TECHNICAL NOTE

Low cost sequencing of mitogenomes from museum samples using baits capture and Ion Torrent

**Spyros Kollias · Marloes Poortvliet ·
Irina Smolina · Galice Hoarau**

Abstract The development of various target enrichment methods in combination with next generation sequencing techniques has greatly facilitated the use of partially degraded DNA samples in genetic studies. We employed the MYbaits target enrichment system in combination with Ion Torrent sequencing on a broad range of DNA quality, extracted from tissues obtained from both natural history archives and through various opportunistic sampling methods, to sequence the mitogenome of 11 mobulid rays and two closely related species. Mobulids are large, elusive pelagic filter feeders, for which conservation concerns have recently be raised in connection to their vulnerable life histories and increasing fishing pressure. We show that the MYbaits target enrichment method can be used to effectively sequence large parts of the mitogenome from heavily degraded DNA samples, and provide a time and cost effective alternative for genetic studies of rare and/or difficult to sample species.

Keywords Next generation sequencing · Targeted sequencing · MYbaits · Mitogenome · *Manta* · *Mobula*

S. Kollias · M. Poortvliet · I. Smolina · G. Hoarau (✉)
Faculty of Biosciences and Aquaculture, University of Nordland,
Universitetsalleen 11, 8049 Bodø, Norway
e-mail: gah@uin.no

M. Poortvliet
Department of Marine Benthic Ecology and Evolution, Centre
for Ecological and Evolutionary Studies, University of
Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

M. Poortvliet
Department of Ecology and Evolutionary Biology, University of
California Santa Cruz, 100 Shaffer Road, Santa Cruz, CA 95060,
USA

Genetic studies within rare or difficult to access taxonomic groups are often limited by the availability of samples. In such cases, specimens archived in natural history collections can, in principle, provide an easy accessible source of DNA. The use of DNA extracted from such specimens is however not without challenges. Indeed, tissues can degrade over time, resulting in highly fragmented DNA, which is often inappropriate for downstream applications (Wandeler et al. 2007). The development of next generation sequencing (NGS) in combination with enrichment methods has allowed sequencing both from ancient (Reich et al. 2010; Summerer 2009) and museum specimens (Mason et al. 2011). Here, we explore the performance of the MYbaits target enrichment system (Mycroarray.com) and NGS techniques to sequence mitogenomes, using DNA of mobulid rays and two closely related species (Table 1).

Mobulids are large pelagic filter-feeding elasmobranchs, for which conservation concerns have recently been raised due to their vulnerable life histories and ongoing fishing pressure (Couturier et al. 2012; Dulvy et al. 2014). However, despite concerns for mobulids, the availability of genetic data is limited, mostly due to the inaccessibility of their habitat and logistic issues with sample collection. We use a broad range of DNA quality, extracted from tissues we managed to obtained from both natural history archives and through various opportunistic sampling methods (Table 1).

Total DNA extraction for all samples was performed using DNeasy Blood and Tissue Kit (Qiagen). All DNA samples were quantified using Qubit dsDNA HS Assay Kit (Life Technologies). Nine of the thirteen DNA samples (Table 1) were fragmented using S2 ultrasonicator (Covaris) to a median fragment size of $\sim$260 bp. The four remaining DNA samples (Table 1) were already heavily fragmented with a median fragment size of $\sim$200 bp, and

Table 1 Information about species, type of tissue and preservative, collector, DNA quality after extraction, total number of sequenced reads, the percentage of reads on target, percentage coverage of the mitogenome with coverage of >8 reads, and average coverage depth

Sample no.	Species	Tissue and preservative	Collected by	DNA quality	Total number of reads	% reads on target	% coverage with >8 reads	Average coverage depth
1	*Manta birostris*	Muscle in EtOH	Felipe Galván Magaña	Some degradation (smear from 12 Kb–200 b)	125,784	12.2	91	148
2	*Mobula tarapacana*	Muscle in EtOH	Inter American Tropical Tuna Commission (IATTC)	Some degradation (smear from 12 Kb–200 b)	223,150	2.7	88	53
3	*M. kuhlii*	Tail in EtOH	Marloes Poortvliet	Some degradation (smear from 12 Kb–200 b)	86,638	19	92	155
4	*M. kuhlii*	Muscle in EtOH	Sabine Wintner; KwaZulu-Natal Sharks Board	High quality	372,887	7	99	143
5	*M. eregoodootenkee*	Muscle in EtOH	Sabine Wintner; KwaZulu-Natal Sharks Board	Some degradation (smear from 12 Kb–200 b)	355,728	3.8	91	123
6	*M. thurstoni*	Muscle in EtOH	Daniel Fernando; Guy Stevens	Some degradation (smear from 12 Kb–200 b)	206,831	18.9	93	361
7	*M. rochebrunei*	Skin and muscle, dried specimen	Agnes Dettaï, MNHN Paris	Degraded, ∼200 b	56,382	1.8	86	9
8	*M. hypostoma*	Skin and muscle, dried specimen	Agnes Dettaï, MNHN Paris	Degraded, ∼200 b	–	–	–	–
9	*M. munkiana*	Tail in EtOH	Monterey Bay Aquarium	High quality	211,226	15.4	92	299
10	*M. mobular*	Muscle, frozen and dried	Giuseppe Notarbartolo di Sciara; Fabrizio Serena	Degraded, ∼200 b	480,015	19.2	98	833
11	*M. mobular*	Tissue in formalin	Smithsonian Institute	Degraded, ∼200 b	–	–	–	–
12	*Myliobatis californica*	Tail in EtOH	Daniel Fernando; Poortvliet	High quality	185,352	6.4	84	113
13	*Rhinoptera steindachneri*	Muscle in EtOH	Marloes Poortvliet	Some degradation (smear from 12 Kb–200 b)	310,108	2.3	89	66
	Average	–	–	–	237,646	9.9	91	209

Coverage date have been calculated for the mitogenome excluding the Control Region (15,500 bp)

were therefore used without any additional fragmentation. The size distribution of all DNA samples was assessed on a 2100 Bioanalyzer using HS DNA Kit (Agilent).

We used ∼200 ng of total DNA in 50 µl of 10 mM Tris–HCl as starting material for each library. Library preparation was conducted using the AB Library Builder System with the Ion Plus Library Kit in combination with the Ion Xpress Barcode Adapters (Life Technologies).

Libraries were then amplified for five cycles, and purified using the Agencourt AMPureXP Kit (Beckman Coulter). The purified libraries were quantified using the Ion Library Quantitation Kit (Life Technologies), pooled in equimolar amounts and concentrated using the PureLink PCR Purification Kit (Life Technologies). The library pool was eluted using ultra pure water in two combined elutions of 10 µl each. The concentrated library pool was quantified

using the Qubit dsDNA HS Assay Kit (Life Technologies), 500 ng of pooled libraries were dried using a vacuum concentrator (Eppendorf) and the pellet was resuspended in 3.4 µl ultra pure water.

Sequence enrichment for targeted sequencing was achieved using a customizable liquid-phase DNA capture system under the commercial name MYbaits (MYcroarray.com). More specifically, we used the Mybaits1 system, which contains a custom library of 20.000 biotinylated 120mer single stranded DNA baits designed against a specific reference sequence. As reference sequence for the current study we used the *Mobula japanica* complete mitochondrial genome (NC_018784), which is 18.880 bp long (Poortvliet and Hoarau 2013). Two custom oligonucleotide blocking-probes were used to prevent the cross-hybridization between Ion Torrent adapters during the hybridization step (BlockingAbc: ATCIIIIIIIIICT-GAGTCGGAGACACGCAGGGATGAGATGG 3′-PHO and BlockingP1: ATCACCGACTGCCCATAGA-GAGGAAAGCGGAGGCGTAGTGG). Hybridization was performed for 36 h at 65 °C. Recovery of the captured targets with MyOne Streptavidin C1 magnetic beads (Life Technologies) was followed by elution and cleanup of the enriched library pool. Post capture amplification for ten cycles was performed using the Library Amplification Primer Mix supplied in the Ion Plus Fragment Library Adapters Kit (Life Technologies). The enriched amplified library pool was purified using the PureLink PCR Purification Kit (Life Technologies), quantified and diluted appropriately.

Template preparation of the diluted library pool was performed with an IonOneTouch2 Instrument using the Ion PGM Template OT2 200 Kit (Life Technologies). Sequencing of templated spheres was conducted using Ion PGM200 Sequencing Kit and an Ion316 Chip on an IonTorrent Personal Genome Machine (PGM) System (Life Technologies),following the manufacturer instructions.

Sequence reads belonging to each barcoded library were mapped to the reference genome of *M. japanica* in CLC Genomics Workbench v.6 (CLC bio, Denmark) using default mapping parameters (Mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.5, similarity fraction 0.8). The mitochondrial Control Region and flanking regions were covered by very few reads in all species, due to the presence of long AT-rich tandem repeat regions (Poortvliet and Hoarau 2013), which are known to lead to coverage bias on the Ion Torrent sequencing platform (Ross et al. 2013). Most of the remaining 15.500 bp of the mitogenome (84 % or more) was successfully sequenced for 11 of the 13 species, including two out of four heavily degraded samples (for which PCR amplifications of short mtDNA fragments were unsuccessful). Sequencing of the remaining two degraded samples produced no barcoded

reads. Although the percentage of on-target reads was overal quite low (1.8–19.2 %); the sequencing run resulted in adequate coverage (average coverage depth = 9–833 reads) of large part of the mitogenomes (coverage = 84–99 %) of the 11 species (Table 1). The percentage of mtDNA read was 1–2 order of magnitude higher than what was obtained by direct sequencing without targeted enrichment. Optimization of the hybridization step could results in improved enrichment efficiency, and increased coverage. However, the efficiency of hybridization will be taxa specific, depending on the phylogenetic distance between the taxa use to design the baits and the target but also on features of the mitogenomes such as GC contents and gene rearrangements. Furthermore, given the low cost of NGS, the percentage of reads on target should be adequate for most research.

Our study demonstrates that MYbaits targeted capture system in combination with NGS can be used to sequence large parts of the mitochondrial genome from highly degraded samples. Moreover, the described method is a time and cost-effective way of sequencing mitogenomes in non-model organisms: the experimental procedures can be completed within 2 weeks, and the total cost of all described procedures was around 2500 USD, <200 USD per sample.We conclude that our approach can be a useful tool for researchers studying rare or difficult to sample species, by making samples from for example natural history archives accesable for genetic studies.

References

Couturier LI, Marshall AD, Jaine FR, Kashiwagi T, Pierce SJ, Townsend KA, Weeks SJ, Bennett MB, Richardson AJ (2012) Biology, ecology and conservation of the Mobulidae. J Fish Biol 80:1075–1119

Dulvy NK, Fowler SL, Musick JA, Cavanagh RD, Kyne PM, Harrison LR, Carlson JK, Davidson LN, Fordham SV, Francis MP, Pollock CM, Simpfendorfer CA, Burgess GH, Carpenter KE, Compagno LJ, Ebert DA, Gibson C, Heupel MR, Livingstone SR, Sanciangco JC, Stevens JD, Valenti S, White WT (2014) Extinction risk and conservation of the world's sharks and rays. eLife 3:e00590

Mason VC, Li G, Helgen KM, Murphy WJ (2011) Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. Genome Res 21:1695–1704

Poortvliet M, Hoarau G (2013) The complete mitochondrial genome of the Spinetail Devilray, *Mobula japanica*. Mitochondrial DNA 24:28–30

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Paabo S (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468:1053–1060

Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB (2013) Characterizing and measuring bias in sequence data. Genome Biol 14:R51

Summerer D (2009) Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. Genomics 94:363–368

Wandeler P, Hoeck PE, Keller LF (2007) Back to the future: museum specimens in population genetics. Trends Ecol Evol 22:634–642