

A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species

Marie-Ka Tilak, Fabienne Justy, Mélanie Debiais-Thibaud, Fidel Botero-Castro, Frédéric Delsuc, Emmanuel Douzery

► To cite this version:

Marie-Ka Tilak, Fabienne Justy, Mélanie Debiais-Thibaud, Fidel Botero-Castro, Frédéric Delsuc, et al.. A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. Conservation Genetics Resources, Springer, 2015, 7, pp.37 - 40. <hal-01925895>

HAL Id: hal-01925895

<https://hal.archives-ouvertes.fr/hal-01925895>

Submitted on 29 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A COST-EFFECTIVE STRAIGHTFORWARD PROTOCOL FOR SHOTGUN ILLUMINA LIBRARIES DESIGNED TO ASSEMBLE COMPLETE MITOGENOMES FROM NON-MODEL SPECIES

Marie-ka Tilak ^(*,*), Fabienne Justy ^(*), Mélanie Debiais-Thibaud, Fidel Botero-Castro, Frédéric Delsuc, Emmanuel J. P. Douzery ^(1,2)

⁽¹⁾ Université Montpellier 2

⁽²⁾ CNRS & IRD, Institut des Sciences de l'Evolution (ISE-M, UMR 5554)

CC064 - Place Eugène Bataillon

34 095 Montpellier Cedex 05 — France

^(*) MT and FJ contributed equally to this work

^(*) Corresponding author

KEYWORDS

Mitochondrial DNA — Next Generation Sequencing — Degraded DNA — Non-model species — Chiroptera — Tunicata.

ABSTRACT

The mitogenome is an inescapable tool in conservation biology studies. Yet, its routine sequencing may remain tricky despite next-generation sequencing (NGS) technologies. An enrichment step is often necessary but not always straightforward depending on the initial DNA quality or quantity. Furthermore, the availability of close mitochondrial DNA reference sequences for non-model species limits the primer design for long-range PCR or bait synthesis. Here we propose an easy and cost-effective protocol without enrichment step for building and sequencing multiplexed Illumina libraries from small quantities of either high-quality or degraded genomic DNA. We validated the approach through the successful assembly of the complete mitogenome of 60 bats and 7 tunicates. Our protocol allows the sequencing and assembly of mitochondrial genomes from non-model species with sufficient coverage for applications in conservation genetics.

MAIN TEXT

Complete mitogenomes play a central role in phylogenetics (Janke et al., 2002), phylogeography (Morin et al., 2010), ancient DNA (Cooper et al., 2001) and conservation biology (Bagatharia et al., 2013) studies. Recent technological advances, especially Illumina sequencing, provide unprecedented amounts of genetic data but often require prior enrichment steps — long-range PCR based capture probes (Maricic et al., 2010) or custom capture probe arrays (Hancock-Hanser et al., 2013). Because of the variable preservation level of biological samples, protocols to obtain mitogenomes from low quality and / or quantity DNA extracts are required. Yet, enrichment steps for non-model species are not always feasible due to their critical reliance on the availability of reference sequences. Using standard laboratory equipment, we propose a cost-effective and straightforward protocol adapted from Meyer & Kircher (2010) to prepare shotgun Illumina libraries from genomic DNA. This allowed pooling up to 48 DNA libraries from animal taxa for single-read sequencing on an Illumina HiSeq 2000 lane, and assembling the corresponding mitogenomes.

We used modern samples preserved in 95% ethanol or DMSO/salt of 60 bats (liver, heart, kidney or muscles) and 24 tunicates (gonads or whole individuals). DNA extractions with negative controls were performed with the DNAeasy Blood and Tissue kit (QIAGEN) following manufacturer's instructions, except decreasing the elution volume to 100 μ l. DNA quality varied markedly among chiropterans : 57% of the samples yielded high molecular weight DNA, 9% led to partially degraded DNA extracts, while 34% had too low concentration to be visualized on an agarose gel (Online Resource 1). Prior to library preparation, PCR amplification, Sanger sequencing and phylogenetic tree reconstruction of a barcoding fragment of the mitochondrial 12S rRNA gene (Online Resource 2), showed that 14% of the bat species were mislabelled or misidentified (see also Shen et al. 2014).

Total DNA was sheared for 20 min using an ultrasonic cleaning unit (Elmasonic one) whose frequency was 35 kHz. Conditions were set to get the best level of reproducibility (Online Resource 3). Sheared genomic DNA was sized and concentrated by adding 1.7 volume of SPRI bead suspensions (Agencourt® AMPure® XP) per volume of sample and eluted in 25 μ l of ultra-pure water. For chiropteran samples, the amount of sheared and sized DNA ranged from 60 ng to 1.4 μ g with 34% being less than 120 ng, 48% between 120 ng and 700 ng, and 18% between 700 ng and 1.4 μ g.

We followed the Illumina library preparation procedure with blunt-end repair, adapter ligation, adapter fill-in and indexing PCR steps developed by Meyer & Kircher (2010), with slight modifications aiming at decreasing the overall cost. Since DNA templates were two-fold concentrated during the sizing step, the amounts of reagents and suspension beads for purification were halved. During the adapter ligation step, we substituted 4 μ l of PEG-4000 (50%) with 4 μ l of ultra-pure water and incubated at 16°C overnight. Before the PCR indexing step, we performed a purification adding 1.7 volume of Agencourt® AMPure® XP reagent to the sample to remove adapter dimers, with a 24 μ l of ultra-pure water elution. We also substituted the library characterization step described in Meyer & Kircher (2010) by the quantification of DNA libraries with a Nanodrop ND-800 spectrophotometer (Nanodrop technologies). All steps were conducted with dedicated material, equipment, and laboratory space to reduce the risk of library contamination.

Libraries were PCR-indexed (14 cycles) with primers 1 to 24 according to Meyer & Kircher (2010). After SPRI bead suspension purification, indexed libraries were quantified with Nanodrop ND-800, and pooled using their relative concentrations to ensure equimolarity. We generated three pools of indexed libraries including respectively 15, 21 and 48 species samples. In the latter case, we multiplexed 24 chiropteran and 24 tunicate samples with the same 24 indexes, thanks to the high mitogenomic divergence of these two groups (Rubinstein et al. 2013). Each pool of indexed libraries was independently single-read sequenced on one lane of Illumina HiSeq 2000 at GATC-Biotech (Konstanz, Germany).

For the 60 chiropterans, 0.03% to 4% of the total read number per species corresponded to mitochondrial sequences. We successfully assembled all complete mitogenomes following the bioinformatics pipeline of Botero-Castro et al. (2013), with a combination of *de novo* assembly using ABySS (Simpson et al., 2009) and, when possible, read mapping on a phylogenetically related mitogenome using Geneious Pro (Drummond et al., 2011). The average nucleotide coverage decreased with increasing number of multiplexed species per sequencing lane but it remained sufficient for bats (see Table1). We did not observe any significant correlation between the number of reads obtained and the amount of DNA used for each library construction (Fig. 1a), nor between the mitogenome coverage and the total number of reads (Fig. 1b). However, we found a significant negative correlation between either the percentage of mitochondrial reads obtained (Fig. 1c) or the mitogenome coverage (Fig. 1d) and the amount of DNA used. These results suggest that (i) the ratio of mitochondrial to nuclear DNA is more favourable to the former at lower amounts of DNA used, and (ii) complete mitogenomes can be assembled with reasonable coverage even when indexed Illumina libraries are built from poor quality and / or low quantity of the initial DNA extract.

For the 24 tunicates, 0.02% to 0.07% of the total read number per species corresponded to mitochondrial sequences. The overall mitogenome coverage was much lower compared to bats despite similar initial DNA quality and quantity. Only 7 mitogenomes were successfully assembled, with two of them published in Griggio et al. (2014), and the remaining 17 samples yielded only partial mitochondrial contigs. These results suggest a superimposition of biological factors and bioinformatics issues: (i) a lower ratio of mitochondria versus nuclei in tunicate tissues reduces the number of available mitochondrial reads, (ii) the rapid rate of mitogenome evolution hampers read mapping on closely related reference sequences, and (iii) shuffled gene orders among urochordates preclude the assembly of contigs into supercontigs (Rubinstein et al. 2013).

In conclusion, we validated a cost-effective protocol allowing the efficient assembly of numerous complete mitogenomes for non-model species, starting from heterogeneous DNA extracts without prior enrichment. Overall costs can be adjusted by choosing different multiplexing (e.g., single versus double indexing) and sequencing (single versus paired-end) strategies. Our procedure could potentially be useful for conservation studies requiring DNA isolation from museum specimens or non-invasive sampling of endangered species.

ACKNOWLEDGMENTS

This work was supported by the Centre National de la Recherche Scientifique (CNRS), the Scientific Council of Université Montpellier 2, and the Agence Nationale de la Recherche "ANR Programme Blanc" (Tunicate Evo-Devo "TED", ANR-13-BSV2-0011). We benefited from the technical facilities of the "Centre Méditerranéen de l'Environnement et de la Biodiversité" (Labex CeMEB) and the "Plateau technique : construction de banques d'ADN pour technique Illumina".

For bat and tunicate tissues, we are indebted to the following institutions, curators and collectors: the Ambrose Monell Cryo Collection (AMCC) at the American Museum of Natural History (AMNH, New York: Neil Duncan), the Museum of Southwestern Biology - University of New Mexico (Joseph A. Cook), the Museum of Vertebrate Zoology - University of California (MVZ, Berkeley: Christopher J. Conroy), the National Museum of Natural History – Smithsonian Institution (Darrin P. Lunde and Kristofer M. Helgen), the Museum of Texas Tech University (Robert J. Baker), François Catzeflis (ISE-M), and Xavier Turon (Center for Advanced Studies of Blanes, CSIC, Spain). We also thank Hendrik Poinar for discussion about next generation sequencing, and two anonymous referees for comments on the manuscript. This publication is contribution ISEM 2014-152 of the Institut des Sciences de l'Evolution de Montpellier.

REFERENCES

- Bagatharia SB, Joshi MN, Pandya RV, Pandit AS, Patel RP, Desai SM, Sharma A, Panchal O, Jasmani FP, Saxena AK (2013) Complete mitogenome of asiatic lion resolves phylogenetic status within *Panthera*. *BMC Genomics*, 14, 572.
- Botero-Castro F, Tilak M, Justy F, Catzeflis F, Delsuc F, Douzery EJP (2013) Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol Phylogenet Evol*, 69, 728-739.
- Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, 409, 704-707.
- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2011) Geneious v6. Available from <http://www.geneious.com/>.
- Griggio F, Voskoboynik A, Iannelli F, Justy F, Tilak M, Turon X, Pesole G, Douzery EJP, Mastrototaro F, Gissi C (2014) Ascidian mitogenomics: comparison of evolutionary rates in closely related taxa provides evidence of ongoing speciation events. *Genome Biol Evol*, 6, 591-605.
- 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. *Mol Ecol Resources*, 13, 254-268.
- Janke A, Magnell O, Wiczorek G, Westerman M, Arnason U (2002) Phylogenetic analysis of 18S rRNA and the mitochondrial genomes of the wombat, *Vombatus ursinus*, and the spiny

anteater, *Tachyglossus aculeatus*: increased support for the Marsupionta hypothesis. *J Mol Evol*, 54, 71-80.

Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, 5, e14004.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*, 6, 1-10.

Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Nielsen SCA, Rasmussen M, Willerslev E, Gilbert MTP, Harkins T (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res*, 20, 908-916.

Rubinstein ND, Feldstein T, Shenkar N, Botero-Castro F, Griggio F, Mastrototaro F, Delsuc F, Douzery EJP, Gissi C, Huchon D (2013) Deep sequencing of mixed total DNA without barcodes allows efficient assembly of highly plastic ascidian mitochondrial genomes. *Genome Biol Evol*, 5, 1185-1199.

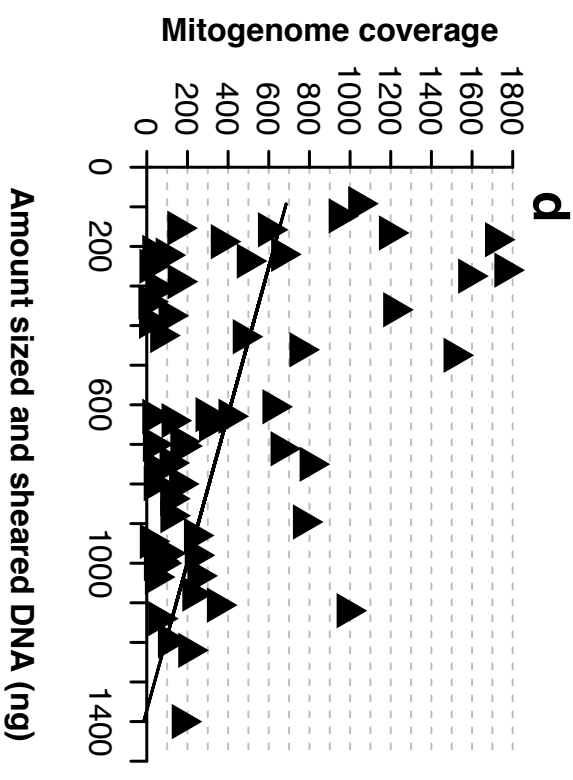
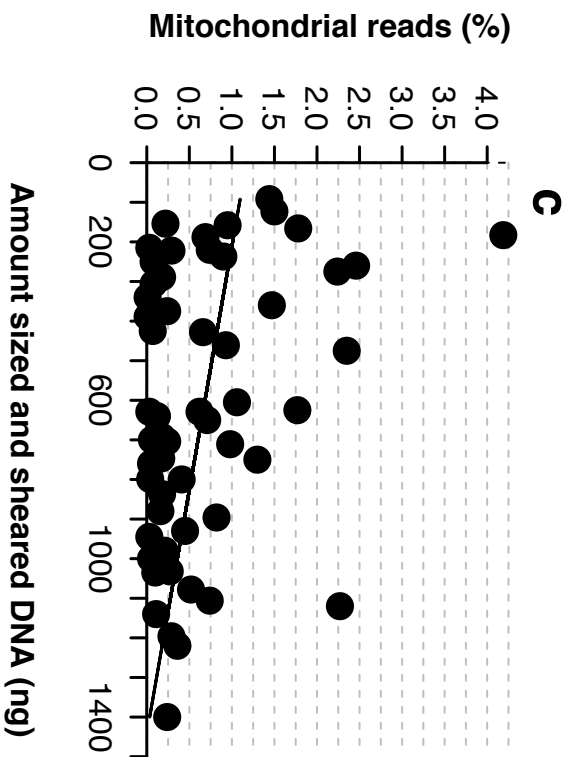
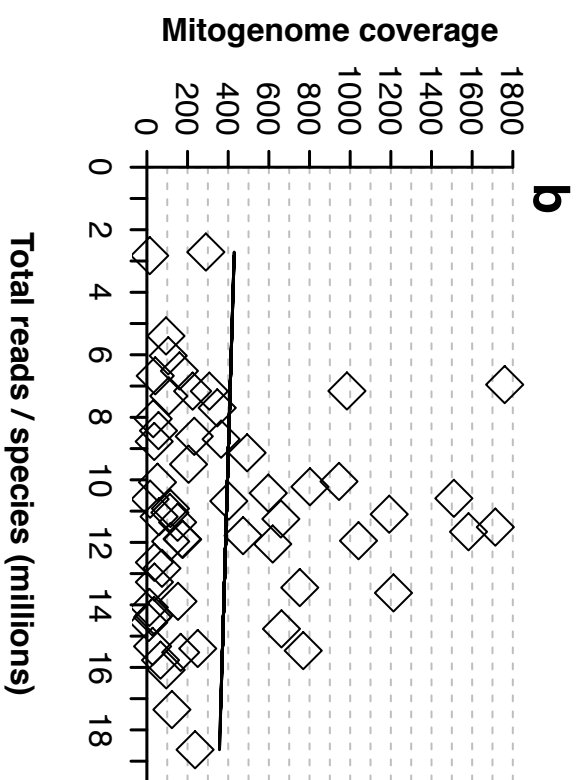
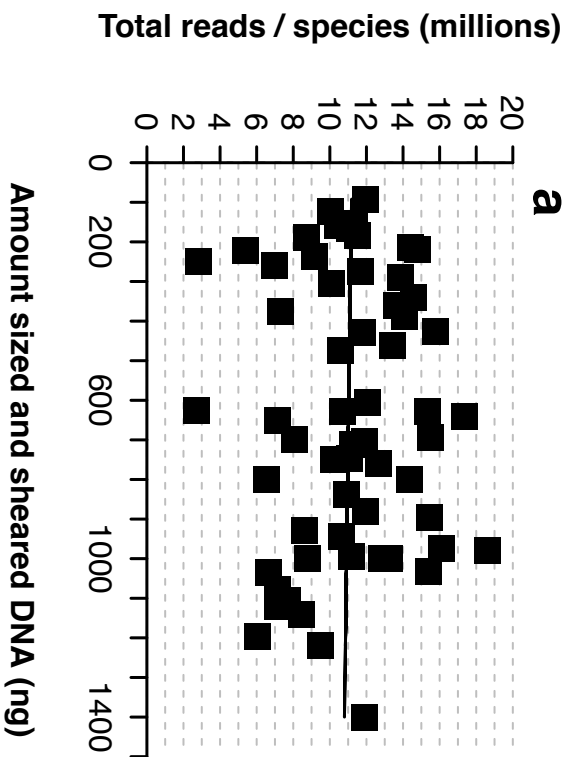
Shen Y-Y, Chen X, Murphy RW (2013) Assessing DNA barcoding as a tool for species identification and data quality control. *PLoS ONE*, 8, e57125.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) ABySS: A parallel assembler for short read sequence data. *Genom Res*, 19, 1117-1123

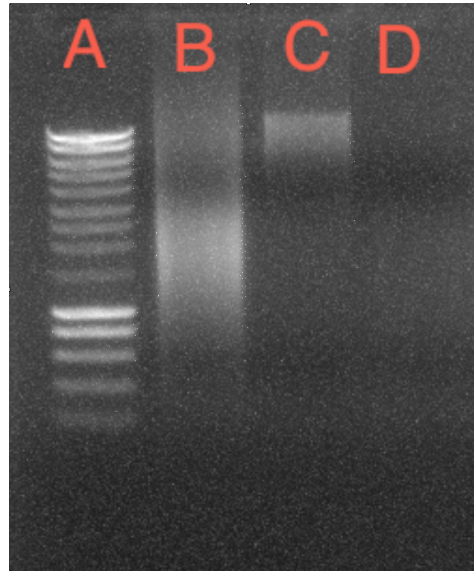
Fig. 1: Correlation plots based on 60 bat samples between (a) total million reads per species and initial amount (in ng) of sized and sheared DNA ($R^2=0.00076$, $P\text{-val}=0.84$), **(b)** average mitogenome nucleotide coverage and total million reads per species ($R^2=0.0019$, $P\text{-val}=0.74$), **(c)** percent of mitochondrial reads with respect to initial amount of sized and sheared DNA ($R^2=0.12$, $P\text{-val}<0.01$), and **(d)** average mitogenome nucleotide coverage and initial amount of sized and sheared DNA ($R^2=0.14$, $P\text{-val}<0.01$).

Table 1: Statistics on the mitogenome assemblies for three pools of 15, 21 and 48 libraries single reads sequenced on one of Illumina HiSeq 2000.

Taxa	Pool	Species per lane	Number of indexes	Complete mitogenomes	Median coverage	Minimum coverage	Maximum coverage
Chiroptera	#1	15	15	15	639	28	1760
Chiroptera	#2	21	21	21	233	14	1582
Chiroptera + Tunicata	#3	48	24	31	—	—	—
	Chiroptera	#3	24	24	116	9	1213
	Tunicata	#3	24	7	22	14	25



ONLINE RESOURCE 1



Agarose gel electrophoresis of 5 μ l of bat DNA extracts, with ethidium bromide staining.

- A: 5 μ l of 100 bp ladder (Eurogentec) ;
- B: Partially degraded DNA extract ;
- C: DNA extract with high molecular weight ;
- D: DNA extract with low concentration.

ONLINE RESOURCE 2

12S ribosomal RNA (rRNA) PCR primers and amplification conditions used for the barcoding step.

	Primer sequence (5' → 3')	Annealing temperature
S2	TCTTCTGGGTGTAGGCCAGATGCTTT	50°C
2084	CAAACCTGGGATTAGATACCCCACTAT	50°C

12S rRNA PCR amplification consisted of an initial denaturation step of 95°C for 3 min, followed by 30 cycles of 95°C for 20 sec, 50°C for 20 sec, and 72°C for 30 sec, with a final extension of 72°C for 10 min.

PCR products were then purified with magnetic beads (Agencourt® AMPure® XP) following the manufacturer's instructions.

Purified PCR amplicons were sequenced in both directions with a BigDye® Terminator v3.1 kit (Applied Biosystems) on an Applied ABI Prism 3130 XL sequencer.

ONLINE RESOURCE 3

DNA fragmentation

We transferred 40-60 μ l of genomic DNA (less than 100 ng to 1 μ g) in 0.2 ML tubes (Thermo scientific AB-0620).

A 96-position rack was pasted onto an Ultrasonic Cleaning Unit (Elmasonic one) tank filled with 250 ml of 4°C sterilized water.

We have positioned 8 tubes in the middle of the tank. After 20 min in one stage, eight genomic DNA extracts were sheared down in the same time between 100 to 400 bp.

To check post-fragmentation fragment sizes, 5 μ l of the sheared DNA were electrophoresed for 40 min at 100 volts in 2% agarose gel stained with ethidium bromide and using a 100 bp ladder (Eurogentec).

The sheared DNAs sized and concentrated were quantified on a Nanodrop ND-800 spectrophotometer (Nanodrop technologies)