

RESEARCH ARTICLE

# Ultraconserved Elements Sequencing as a Low-Cost Source of Complete Mitochondrial Genomes and Microsatellite Markers in Non-Model Amniotes

Fábio Raposo do Amaral<sup>1\*</sup>, Leandro G. Neves<sup>2</sup>, Márcio F. R. Resende, Jr<sup>2</sup>, Flávia Mobili<sup>1</sup>, Cristina Y. Miyaki<sup>3</sup>, Katia C. M. Pellegrino<sup>1</sup>, Cibele Biondo<sup>4</sup>

**1** Universidade Federal de São Paulo, Instituto de Ciências Ambientais, Químicas e Farmacêuticas, Laboratório de Genética Evolutiva, Rua Professor Artur Riedel, 275, Diadema, SP, 09972–270, Brazil,

**2** RAPiD Genomics, LLC, 747 SW 2nd Avenue, Gainesville, FL, 32601, United States of America,

**3** Universidade de São Paulo, Departamento de Genética e Biologia Evolutiva, Rua do Matão, 277, Cidade Universitária, São Paulo, SP, 05508–090, Brazil, **4** Universidade Federal do ABC, Centro de Ciências Naturais e Humanas, Rua Arcturus 03, Jardim Antares, São Bernardo do Campo, SP, 09606–070, Brazil

\* [fabioraposo@gmail.com](mailto:fabioraposo@gmail.com)



**OPEN ACCESS**

**Citation:** Raposo do Amaral F, Neves LG, Resende MFR, Jr, Mobili F, Miyaki CY, Pellegrino KCM, et al. (2015) Ultraconserved Elements Sequencing as a Low-Cost Source of Complete Mitochondrial Genomes and Microsatellite Markers in Non-Model Amniotes. PLoS ONE 10(9): e0138446. doi:10.1371/journal.pone.0138446

**Editor:** Gen Hua Yue, Temasek Life Sciences Laboratory, SINGAPORE

**Received:** April 14, 2015

**Accepted:** August 31, 2015

**Published:** September 17, 2015

**Copyright:** © 2015 Raposo do Amaral et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Sequences and raw data were deposited in Genbank (mitogenomes: KT272188-KT272190; microsatellites with designable primers: KT272191-KT272395) and Dryad (raw Illumina sequences, assemblies and probes used in sequence capture: doi:10.5061/dryad.33656).

**Funding:** Funding was provided by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP, grant numbers 2011-50143-7 and 2011/23155-4 to FR, website: <http://www.bv.fapesp.br/21144>; grant number 2013/50297-0 to CYM, website:

## Abstract

Sequence capture of ultraconserved elements (UCEs) associated with massively parallel sequencing has become a common source of nuclear data for studies of animal systematics and phylogeography. However, mitochondrial and microsatellite variation are still commonly used in various kinds of molecular studies, and probably will complement genomic data in years to come. Here we show that besides providing abundant genomic data, UCE sequencing is an excellent source of both sequences for microsatellite loci design and complete mitochondrial genomes with high sequencing depth. Identification of dozens of microsatellite loci and assembly of complete mitogenomes is exemplified here using three species of *Poospiza* warbling finches from southern and southeastern Brazil. This strategy opens exciting opportunities to simultaneously analyze genome-wide nuclear datasets and traditionally used mtDNA and microsatellite markers in non-model amniotes at no additional cost.

## Introduction

Mitochondrial DNA and microsatellites have been the markers of choice since the emergence of molecular data in studies of ecology and evolution [1–2]. Sequencing of mitochondrial DNA and isolation of microsatellites have relied for a long time on classic methods such as Sanger sequencing of amplicons [3], library enrichment [4] or cloning [5]. More recently, next-generation sequencing (NGS) has replaced Sanger sequencing in both prospection of microsatellite loci [6] and assembly of mitogenomes [7] due to its higher efficiency and lower costs per base pair.

<http://www.bv.fapesp.br/7387>) and National Science Foundation (DOB 1343578 to CYM, [http://nsf.gov/awardsearch/showAward?AWD\\_ID=1343578](http://nsf.gov/awardsearch/showAward?AWD_ID=1343578)). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. RAPID Genomics provided support in the form of salaries for authors LGN and MFRR, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

**Competing Interests:** LGN and MFRR are co-owners of RAPID Genomics, which generated the sequences used in this work. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

The benefits of using multilocus data in estimates of historical demography, population structure and kinship have received strong theoretical and empirical support [8–12]. Recent emergence of massively parallel sequencing associated with techniques of genome reduction provide a cost-effective method to acquire large numbers of independent loci in a short time-frame. Consequently, studies based on techniques such as RAD-seq and sequence capture are becoming increasingly common [13–14]. Despite rapid increase in use of genomic data in evolutionary studies, mtDNA and microsatellite markers are still commonly used, and they will likely be employed in biodiversity studies along with NGS data for years to come.

Among the protocols currently used to obtain genomic variation, next-generation sequencing of ultraconserved elements has become a major source of data for animal phylogenetics in the last half decade. Nuclear ultraconserved elements (referred to henceforth as “UCEs”) are genomic regions first discovered in humans and which are highly conserved (>95% of identity) across distantly related organism, such as mammals, birds and fishes [15]. Such regions have been recently adopted as molecular markers for phylogenetic studies [16] due to their universality, which allows sequencing of thousands of homologous markers from sets of species with little genomic resources available. In addition, despite their conservatism, UCE flanking regions may contain sufficient information to study both deep [17] and shallow divergences [18]. The high similarity of UCE sequences allows the same set of target regions to be used across a number of distantly related amniotes [16] and more recently, fishes [19] and arthropods [20]. Despite its targeted nature, sequence capture is not 100% specific, and sequences located outside the target regions—called off-target sequences—may reach levels as high as 60% of all sequences obtained using commercial exome-capture kits [21,22]. Those unexpected sequences may be assembled, for example, into large mtDNA contigs [18], sometimes being sufficient to assemble an entire mitogenome (as briefly mentioned with UCEs in [23] and [24]). Mitogenomes are routinely assembled in human exome studies [25–27], but are still little explored subproducts of UCE studies.

Here we show in detail that the increasingly common sequence capture of UCEs may be used simultaneously as a source of dozens to hundreds of microsatellite loci and mitogenomes with high sequencing depth, rendering UCE sequencing a low-cost, hybrid strategy to obtain traditional markers alongside powerful genomic data. We use as exemplars three species of *Poospiza* warbling finches from the subtropical Brazilian Atlantic Forest, from which we obtained dozens of microsatellite loci and complete mtDNA genomes.

## Material and Methods

### Tissue sampling

We included three samples in this study, which represent one sample from each of three species of warbling-finches: the Gray-Throated Warbling Finch (*Poospiza cabanisi*), Buff-Throated Warbling Finch (*Poospiza lateralis*) and Bay-chested Warbling Finch (*Poospiza thoracica*). These species inhabit montane and subtropical forests in SE and S Brazil, and they are being studied as part of a larger phylogeography study (Amaral *et al.* in prep). A recent phylogenetic study [28] including two of these species (*P. thoracica* and *P. cabanisi*) suggests that they are not close relatives, and that *Poospiza* may be polyphyletic. *Poospiza lateralis* and *P. cabanisi*, in turn, appear to have split recently, and their divergence may be among the most recent in the Neotropics (Amaral *et al.*, unpublished data). *Poospiza lateralis* and *P. cabanisi* are allopatric, while *P. thoracica* is sympatric to both. Specimens were attracted in the field using playback recordings and collected using an air shotgun [29]. This research has been approved by the ethics committee of the Federal University of São Paulo (CEP 0069/12), including the adopted sampling procedure. Field expeditions and specimen collection permits for all visited localities

were granted by ICMBio (permit numbers 14673 and 30840). All specimens were housed at the ornithological collection of the Museu de Zoologia da Universidade de São Paulo (MZUSP), and tissue samples deposited at the Laboratório de Genética e Evolução de Aves da Universidade de São Paulo (LGEMA-USP, see [Table 1](#)).

## Laboratory methods

We extracted total DNA from muscle samples using the Qiagen DNeasy kit (Valencia, CA) according to the manufacturer's protocol, including the suggested RNase treatment. Sequence capture and sequencing of ultraconserved elements were performed according to the original UCE capture and sequencing protocols [16] by RAPiD Genomics (Gainesville, FL, USA). Modifications made to the original protocol consisted of the reduction from 2,560 to 650 probes (targeting 634 loci selected to cover all chromosomes of the *Gallus* genome), use of 100 bp paired-end Illumina HiSeq 2000 sequencing run, and use of 16 cycles in both pre- and post-capture PCR reactions. We sequenced the three specimens used here in a multiplexed batch of 48 individuals (i.e. sequencing included 45 individuals from other studies).

## Sequence quality control and assembly

We separated raw sequences by individual tags using Illumina's Casava software. Sequences were initially evaluated using FastQC 0.10.1 [30]. Adapters, barcodes and low quality regions were removed using Illumiprocessor 2.0.7 [31], which allows processing of Illumina sequencing reads using the trimming tool Trimmomatic 0.32.1 [32]. The resulting reads were processed using Phyluce 1.4 [33], which is a software package that allows processing of trimmed reads into aligned loci. We performed processing in Phyluce up to the assembly stage, using the Trinity assembler v20140717 [34]. All analyses were performed using default parameters. Contigs resulting from assembly were used for downstream analyses.

## mtDNA assembly, variant calling, annotation and phylogenetic inference

Contigs were compared to all avian mtDNA genomes at NCBI as of August 6<sup>th</sup> 2014 using the standalone version of BLAST 2.2.28 [35]. Contigs with a strong match (e-value < 0.1) to avian mtDNA were used to manually assemble the mitogenome in Bioedit 7.1.3 [36]. Despite the haploid nature of mtDNA, we called SNPs in order to detect potential heteroplasmy and/or PCR errors. Paired reads processed in Illumiprocessor were resynchronized using custom scripts and then mapped to the respective mitogenome using Bowtie2 v2.2.4 [37]. PCR duplicates, which consist of identical sequences that arise due to amplification during library preparation, were removed using Samtools 0.1.19 [38], and variants and indels were called with Freebayes 0.9.15 [39] using the diploid option and standard filters (-m 30-q 20-R 0-S 0). Sequencing depth statistics were obtained using Bedtools 2.17 [40].

Complete mitochondrial genomes were annotated based on the automatic methods of MITOS [41] and DOGMA [42] using default parameters. Incongruences in the results obtained between the two methods for each species were manually adjusted based on comparisons to the annotated mtDNA genomes of *Gallus gallus* (NC\_001323) and *Emberiza pusilla* (NC\_021408). Final annotation of the mitogenomes was performed using Geneious 6.1.8 [43].

In order to evaluate mitochondrial sequence authenticity, we performed a phylogenetic analysis including cytochrome b sequences from the mitogenomes obtained here and public sequences from species attributed to the Poospizinae sub-family [28], which were sequenced by those authors using standard Sanger sequencing. We included representatives of the genera *Poospiza*, *Compsospiza*, *Cnemoscopus*, *Cypsnagra*, *Donacospiza*, *Hemispingus*, *Nephelornis*, *Piezorina*, *Pyrrhocoma*, *Thlypopsis*, *Urothraupis* and *Xenospingus*. We generated sequence

**Table 1. Sample information and descriptive sequence statistics.**

	<i>Poospiza cabanisi</i>	<i>Poospiza lateralis</i>	<i>Poospiza thoracica</i>
Field number	FRA67	FRA46	SC04
Locality	Pelotas, RS, Brazil	Dores do Rio Preto, ES Brazil	Urubici, SC, Brazil
Coordinates	-31.5265, -52.5505	-20.5125; -41.8075	-28.1449; -49.6324
Read pairs	2527058	3105898	2935389
Contigs assembled	8638	11392	9365
UCE loci captured	625	632	626
Mitogenome length (bp)	16768	16773	16772
Mitogenome GC content (%)	47.1	47.1	46.6
mtDNA Sequencing depth (X) with duplicates (Mean, st. dev., min, max)	1161, 263, 37, 2258	1081, 294, 9, 2419	588, 138, 32, 1083
mtDNA Sequencing depth (X) without duplicates (Mean, st. dev., min, max)	95, 10, 6, 122	74, 9, 4, 95	81, 11, 2, 112
Number of microsatellite loci (total / with primer designed)			
Dinucleotide	250 / 54	307 / 57	277 / 52
Trinucleotide	85 / 6	105 / 15	87 / 13
Tetranucleotide	23 / 4	30 / 3	27 / 3
Pentanucleotide	2 / 0	6 / 0	4 / 0
Hexanucleotide	1 / 0	3 / 2	2 / 0
All repetition types	361 / 64	451 / 77	397 / 68

MZUSP and LGEMA numbers pending. “Read pairs” refer to number of pairs resulting from resynchronization of files processed with Illumiprocessor.

doi:10.1371/journal.pone.0138446.t001

alignments in Muscle [44] and performed model selection based on the Akaike Information criterion as implemented in MrModeltest 2 [45], which indicated HKY+I+G as the best-fit model. We performed Bayesian phylogenetic inference in MrBayes 3.2.3 [46] using the CIPRES infrastructure [47], adopting a conservative burnin of 25% and *Xenospingus concolor* as out-group. MrBayes runs were performed twice to verify convergence of topology and posterior probabilities obtained from independent runs.

### Identification of microsatellite loci candidates

Contigs resulting from the Trinity assembler (implemented in Phyluce) were used as input for QDD 1.3 [48], which is a software package that detects microsatellites and allows primer design. We searched for pure di-, tri-, tetra-, penta- and hexa-nucleotides with at least five repeats. Potential primers were designed using default settings in Primer3 [49] as implemented in QDD. Sequences used for primer design were compared to the corresponding species’ mitogenome using BLAST, and removed in case of matches. The same procedure was also performed using sex chromosomes in order to identify sex-linked microsatellites. Comparisons to sex chromosomes were based on both chicken and zebra finch Z chromosomes, while only the chicken W chromosome was used due to the lack of currently available zebra finch W chromosome sequences. Identification of homologous loci across species was based on alignment of sequences used for primer design in Geneious. Alignments with a minimum of two species were used to evaluate microsatellite polymorphism.

## Results and Discussion

### Mitogenomes

Here we show that UCE sequencing may be a low-cost (~ US\$ 100 per sample) and efficient method for obtaining high-depth complete mitogenomes based solely on off-target sequences,

with no need for mtDNA-specific baits. The three samples sequenced here provided complete mitogenomes along with at least 625 out of 634 target loci per sample (Table 1). Mitogenome sequencing depth was high in all cases (Table 1), with means higher than 500x when PCR duplicates were considered. Removal of the PCR duplicates considerably dropped depth, but mean depths remained high (>70x), with depth ranging from 2x to 122x. Use of skeletal muscle as source material, which has a high mtDNA/nDNA ratio, may be responsible for such high mean depths [25,50], and further tests will be important to evaluate the efficiency of this protocol with other types of tissue. Blood samples, for example, appear to be a poor source of mtDNA off-target in UCE experiments (B. Faircloth, personal communication). Our results in particular showed PCR duplication levels higher than 90% for all three species, and those values may be considerably lowered by decreasing the number of PCR cycles during library amplification (B. Faircloth, personal communication).

Mitogenome completeness of the three warbling finches is supported by the presence of all regions found in two close mitogenomes available (*Thraupis episcopus*, [51]; *Emberiza pusilla*, [52]), including one control region, two rRNAs, 22 tRNAs and 13 protein-coding genes (S1 Table). Gene synteny was observed among the three species (S1 Table), and their gene order is the one present in most oscines studied so far. Genome sizes and G+C contents were also similar to those of published avian mitogenomes, ranging between 16,768–16,773 bp and 46.6–47.1%, respectively. Five mitochondrial polymorphic sites were found (*P. lateralis*, T and C at position 7847 and A and T at position 15290; *P. cabanisi*, C and A at position 15600; *P. thoracica*, A and T at position 14, and T and C at position 15606), and they can be explained by PCR error or heteroplasmy, which is an increasingly common finding in mtDNA studies [53–55]. Although an early diverging numt could also explain these polymorphic sites [56], we expect that use of muscle as DNA source makes this hypothesis unlikely due to its high mtDNA/nDNA ratio. Bayesian phylogenetic inference based on cytochrome b sequences also supported authenticity of the mitochondrial sequences with maximum posterior probabilities (S1 Fig). *Poospiza lateralis* was sister to *P. cabanisi*, a finding in line with their conspecificity according to some authors (see [57] for a taxonomic review). In addition, both sequences of *P. cabanisi* and *P. thoracica* collected here were sister to conspecific sequences obtained from Genbank.

## Microsatellite loci

We found a minimum of 361 microsatellites per species (Table 1), with primer design being possible for at least 64 loci per species (S1 Table). These loci numbers are higher than those recommended by Gardner *et al.* [58] for typical molecular ecology studies (~40 designable, unique loci). We believe that the number of microsatellites for which primers cannot be designed may decrease with use of inserts larger than the ones used here (~300 bp), which could translate into higher average assembly length and consequently longer flanking regions for primer design. We found microsatellite loci in both UCE and non-UCE data. Purifying selection may be a concern in UCE regions [59], which may include linked microsatellite loci. However, less than one fifth of the loci with flanking sequence suitable for primer design found here matched UCEs (*P. cabanisi*: 11 of 64; *P. lateralis*: 11 of 77; *P. thoracica*: 10 of 68). No mitochondrial or W-linked microsatellite loci were found, but a minor proportion (eight in *P. cabanisi*, eight in *P. lateralis*, and nine in *P. thoracica*) were Z-linked. From the microsatellite loci alignable between at least two species (23), 10 matched UCE loci and 13 consisted of non-UCE loci, with one of each category being variable. Although the small panel of individuals (2–3) may not be sufficient to reliably explore polymorphism, most of the sequences found were outside UCE regions, and thus should present levels of variation similar to those of microsatellite loci isolated using standard techniques (e.g. microsatellite enrichment).



## Advantages and potential pitfalls

Obtaining mitogenomes and microsatellite loci from off-target sequences is an opportunistic approach that depends on the incomplete efficiency of sequence capture protocols. Thus, future increases in sequence capture efficiency can negatively affect any application based on off-target sequences. However, despite the rising efficiency in sequence capture, off-target sequences are still common (e.g. protocols with only 40% on target in some cases, [26]), providing abundant mtDNA sequences and microsatellite loci. Even in cases of high specificity, increasing sequencing depth may still provide sufficient off-target sequences for diverse applications at reasonable costs. When the intended number of microsatellite loci or the desired mitogenome completeness is not met, the low cost associated with UCE sequencing (~ US\$ 100/sample) makes replicated sequencing of the same sample or sequencing of multiple individuals of a given species still a cost-effective strategy. Use of mixed mtDNA and nuclear baits (as suggested by Falk et al. [60]) is certainly an exciting and promising alternative to simultaneous sequencing of mtDNA and UCEs, but may also have its downsides, such as the additional costs of design and synthesis of compatible baits at different concentrations. In addition, designing baits for fast-evolving mitogenomes may be challenging when sequencing distant species, which is a common requirement of phylogenetic studies. Finally, use of off-target sequences may facilitate locus design of non-model organisms, since they contain non-UCE microsatellite markers.

## Conclusion

Here we show that the increasingly common sequencing of UCE loci may be a low-cost and simultaneous source of mitogenomes and sequences for microsatellite primer design with good sequencing depth. Since UCE sequencing has been successfully used in many non-model species of amniotes—and more recently, fishes and arthropods—the mitogenomes and microsatellite loci can be obtained for thousands of animal species using the strategy presented here, including species with limited genomic resources available.

## Supporting Information

**S1 Table. Characterization of the mitogenomes of *Poospiza cabanisi*, *P. lateralis* and *P. thoracica*.**

(XLS)

**S2 Table. List of microsatellite with designable primers.**

(XLS)

**S1 Fig. Bayesian phylogeny of *Poospizinae* thraupids based on cytochrome b sequences.**

Bold names with asterisks indicate sequences obtained here. Numbers on branches correspond to posterior probabilities.

(TIF)

## Acknowledgments

Special thanks to Brant Faircloth for his extensive support with UCE bioinformatics, as well as for sharing unpublished scripts and lab protocols. Michael Harvey and Brian Smith kindly shared unpublished information on UCEs. Editor Gen Hua Yue, Brant Faircloth, Andrew DeWoody and anonymous reviewers provided excellent comments in earlier versions of this manuscript. Jim Hesson (Academic English Solutions) revised the English. Thanks for the staff at P. N. de São Joaquim and owners of private properties for support and access to the land

under their care. C. Assis, B. Genevcius, G. Macedo, J. Montesanti, R. Quagliano, C. F. Schwertner and G. Thom provided excellent field assistance. This work was developed in the Research Center on Biodiversity and Computing (BioComp) of the Universidade de São Paulo (USP), supported by the USP Provost's Office for Research.

## Author Contributions

Conceived and designed the experiments: FRA CB. Performed the experiments: FRA LGN MFRR. Analyzed the data: FRA FM CB. Contributed reagents/materials/analysis tools: FRA CYM KCMP. Wrote the paper: FRA FM LGN MFRR CYM KCMP CB.

## References

1. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Reviews of Ecology and Systematics*. 1987; 18: 489–522.
2. Estoup A, Presa P, Krieg F, Vaiman D, Guyomard R. (CT)<sub>n</sub> and (GT)<sub>n</sub> microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity*. 1993; 71: 488–496. PMID: [8276633](#)
3. Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution*. 1999; 12: 105–114. PMID: [10381314](#)
4. Armour JA, Neumann R, Gobert S, Jeffreys AJ. Isolation of human simple repeat loci by hybridization selection. *Human Molecular Genetics*. 1994; 3: 599–565. PMID: [8069306](#)
5. Brown WM, Prager EM, Wang A, Wilson AC. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *Journal of Molecular Evolution*. 1982. pp. 225–239. PMID: [6284948](#)
6. Jennings TN, Knaus BJ, Mullins TD, Haig SM, Cronn RC. Multiplexed microsatellite recovery using massively parallel sequencing. *Molecular Ecology Resources*. 2011; 11: 1060–1067. doi: [10.1111/j.1755-0998.2011.03033.x](#) PMID: [21676207](#)
7. Jex AR, Hall RS, Littlewood DTJ, Gasser RB. An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes. *Nucleic Acids Research*. 2010; 38: 522–533. doi: [10.1093/nar/gkp883](#) PMID: [19892826](#)
8. Edwards SV, Bensch S. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*. 2009. pp. 2930–2933. doi: [10.1111/j.1365-294X.2009.04270.x](#) PMID: [19552688](#)
9. Edwards SV, Beerli P. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*. 2000; 54: 1839–1854. PMID: [11209764](#)
10. Koskinen MT, Hirvonen H, Landry PA, Primmer CR. The benefits of increasing the number of microsatellites utilized in genetic population studies: An empirical perspective. *Heredity*. 2004; 141: 61–67. PMID: [15383073](#)
11. Knowles LL. Statistical phylogeography. *Annual Reviews in Ecology, Evolution and Systematics*. 2009; 40: 593–612.
12. Amaral FR, Albers PK, Edwards S V, Miyaki CY. Multilocus tests of Pleistocene refugia and ancient divergence in a pair of Atlantic Forest antbirds (Myrmeciza). *Molecular Ecology*. 2013; 22: 3996–4013. doi: [10.1111/mec.12361](#) PMID: [23786305](#)
13. Etter PD, Preston JL, Bassham S, Cresko W, Johnson E. Local de novo assembly of RAD paired-end contigs using short sequencing reads. *PLoS One*. 2011; 6: e18561. doi: [10.1371/journal.pone.0018561](#) PMID: [21541009](#)
14. McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*. 2013; 66: 526–538. doi: [10.1016/j.ympev.2011.12.007](#) PMID: [22197804](#)
15. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, et al. Ultraconserved Elements in the Human Genome. *Science*. 2004. 304: 1321–1325. PMID: [15131266](#)
16. Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*. 2012; 61: 717–726. doi: [10.1093/sysbio/sys004](#) PMID: [22232343](#)
17. Crawford NG, Faircloth BC, McCormack JE, Brumfield RT, Winker K, Glenn TC. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters*. 2012; 8: 783–786. doi: [10.1098/rsbl.2012.0331](#) PMID: [22593086](#)

18. Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*. 2014; 63: 83–95. doi: [10.1093/sysbio/syt061](https://doi.org/10.1093/sysbio/syt061) PMID: [24021724](https://pubmed.ncbi.nlm.nih.gov/24021724/)
19. Faircloth BC, Sorenson L, Santini F, Alfaro ME. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). *PLoS One*. 2013; 8: e65923. PMID: [23824177](https://pubmed.ncbi.nlm.nih.gov/23824177/)
20. Faircloth BC, Branstetter MG, White ND, Brady SG. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources* 2015; 15: 489–501 doi: [10.1111/1755-0998.12328](https://doi.org/10.1111/1755-0998.12328) PMID: [25207863](https://pubmed.ncbi.nlm.nih.gov/25207863/)
21. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. 2010; 329: 75–78. doi: [10.1126/science.1190371](https://doi.org/10.1126/science.1190371) PMID: [20595611](https://pubmed.ncbi.nlm.nih.gov/20595611/)
22. Chilamakuri CSR, Lorenz S, Madoui M- A, Vodák D, Sun J, Hovig E, et al. Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics*. 2014; 15: 449. doi: [10.1186/1471-2164-15-449](https://doi.org/10.1186/1471-2164-15-449) PMID: [24912484](https://pubmed.ncbi.nlm.nih.gov/24912484/)
23. Miller MJ, Aguilar C, De León LF, Loaiza JR, McMillan WO. Complete mitochondrial genomes of the New World jacanas: *Jacana spinosa* and *Jacana jacana*. *Mitochondrial DNA*; 2014; 1–2.
24. Meiklejohn KA, Danielson MJ, Faircloth BC, Glenn TC, Braun EL, Kimball RT. Incongruence among different mitochondrial regions: a case study using complete mitogenomes. *Molecular Phylogenetics and Evolution*. 2014; 78: 314–23. doi: [10.1016/j.ympev.2014.06.003](https://doi.org/10.1016/j.ympev.2014.06.003) PMID: [24929245](https://pubmed.ncbi.nlm.nih.gov/24929245/)
25. Picardi E, Pesole G. Mitochondrial genomes gleaned from human whole-exome sequencing. *Nature Methods*. 2012. pp. 523–524. doi: [10.1038/nmeth.2029](https://doi.org/10.1038/nmeth.2029) PMID: [22669646](https://pubmed.ncbi.nlm.nih.gov/22669646/)
26. Guo Y, Li J, Li C-I, Shyr Y, Samuels DC. MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics*. 2013; 29: 1210–1211. doi: [10.1093/bioinformatics/btt118](https://doi.org/10.1093/bioinformatics/btt118) PMID: [23471301](https://pubmed.ncbi.nlm.nih.gov/23471301/)
27. Samuels DC, Han L, Li J, Quanguo S, Clark TA, Shyr Y, et al. Finding the lost treasures in exome sequencing data. *Trends in Genetics*; 2013; 29: 593–599. doi: [10.1016/j.tig.2013.07.006](https://doi.org/10.1016/j.tig.2013.07.006) PMID: [23972387](https://pubmed.ncbi.nlm.nih.gov/23972387/)
28. Burns KJ, Shultz AJ, Title PO, Mason NA, Barker FK, Klicka J, et al. Phylogenetics and diversification of tanagers (Passeriformes: Thraupidae), the largest radiation of Neotropical songbirds. *Molecular Phylogenetics and Evolution*; 2014; 75: 41–77. doi: [10.1016/j.ympev.2014.02.006](https://doi.org/10.1016/j.ympev.2014.02.006) PMID: [24583021](https://pubmed.ncbi.nlm.nih.gov/24583021/)
29. Amaral FR, Piacentini VQ, Brito GRR, Curcio FF. A simple and effective air shotgun for collecting small birds and other vertebrates at close range. *Journal of Field Ornithology*. 2012; 83: 403–406.
30. Andrews S. FastQC: A quality control tool for high throughput sequence data. 2014. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
31. Faircloth BC. Illumiprocessor: a Trimmomatic wrapper for parallel adapter and quality trimming. <https://github.com/faircloth-lab/illumiprocessor>. 2014.
32. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. In press. *Bioinformatics*.
33. Faircloth BC. Phyluce: phylogenetic estimation from ultraconserved elements. <https://github.com/faircloth-lab/phyluce>. 2014
34. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. *Nature* 2011; 29: 644–652.
35. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*; 1997:3389–3402. PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
36. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999; 41: 95–98.
37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012; 357–359. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
39. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv:12073907. 2014;
40. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842. doi: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/)



41. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*. 2013; 69: 313–319. doi: [10.1016/j.ympev.2012.08.023](https://doi.org/10.1016/j.ympev.2012.08.023) PMID: [22982435](https://pubmed.ncbi.nlm.nih.gov/22982435/)
42. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20: 3252–3255. PMID: [15180927](https://pubmed.ncbi.nlm.nih.gov/15180927/)
43. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28: 1647–1649. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199) PMID: [22543367](https://pubmed.ncbi.nlm.nih.gov/22543367/)
44. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004; 5: 113. PMID: [15318951](https://pubmed.ncbi.nlm.nih.gov/15318951/)
45. Nylander JA. MrModeltest v2. Program distributed by the author. 2004.
46. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003; 19: 1572–1574. PMID: [12912839](https://pubmed.ncbi.nlm.nih.gov/12912839/)
47. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop, GCE 2010.
48. Megléc E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, et al. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*. 2010; 26: 403–404. Available: doi: [10.1093/bioinformatics/btp670](https://doi.org/10.1093/bioinformatics/btp670) PMID: [20007741](https://pubmed.ncbi.nlm.nih.gov/20007741/)
49. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. *Nucleic Acids Research*. 2012; 40: e115. PMID: [22730293](https://pubmed.ncbi.nlm.nih.gov/22730293/)
50. Griffin HR, Pyle A, Blakely EL, Alston CL, Duff J, Hudson G, et al. Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genetics in Medicine*. 2014; 16: 962–971. doi: [10.1038/gim.2014.66](https://doi.org/10.1038/gim.2014.66) PMID: [24901348](https://pubmed.ncbi.nlm.nih.gov/24901348/)
51. Lerner HRL, Meyer M, James HF, Hofreiter M, Fleischer RC. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Current Biology*. 2011; 21: 1838–1844. doi: [10.1016/j.cub.2011.09.039](https://doi.org/10.1016/j.cub.2011.09.039) PMID: [22018543](https://pubmed.ncbi.nlm.nih.gov/22018543/)
52. Pan T, Wang H, Zhu X, Zhou W, Hou Y, Yu L, et al. Mitochondrial genome of the *Emberiza pusilla* (Emberizidae: Emberiza). *Mitochondrial DNA*. 2013; 24: 382–384. doi: [10.3109/19401736.2013.763238](https://doi.org/10.3109/19401736.2013.763238) PMID: [23391161](https://pubmed.ncbi.nlm.nih.gov/23391161/)
53. Kmiec B, Woloszyńska M, Janska H. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics*. 2006:149–159. PMID: [16763846](https://pubmed.ncbi.nlm.nih.gov/16763846/)
54. Vollmer NL, Viricel A, Wilcox L, Katherine Moore M, Rosel PE. The occurrence of mtDNA heteroplasmy in multiple cetacean species. *Current Genetics*. 2011; 57: 115–131. doi: [10.1007/s00294-010-0331-1](https://doi.org/10.1007/s00294-010-0331-1) PMID: [21234756](https://pubmed.ncbi.nlm.nih.gov/21234756/)
55. Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, et al. Universal heteroplasmy of human mitochondrial DNA. *Human Molecular Genetics*. 2013; 22: 384–390. doi: [10.1093/hmg/dds435](https://doi.org/10.1093/hmg/dds435) PMID: [23077218](https://pubmed.ncbi.nlm.nih.gov/23077218/)
56. Sorenson MD, Quinn TW. Numts: a challenge for avian systematics and population biology. *Auk*. 1998; 115: 214–221.
57. Assis CP, Raposo MA, Parrini R. Validação de *Poospiza cabanisi* Bonaparte, 1850 (Passeriformes: Emberizidae). *Revista Brasileira de Ornitologia*. 2007; 15: 103–112.
58. 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: 1093–1101. doi: [10.1111/j.1755-0998.2011.03037.x](https://doi.org/10.1111/j.1755-0998.2011.03037.x) PMID: [21679314](https://pubmed.ncbi.nlm.nih.gov/21679314/)
59. Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT. Sequence capture versus restriction site associated DNA sequencing for phylogeography. *arXiv Prepr*. 2013; 1–53.
60. Falk MJ, Pierce Ea, Consugar M, Xie MH, Guadalupe M, Hardy O, et al. Mitochondrial disease genetic diagnostics: optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. *Discovery Medicine*. 2012; 14: 389–99. PMID: [23272691](https://pubmed.ncbi.nlm.nih.gov/23272691/)