtDNA genome ('reference based'); (2) *de novo*; and (3) baiting and iterative mapping. To implement each strategy, we selected the best-commented and most frequently used software in the specialized literature.

Reference-based assembly (1) was performed using BOWTIE2 v2.2.3 (Langmead & Salzberg 2012). The mitogenome of the Tibetan toad *Bufo tibetanus* (NCBI accession number NC 020048; Wang *et al.* 2013), which is currently a junior synonym of *B. gargarizans* (for taxonomic comments see Frost 2015), was selected as reference due to its completeness and phylogenetic position and the reliability of the long PCR-based amplification method used to sequence it.

For *de novo* sequence assembly (2), the programs SOAPDENOVO2 v2.04 (Luo *et al.* 2012), ABYSS v1.5 (Simpson *et al.* 2009) and VELVET v1.2.10 (Zerbino & Birney 2008) were used. SOAPDENOVO2 v2.04 was run with average insert sizes of 150, 200 and 250 bp. ABYSS and VELVET were run for all k-mer sizes from 21 to 63, with incremental steps of 2. BLAT (Kent 2002) was used to map contigs and scaffolds against the reference genome of *B. tibetanus*.

For the baiting and iterative mapping strategy (3), we used MIRA v4.0 (Chevreux *et al.* 1999) and a modified version of MITOBIM.PL v1.6 (Hahn *et al.* 2013). This strategy has two main steps (Hahn *et al.* 2013). First, reads are mapped against a reference sequence in MIRA, effectively generating a new reference based on the most conserved regions. New reads with overlap are then iteratively fished from the read-pool and mapped against the previous reference using MITOBIM. Each iteration in MITOBIM expands the novel reference sequence until reaching a stationary number of reads. This approach only returns a single-padded consensus sequence in the end, but sequences can be connected by 'N' to indicate that the fragments are not connected by reads are probably not contiguous.

Four baiting and iterative mapping strategies were employed: (i) mapping to the complete mitogenome of a closely related species (*B. tibetanus*); (ii) mapping to the mitogenome of a more distantly related species (a salamander, *Rhyacotriton variegatus*; NCBI accession number NC 006331; Mueller *et al.* 2004); (iii) baiting with a barcode seed (the cytochrome C oxidase subunit I [COI] gene sequence from *B. tibetanus*, NCBI accession number NC 020048, 5533–7044 bp) with the *de novo* option off; and (iv) same as (iii) but with the *de novo* option on. Only consensus sequences with average coverage >20 and average quality >80 were accepted. If more than one consensus sequence was recovered, the longest one was chosen for further analysis.

The optimal mitogenome assembly strategy was selected according to the number of reads used, total ungapped sequence size, average coverage and consensus quality. This strategy was then applied to assemble mitogenomes using the libraries *of Hyloxalus yasuni*, *Pristimantis fenestratus*, *Melanophryniscus simplex* and *Rhinella* sp. C. The complete bioinformatics protocol for assembly is available in Appendix S2 (Supporting information).

Mitogenome annotation and comparison

Assemblies in CAF format were parsed using a home-made Python script (parseCaf.py) to extract DNA data and evaluate the coverage and quality of each mtDNA element. Preliminary *de novo* mitogenome annotation used the mitochondrial genome annotation server MITOS (Bernt *et al.* 2013) with default parameters. Additional search and validation of tRNA sequences were performed using ARWEN (Laslett & Canbäck 2008) and tRNAscan-SE (Lowe & Eddy 1997; Schattner *et al.* 2005). Automated annotation was confirmed and edited manually by comparison to published anuran mitogenomes (Table S1, Supporting information). The control region (CR), which typically lies between cytochrome B (CytB) and the LTPF tRNA cluster in neobatrachians (Zhang *et al.* 2013), was annotated using sequence similarity searching with BLAST using default parameters (Altschul *et al.* 1990).

Results

Software

Home-made PYTHON scripts (PATO-FU, selectTiles and parseCaf) are available at http://www.ib.usp.br/grant/anfibios/researchSoftware.html and https://gitlab.com/MachadoDJ/ under the GNU General Public License version 3.0 (GPL-3.0). We modified the MITOBIM original script so it would create manifest files for MIRA pointing to a directory in a local file system in a cluster environment. Modifications to MITOBIM allow multiple mitogenomes to be reconstructed simultaneously using the same compute node. The modified MITOBIM script is available at http://www.ib.usp.br/grant/anfibios/researchHPC.html.

Quality control results

Comparison of quality reports before and after quality control shows major improvements in per base/tile sequencing quality and over-represented sequences (see

Table S2, Supporting information). Some filtered sequence files still failed per base sequence content and k-mer content tests; however, according to the FASTQC help page (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/; last access: May 4, 2015), libraries derived from random priming will nearly always show k-mer bias, and sequences subjected to aggressive trimming are more likely to present per base sequence content bias. Quality control took <1 h total computation time using ACE, with <5 min hands-on time.

Comparison of assembling strategies

Reference-based genome assembly using BOWTIE2 failed to align mtDNA sequence reads to the *B. tibetanus* reference genome. Likewise, BLAT mapping failed to find mtDNA sequences within contigs and scaffolds generated using the *de novo* sequencing strategy with ABYSS, SOAPDENOVO2 and VELVET. Only the baiting and iterative mapping strategy with MIRA and MITOBIM succeeded in assembling mtDNA sequences.

We were able to assemble mtDNA for all three variations of the baiting and iterative mapping protocol. However, the consensus sequences generated using a barcode seed with the '–denovo' option in MITOBIM did not pass our minimum quality criteria. The remaining assembled consensus sequences were compared according to the number of reads used, total ungapped sequence size, average coverage and consensus quality (see Table 1). The N50 and N90 values are incalculable because only one contig remains in the last iteration. The longest ungapped consensus sequences that passed minimum quality criteria were achieved by mapping to the complete mitogenome of the more closely related species (*B. tibetanus*).

Assembly of mitogenomes using MIRA and MITOBIM took variable amounts of time depending on the reference used and the number of iterations required by MITOBIM. However, mitogenome assembly using the complete frog mitogenome as reference required fewer iterations and <3-h computation time, with <5 min hands-on time.

Mitogenomic sequences and gene rearrangements

We recovered the nearly complete mitogenome of all five species of frogs, including the standard 13 protein-coding genes, 2 ribosomal subunits and 21–22 tRNAs (Fig. 1). The number of reads used for assembling mitogenomes and the size of each ungapped consensus sequenced are shown in Table 1.

Table 1 Baiting and iterative mapping assembly statistics. Strategies: (1) closely related mitogenome (the Tibetan toad, *Bufo tibetanus*, NCBI accession number NC_020048); (2) distantly related genome (a salamander, *Rhyacotriton variegatus*, NCBI accession number NC_006331); (3) barcode seed (the COI gene sequence of *B. tibetanus*, NCBI accession number NC_020048, 5533–7044 bp), *de novo* option off; (4) same as previous, *de novo* option on. The chosen mitogenomic sequences for each species are highlighted. See gene order information in Table S2 (Supporting information)

			Reads (×2)					
Species	Strat.	Iterations	Raw	Filtered	Used	Ungapped consensus size (bp)	Avg. coverage	Avg. quality
Hylodes meridionalis	1	18	8 608 779	7 745 168	4389	16 166	26.88	81
	2	49			4122	15 651	20.96	60
	3	84			3389	12 079	29.66	87
	4	37			3261	13 205	23.78	79
Hyloxalus yasuni	1	35	6 894 772	6 160 445	4650	16 052	28.28	80
	2	29			4625	15 946	23.06	64
	3	133			3324	10 330	33.44	87
	4	Fail	–	–	–	–	–	–
M. simplex	1	8	7 958 678	7 166 358	3717	16 498	23.2	81
	2	49			3261	13 633	17.3	55
	3	81			2 503	10 258	26.09	87
	4	38			1690	7404	22.07	69
P. fenestratus	1	17	4 714 625	4 213 416	23 566	17 892	130.46	87
	2	47			16 890	15 880	76.13	63
	3	118			17 483	15 966	107.52	88
	4	Fail	–	–	–	–	–	–
Rhinella sp. C.	1	13	9 874 464	8 738 815	4765	17 050	28.79	84
	2	39			4081	15 879	20.58	62
	3	68			1700	6912	25.91	86
	4	36			1641	7608	0	83

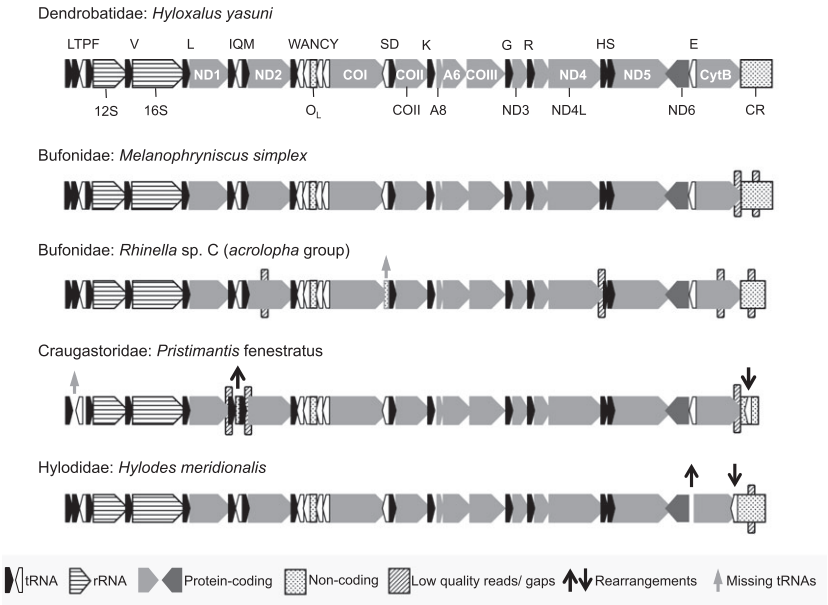

Fig. 1 Gene order and orientation for mitogenomes of five species of South American frogs. The mitogenome of *Hyloxalus yasuni* has all the expected elements in the most common gene order in Neobatrachia. The remaining mitogenomes follow alphabetical order (family: genus). Graphical representation shows elements pointed in the corresponding direction in the mitogenome.

We also recovered partial CR sequences for all five mitogenomes. The partial CR of *Hyloxalus yasuni* was recovered in a single, contiguous sequence with all coding genes. The CR sequences of the remaining four mitogenomes were recovered as non-contiguous sequences, with fragments varying from 145 bp in *Rhinella* sp. C. to 2302 bp in *P. fenestratus*.

Screening with parseCaf allowed us to identify only a few poorly sequenced (<10× coverage, quality <40) regions in all mitogenomes, in most cases associated with homopolymeric regions (poly-G or poly-C sequences). In the mitogenome of *M. simplex*, there is a poorly sequenced fragment at the 5′ end of CytB that resulted in a small duplication that was removed manually in the final assembly. In the *Rhinella* sp. C. mitogenome, we found three poorly sequenced regions: two small regions inside the ND2 and CytB gene sequences and one region between ND4 and tRNA-H. The stop codon for ND4 and a fragment of approx. 20 bp of the tRNA-H sequence could not be assembled. These regions were also edited manually and 'Ns' were included in the final assembly. Finally, we found two poorly sequenced regions in the mitogenome of *P. fenestratus*: a small fragment immediately before the tRNA-I sequence and another fragment just after the tRNA-M sequence.

Most genes in the five mitogenomes we report are transcribed from the H-strand, exceptions being ND6 and eight tRNA genes (Fig. 1; Table S1, Supporting information), as described in other anurans (Irisarri *et al.* 2012; Zhang *et al.* 2013). The gene arrangement in *Hyloxalus yasuni* and *Melanopryniscus simplex* mitogenomes follows the most common order of Neobatrachia (Zhang *et al.* 2005, 2013; Kurabayashi & Sumida 2013). In the case of *Rhinella* sp. C., the gene arrangement also matches the arrangement found in most neobatrachian anurans, except that we were unable to find the tRNA-S2 gene. Although there is a non-coding region in the expected position of this tRNA (i.e. just before tRNA-D), the sequence has low similarity with the tRNA-S2 sequence from other anurans and we were unable to predict its secondary structure.

One novel tRNA gene rearrangement was observed in *Hylodes meridionalis*, in which the tRNA-E is located between CytB and the major non-coding region rather than the typical neobatrachian location between ND6 and Cyt B (Fig. 1). Given the high coverage and quality of this fragment, this unique pattern is unlikely to be an artefact of assembly. Similarly, in *P. fenestratus* we found a new arrangement in the LTPF tRNA cluster, the tRNA-T was not recovered, and the IQM gene cluster is modified such that tRNA-Q is absent and now occurs inside the control region.

Discussion

Numerous studies have employed the baiting and iterative mapping using MIRA and MITOBIM (e.g. Doyle *et al.* 2014; Grau *et al.* 2015). Most of these studies share a set of characteristics: species were sequenced one at a time; sequencing the mitochondrial genome was at least one of the main objectives; and the number and quality of the sequence reads were high. However, when libraries are multiplexed (e.g. several libraries of different species are sequenced simultaneously) and/or genomic DNA samples have been enriched for particular loci (see Jones & Good 2015), read number will decrease substantially,

with possible negative effects on overall read quality. The methods described here can be used to assemble organellar genomes in this latter scenario.

Our results show that even a low number of reads can be enough to provide high coverage for most of the mitochondrial genome, allowing organellar genomes to be extracted and assembled as by-catch from any Illumina HiSeq machine run using total genomic libraries, even when libraries are multiplexed. The strategy presented here might also be effective for other technologies, since MIRA and MITOBIM also accept Ion Torrent and 454 data as input. We note that MIRA's manual suggests that the program may not be suitable for data sets with more than 20–40 million reads and that in some cases it may be necessary to randomly sample reads from the original pool, but read number will be naturally reduced for multiplex libraries.

Mitogenomes assembled by mapping to a more closely related reference mitogenome (in this case, another anuran) were longer and required fewer iterations than those generated by mapping to a more distantly related mitogenome (a salamander), and the assemblies obtained by mapping to a complete mitogenome generated longer consensus sequences than by using barcode seeds. However, no other significant differences were observed in sequence order and composition when reference sequences were changed. It should be noted that the anuran reference mitogenome we employed is more closely related to our four test species than is the salamander reference mitogenome, but it is deeply nested within the family Bufonidae and is, therefore, not especially close to any of our test species (Frost *et al.* 2006). Consequently, we suggest choosing references based first on sequence length and second on phylogenetic proximity.

Mitochondrial DNA has historically been the molecule of choice to address problems in phylogenetics and population genetics. The availability of complete or partial mitogenomes from different species provides a unique model to understand mechanisms of genome evolution (Gissi *et al.* 2008). Several genome features, such as molecular evolutionary rates, gene content, gene order and secondary structure of RNAs, can be explored in a phylogenetic context, but the utility of these data sets is fully dependent on taxon sampling (Boore 1999; Gissi *et al.* 2008). By using an optimized *in silico* strategy to recover mitogenomes from NGS data, the available mitogenome data set can be efficiently increased and can enable comparative genomic analysis.

Frog mitogenomics has been a slow yet steadily growing field of research. At the time this paper was submitted for publication, there were 192 complete and 109 partial mitogenomes of different species of Amphibia (Gymnophiona, Caudata and Anura) available in NCBI's Organelle Genome Resources database (Wolfsberg *et al.* 2001), and only 83 complete and 56 near-complete (>14 000 bp) mitogenomes of anurans of 22 different families. This constitutes a very small proportion of the 7395 known species of amphibians and 6500 species of Anura (Frost 2015; accessed July 21, 2015), and the five new mitogenomes we present represent three families and five genera for which mitogenomes were unknown previously.

Among vertebrates, amphibian mitogenomes have the greatest variation in gene order. Gene rearrangements are present in all orders of amphibian. For example, San Mauro *et al.* (2006) found rearrangements in the WANCY tRNA cluster in the caecilian genus *Siphonops*, Mueller & Boore (2005) found rearrangements in ND6-tRNA-E and WANCY tRNA cluster in plethodontid salamanders, and Kurabayashi *et al.* (2008) reported high frequency of genomic reorganization in the mitochondria of members of the anuran family Mantellidae. The increasing number of mitogenomes available for this group contributed to overturning the accepted view that mitochondrial gene organization in vertebrates was stable (Boore 1999; Saccone *et al.* 1999; Gissi *et al.* 2008).

Even though the available anuran mitogenomes are a small sample of the diversity of frogs, numerous mitochondrial gene rearrangements have already been reported for frogs. Irisarri *et al.* (2012) found new arrangements for the ND5 gene and ND6-tRNA-E cluster in the neobatrachian frogs *Lechriodus melanopyga* (Limnodynastidae) and *Heleophryne regis* (Heleophrynidae) and also reported modifications in the tRNA clusters of neobatrachians. Zhang *et al.* (2013) and Xia *et al.* (2014) also found several different gene orders for Neobatrachia that are mainly associated with tRNA clusters LTPF, WANCY, and IQM and the occurrence of pseudogenes. Here, we sequenced five new genomes and found three different gene arrangements associated with tRNA clusters, one in *Hylodes* and two in *Pristimantis*.

The mitogenomes presented here should contribute to future phylogenetic analyses of Amphibia and help improve understanding of the evolution of mitochondrial gene order arrangement in this taxon. At this point, however, the taxonomic and phylogenetic significance of these rearrangements is unclear and requires comparison with additional mitogenomes of closely related frogs.

Conclusions

We have reported the first mitogenomic sequences for the anuran families Craugastoridae and Hylodidae and the genera *Hylodes*, *Hyloxalus*, *Pristimantis*, *Melanophryniscus* and *Rhinella*. The mitogenomes of *M. simplex* and *Rhinella* sp. C. are the first mitogenomes of Neotropical bufonids. *Melanoprhyniscus* is the sister group of all other bufonids (e.g. Peloso *et al.* 2012), making the mitogenome of *M. simplex*

especially important for studies of mitochondrial evolution in this large, nearly cosmopolitan family.

By employing the baiting and iterative mapping strategy tested herein, workers can assemble organelle genomes as by-catch for use in comparative studies. Our results demonstrate that even a low number of reads can be sufficient to assemble high-quality mitogenomes, making any Illumina HiSeq run using libraries prepared with total genomic DNA extractions a potential source of organelle assemblies.

Acknowledgements

We thank the editor and three anonymous reviewers for their insightful suggestions on earlier drafts of the paper. Sequences were generated with the assistance of Mariane Targino and Rachel Montesinos in the course 'Next-Generation Sequencing (NGS) - Applications in Plant Systematics and Evolution' taught by Mônica Carlsen and Lúcia Garcez Lohmann in the Institute of Biosciences, University of São Paulo, May 15–29, 2013. This work was made possible thanks to the financial support from the São Paulo Research Foundation (FAPESP; grants 2012/10000-5, 2013/05958-8) and Brazilian National Counsel of Technological and Scientific Development (CNPq grant 305234/2014-5).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipmanl DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Andrew S (2010) FastQC, a quality control tool for high throughput sequence data. Retrieved October 4, 2015, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bernt M, Donath A, Jühling F *et al.* (2013) MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, **69**, 313–319.

Bertrand B, Alburaki M, Legout H, Moulin S, Mougel F, Garnery L (2015) MtDNA COI-COII marker and drone congregation area: an efficient method to establish and monitor honeybee (*Apis mellifera* L.) conservation centres. *Molecular Ecology Resources*, **15**, 673–683.

Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research*, **27**, 1767.

Cameron SL (2014) How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Systematic Entomology*, **39**, 400–411.

Caramaschi U, Cruz CAG (2002) Taxonomic status of Atelopus pachyrhynus Miranda-Ribeiro, 1920, redescription of Melanophryniscus tumifrons (Boulenger, 1905), and descriptions of two new species of Melanophryniscus from the state of Santa Catarina, Brazil (Amphibia, Anura, Bufonidae). *Arquivos do Museu Nacional*, **60**, 303–314.

Castoe TA, Poole AW, de Koning APJ *et al.* (2012) Rapid microsatellite identification from illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE*, **7**, e30953.

Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Proceedings of the German Conference on Bioinformatics*, **99**, 45–56.

Doyle JM, Katzner TE, Bloom PH, Ji Y, Wijayawardena BK, DeWoody A (2014) The genome sequence of a widespread apex predator, the golden eagle (*Aquila chrysaetos*). *PLoS ONE*, **9**, 20–22.

Frost DR (2015) Amphibian Species of the World: an Online Reference. Retrieved October 4, 2015, from http://research.amnh.org/herpetology/amphibia/index.php.

Frost DR, Grant T, Faivovich J *et al.* (2006) The amphibian tree of life. *The American Museum of Natural History*, **297**, 1–370.

Gan HM, Schultz MB, Austin CM (2014) Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evolutionary Biology*, **14**, 19.

Gissi C, Iannelli F, Pesole G (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, **101**, 301–320.

Grant T, Bolívar-G W (2014) A new species of semiarboreal toad with a salamander-like ear (Anura: Bufonidae: *Rhinella*). *Herpetologica*, **70**, 198–210.

Grau JH, Nuñez JJ, Plötner J, Poustka A (2015) The complete mitochondrial genome of *Telmatobufo australis* (Amphibia: Anura: Calyptocephalellidae). *Mitochondrial DNA*, **1736**, 1–2. doi: 10.3109/19401736.2015.1053082.

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - a baiting and iterative mapping approach. *Nucleic Acids Research*, **41.13**, e129.

Hancock-Hanser BL, Frey A, Leslie MS, Dutton PH, Archer FI, Morin PA (2013) Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology Resources*, **13**, 254–268.

Hulsey DC, Keck BAH, O'Meara BC (2013) Mitochondrial genome primers for Lake Malawi cichlids. *Molecular Ecology Resources*, **13**, 347–353.

Irisarri I, Mauro DS, Abascal F, Ohler A, Vences M, Zardoya R (2012) The origin of modern frogs (Neobatrachia) was accompanied by acceleration in mitochondrial and nuclear substitution rates. *BMC Genomics*, **13**, 626.

Jones MR, Good JM (2015) Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*. doi: 10.1111/mec.13304. [Epub ahead of print].

Kent WJ (2002) BLAT - The BLAST-like alignment tool. *Genome Research*, **12**, 656–664.

Kurabayashi A, Sumida M (2013) Afrobatrachian mitochondrial genomes: genome reorganization, gene rearrangement mechanisms, and evolutionary trends of duplicated and rearranged genes. *BMC Genomics*, **14**, 633.

Kurabayashi A, Sumida M, Yonekawa H *et al.* (2008) Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. *Molecular Biology and Evolution*, **25**, 874–891.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Laslett D, Canbäck B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*, **24**, 172–175.

Lounsberry ZT, Brown SK, Collins PW, Henry RW, Newsome SD, Sacks BN (2015) Next-generation sequencing workflow for assembly of nonmodel mitogenomes exemplified with North Pacific albatrosses (*Phoebastria* spp.). *Molecular Ecology Resources*, **15**, 893–902.

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, **25**, 955–964.

Luo R, Liu B, Xie Y *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience*, **1**, 18.

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, **24**, 133–141.

Mardis ER (2013) Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, **6**, 287–303.

Mertens R (1927) Neue Froschlurch aus Rio Grande do Sul, Brasilien. Blätter für Aquarien- und Terrarien-Kunde. *Stuttgart*, **38**, 287–290.

Mueller RL, Boore JL (2005) Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Molecular Biology and Evolution*, **22**, 2104–2112.

Mueller RL, Macey JR, Jaekel M, Wake DB, Boore JL (2004) Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 13820–13825.

Páez-Vacas M, Coloma LA, Santos JC (2010) Systematics of the Hyloxalus bocagei complex (Anura: Dendrobatidae), description of two new cryptic species, and recognition of *H. Maculosus*. *Zootaxa*, **2711**, 1–75.

Peloso PLV, Faivovich J, Grant T, Luiz J (2012) An extraordinary new species of *Melanophryniscus* (Anura, Bufonidae) from southeastern Brazil. *American Museum Novitates*, **3762**, 1–32.

Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene*, **238**, 195–209.

San Mauro D, Gower DJ, Zardoya R, Wilkinson M (2006) A hotspot of gene order rearrangement by tandem duplication and random loss in the Vertebrate mitochondrial genome. *Molecular Biology and Evolution*, **23**, 227–234.

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research*, **33**, W686–W689.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research*, **19**, 1117–1123.

Steindachner F (1864) Batrachologische Mittheilungen. *Verhandlungen des Zoologisch-Botanischen Vereins in Wien*, **14**, 239–288.

Sun Y, Xiong Z, Xiang X *et al*. (2015) Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proceedings of the National Academy of Sciences*, **112**, E1257–E1262.

Wang X, Wang Y, Yue B, Zhang X, Liu S (2013) The complete mitochondrial genome of the *Bufo tibetanus* (Anura: Bufonidae). *Mitochondrial DNA*, **24**, 186–188.

Wolfsberg TG, Schafer S, Tatusov RL, Tatusova TA (2001) Organelle genome resources at NCBI. *Trends in Biochemical Sciences*, **26**, 199–203.

Xia Y, Zheng Y, Miura I, Wong PB, Murphy RW, Zeng X (2014) The evolution of mitochondrial genomes in modern frogs (Neobatrachia): nonadaptive evolution of mitochondrial genome reorganization. *BMC Genomics*, **15**, 691.

Yang X, Liu D, Liu F *et al*. (2013) HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics*, **14**, 33.

Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821–829.

Zhang P, Zhou H, Chen Y-Q, Liu Y-F, Qu L-H (2005) Mitogenomic perspectives on the origin and phylogeny of living amphibians. *Systematic Biology*, **54**, 391–400.

Zhang P, Liang D, Mao R-L, Hillis DM, Wake DB, Cannatella DC (2013) Efficient sequencing of anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. *Molecular Biology and Evolution*, **30**, 1899–1915.

D.J.M. and T.G. conceived and designed the research. T.G. coordinated the study, provided the biological material and performed taxonomic identifications. D.J.M. carried out bioinformatic analysis, including quality control, assembling and coding. M.L.L. analyzed and annotated the mitogenomes. D.J.M. and M.L.L. prepared tables and figures. D.J.M. and M.L.L. also made data and scripts publicly available. All authors wrote the manuscript and gave final approval for publication together.

Data Accessibility

DNA sequences: GenBank accessions KT221610–KT221614; NCBI SRA SRP061702. Final DNA sequence assembly and homemade scripts: DRYAD entry doi: 10.5061/dryad.pn97c. Home-made scripts are also available at http://www.ib.usp.br/grant/anfibios/researchSoftware.html and https://gitlab.com/MachadoDJ.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Gene order.

Table S2 General FASTQC statistics.

Table S3 Base composition and other features of mitochondrial genomes.

Appendix S1 Bioinformatics protocols for quality control.

Appendix S2 Bioinformatics protocols for sequence assembly.