TECHNICAL NOTE

# 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**

**Abstract** The mitogenome is an inescapable tool in conservation biology studies. Yet, its routine sequencing may remain tricky despite next-generation sequencing 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.

---

Marie-ka Tilak and Fabienne Justy have contributed equally to this work.

M.-K. Tilak (✉) · F. Justy · M. Debiais-Thibaud ·
F. Botero-Castro · F. Delsuc · E. J. P. Douzery
CNRS & IRD, Institut des Sciences de l'Evolution
(ISE-M, UMR 5554), Université Montpellier 2, CC064 - Place
Eugène Bataillon, 34 095 Montpellier Cedex 05, France
e-mail: marie-ka.tilak@univ-montp2.fr

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 and 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 μl. 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. 2013).

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 μl of ultra-pure water. For chiropteran samples, the amount of sheared and sized DNA ranged from 60 ng to 1.4 μg with 34 % being less than 120 ng, 48 % between 120 ng and 700 ng, and 18 % between 700 ng and 1.4 μg.

We followed the Illumina library preparation procedure with blunt-end repair, adapter ligation, adapter fill-in and indexing PCR steps developed by Meyer and Kircher (2010), with slight modifications aiming at decreasing the overall cost. Since DNA templates were twofold concentrated during the sizing step, the amounts of reagents and suspension beads for purification were halved. During the adapter ligation step, we substituted 4 μl of PEG-4000 (50 %) with 4 μl 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 μl of ultra-pure water elution. We also substituted the library characterization step described in Meyer and 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 1–24 according to Meyer and 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, 0.03–4 % 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 Table 1). 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 (1) the ratio of mitochondrial to nuclear DNA is more favourable to the former at lower amounts of DNA used, and (2) 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, 0.02–0.07 % 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: (1) a lower ratio of mitochondria versus nuclei in tunicate tissues reduces the number of available mitochondrial reads, (2) the rapid rate of mitogenome evolution hampers read mapping on closely related reference sequences, and (3)

**Table 1** Statistics on the mitogenome assemblies for three pools of 15, 21 and 48 libraries single-end sequenced on one lane of the Illumina HiSeq 2000

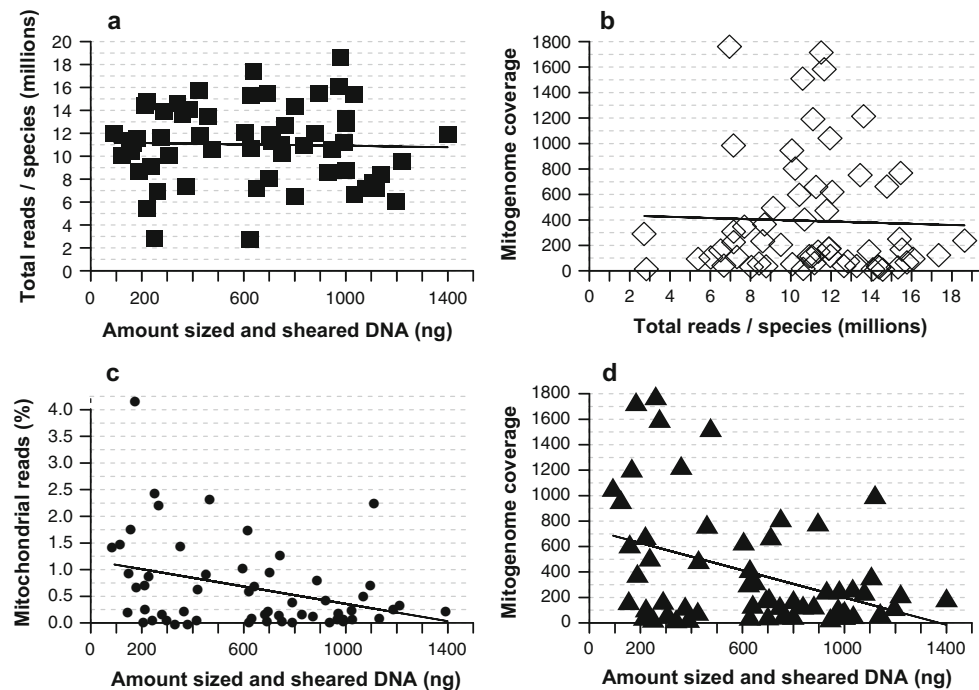| Taxa | Pool | Species per lane | Number of indexes | Complete mitogenomes | Median coverage | Minimum coverage | Maximum coverage |
|---|---|---|---|---|---|---|---|
| Chiroptera | #1 | 15 | 15 | 15 | 639 | 28 | 1,760 |
| Chiroptera | #2 | 21 | 21 | 21 | 233 | 14 | 1,582 |
| Chiroptera + Tunicata | #3 | 48 | 24 | 31 | – | – | – |
| Chiroptera | #3 | 24 | 24 | 24 | 116 | 9 | 1,213 |
| Tunicata | #3 | 24 | 24 | 7 | 22 | 14 | 25 |

**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 ($R^2 = 0.00076$, $P$ value = 0.84), **b** average mitogenome nucleotide coverage and total million reads per species ($R^2 = 0.0019$, $P$ value 0.74), **c** percent of mitochondrial reads with respect to initial amount of sized and sheared DNA ($R^2 = 0.12$, $P$ value $< 0.01$), and **d** average mitogenome nucleotide coverage and initial amount of sized and sheared DNA ($R^2 = 0.14$, $P$ value $< 0.01$)

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.

# 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 Genom 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. 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. Genom 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 Resour 13:254–268

Janke A, Magnell O, Wieczorek 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. Genom 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. Genom 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