Gene 573 (2015) 171–175

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short communication

The use of transcriptomic next-generation sequencing data to assemble mitochondrial genomes of *Ancistrus* spp. (Loricariidae)☆

Daniel A. Moreira [a], Carolina Furtado [b], Thiago E. Parente [a,*]

[a] Laboratório de Toxicologia Ambiental, DCB, Escola Nacional de Saúde Pública — ENSP, Fundação Oswaldo Cruz — FIOCRUZ, Rio de Janeiro, Brasil
[b] Divisão de Genética, Instituto Nacional do Cancer — INCA, Rio de Janeiro, Brasil

ARTICLE INFO

Article history:
Received 28 April 2015
Received in revised form 20 August 2015
Accepted 28 August 2015
Available online 3 September 2015

Keywords:
Mitogenome
Mitochondrial transcripts
Transcriptome
RNA
Illumina
Siluriformes
Biodiversity

ABSTRACT

Mitochondrial genes and genomes have long been applied in phylogenetics. Current protocols to sequence mitochondrial genomes rely almost exclusively on long range PCR or on the direct sequencing. While long range PCR includes unnecessary biases, the purification of mtDNA for direct sequencing is not straightforward. We used total RNA extracted from liver and Illumina HiSeq technology to sequence mitochondrial transcripts from three fish (*Ancistrus* spp.) and assemble their mitogenomes. Based on the mtDNA sequence of a close related species, we estimate to have sequenced 92%, 95% and 99% of the mitogenomes. Taken the sequences together, we sequenced all the 13 protein-coding genes, two ribosomal RNAs, 22 tRNAs and the D-loop known in vertebrate mitogenomes. The use of transcriptomic data allowed the observation of the punctuation pattern of mtRNA maturation, to analyze the transcriptional profile, and to detect heteroplasmic sites. The assembly of mtDNA from transcriptomic data is complementary to other approaches and overcomes some limitations of traditional strategies for sequencing mitogenomes. Moreover, this approach is faster than traditional methods and allows a clear identification of genes, in particular for tRNAs and rRNAs.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The many biological functions of mitochondria and the uses of mitochondrial genes have created a huge interest in sequencing mitogenomes (Huang, 2011). The initial approach to sequence mtDNA was to isolate and fragment it with restriction enzymes, clone the fragments in bacterial plasmids and expand the fragment-plasmid construct in bacteria to obtain enough DNA to be sequenced by the Sanger method (Anderson et al., 1981). Although many different approaches have been developed to isolate mtDNA, it is still a challenge to obtain this molecule with enough quantity, quality and purity for sequencing purposes (Jayaprakash et al., 2015). The sequencing of mitogenomes was made a lot easier by the establishment of the Polymerase Chain Reaction (PCR) as a powerful molecular biology technique. Using long range PCR, overlapping regions of mtDNA could be amplified and sequenced, as before, by the Sanger method (Hu et al., 2002). The complete mtDNA molecule could be assembled by joining the overlapping regions of each PCR amplicon (Hu et al., 2007).

First attempts to use mitochondrial transcripts to sequence mitogenomes were based on Sanger sequencing of expressed sequence tags (ESTs) (Gissi and Pesole, 2003; Samuels et al., 2005). Mitochondrial genomes are transcribed as a single polycistronic transcript, which has the coding sequence of all the 13 intron-less proteins, 22 tRNAs and the two subunits of the ribosomal RNA coded by most metazoan mitogenomes (Bernt et al., 2013a; Friedman and Nunnari, 2014). After transcription, the polycistronic transcript is cleaved on "punctuation marks", giving origin to mature mitochondrial tRNA, rRNA and mRNA (Ojala et al., 1981). Mature mitochondrial mRNAs are monocistronic or bicistronic and polyadenilated (Bernt et al., 2013a).

More recently, Next-Generation Sequencing (NGS) technologies are being used to study heteroplasmy in humans (Huang, 2011; Jayaprakash et al., 2015) and to sequence mitogenomes from model and non-model organisms (Besnard et al., 2014; Dames et al., 2015; Hahn et al., 2013; Jex et al., 2010), including museum specimens (Fabre et al., 2013). However, most descriptions of mitogenomes by NGS uses mtDNA originated from long range PCR or directly isolated from animal tissues (Dames et al., 2015; Payne et al., 2015; Quispe-Tintaya et al., 2015), and hence, the challenge to isolate this molecule with enough quantity, quality and purity, along with the biases introduced by PCR reactions, remain largely unchanged.

Here, we used mitochondrial RNA (mtRNA), rather than mtDNA, to sequence mitogenomes. The recent advances on RNA-Seq allowed the generation of more than 40 millions 100 bp paired-end Illumina HiSeq2500 reads per sample, using as start material total RNA extracted

Abbreviations: mtDNA, mitochondrial DNA; rRNA, ribosomal RNA; tRNA, transfer RNA; mRNA, messenger RNA; PCR, Polymerase Chain Reaction; ESTs, expressed sequence tags; NGS, Next-generation Sequencing; CDS, coding sequence.

☆ The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

* Corresponding author.
E-mail address: parente@ensp.fiocruz.br (T.E. Parente).

http://dx.doi.org/10.1016/j.gene.2015.08.059
0378-1119/© 2015 Elsevier B.V. All rights reserved.

from liver by the simple phenol:chloroform extraction. Based on transcriptomic NGS data, the mitogenomes of three *Ancistrus* spp. individuals were assembled. The use of mtRNA to sequence mitogenomes is straightforward and overcomes some of the current challenges to isolate mtDNA and the biases introduced by PCR. More importantly, the use of transcriptomic data enables the sequencing of mitogenomes, while still sequence the complete coding region of thousands of nuclear protein-coding genes, and is a valuable approach in face of the fast increasing number of transcriptomic researches using non-model species.

2. Material and methods

2.1. RNA extraction

RNA samples were extracted from the liver of three individual fish of the genus *Ancistrus*, promptly preserved in RNA later and kept at −20°C. The fish were deposited at the Museu Nacional do Rio de Janeiro under the voucher number MNRJ42890. RNA extractions were performed using either TRIzol (Invitrogen) or TRI Reagent (Life technologies) following manufacturer's instructions. After extraction, the RNA preparations were quantified using a BioDrop ulite spectrophotometer (Biodrop). RNA quality was evaluated using the kit RNA 6000 Nano for Bioanalyzer (Agilent).

2.2. Library preparation and sequencing

The complementary DNA (cDNA) libraries were prepared using 1000 ng of total RNA strictly following the instructions of the TrueSeq RNA Sample kit v2 (Illumina). Each of the three libraries was uniquely identified using specific barcodes. Quality of library preparations was accessed using the DNA 1000 kit for Bioanalyzer (Agilent). Libraries were quantified by qPCR using the Library quantification kit for Illumina (Kapa Biosystems). The three libraries were clustered, using the TrueSeq PE Cluster kit v3 for cBot (Illumina), in the same lane together with six other samples used in other projects. A 100 bp single-end and another 100 bp paired-end sequencing reactions were performed in a HiSeq2500 using the TrueSeq SBS kit v3 (Illumina). The library preparation and sequencing reaction were performed at the Divisão de Genética of the Instituto Nacional do Cancer (INCA) in Rio de Janeiro, Brazil.

2.3. Bioinformatic processing

Raw Illumina data were demultiplexed using the BCL2FASTQ software (Illumina). Reads were trimmed for Illumina adaptors by Trimmomatic (Bolger et al., 2014) and its quality were evaluated using FastQC (Babraham Bioinformatics). Only reads with PHRED score > 30 were used for the transcriptome assembly, which was performed using the concatenated reads from the single and paired-ends reactions and the default parameters of Trinity v. 2.0.2 (Grabherr et al., 2011; Haas et al., 2013). Although Trinity was used to assemble the transcriptomes, many *de novo* assemblers are available and should have similar performances.

Each transcriptome was subjected to BLASTX searches against two databases, the Uniprot entries of humans (*Homo sapiens*), and the Uniprot entries of zebrafish (*Danio rerio*) (Altschul et al., 1990; Consortium, 2015). An additional BLASTN was performed against the complete mitogenome of the closest related species, whose mitogenome is publically available, *Pterygoplichthys disjunctivus* (GI: 339506171) (Nakatani et al., 2011). The transcripts aligned with the reference mitogenome were used for mitogenomes assembly. Selected transcripts were edited according to the information of strand orientation given by the BLASTN result, and aligned by SeaView using the built-in CLUSTAL alignment algorithm to the reference mitogenome (Gouy et al., 2010). A CONTIG sequence was generated using the sequence information of just the transcripts of each individual fish. The CONTIG sequence was then manually checked for inconsistencies and gaps. The mitogenomes were annotated using the web-based services MitoFish and MITOS (Bernt et al., 2013b; Iwasaki et al., 2013). In order to estimate the support of each base of the mitogenomes, Bowtie v. 1.0.0 was used to align the reads of each fish on its own assembled mitogenome, and this mapping was viewed using the Integrated Genome Viewer (IGV) or the Tablet (Langmead et al., 2009; Milne et al., 2009; Robinson et al., 2011; Thorvaldsdóttir et al., 2013). Heteroplasmic sites were detected using IGV, setting the software to show positions in which the frequency of the second most frequent base was equal to or higher than 10% and the total reads number were higher than 100.

3. Results

3.1. Library construction, Illumina sequencing and transcriptome assemble

High quality RNA (RIN > 7.0) was extracted from the three fishes (Table 1). The size of the three libraries ranged from 230 to 800 base pairs (bp), and the total number of HiSeq reads from 43.5 to 60.1 million (Table 1). Each of the three transcriptomes is composed by more than 60 thousand transcripts (Table 1). In all cases, around 50% of the assembled transcripts have a BLASTX hit (Table 1).

3.2. Mitogenome assembling and annotation

In order to assemble the mitogenomes, 7 to 13 transcripts of each fish were used (Table 1 and see Table 1 in (Daniel et al., submitted for publication)). In comparison to the complete mitogenome of *Pterygoplichthys disjunctivus*, which is the closest relative of *Ancistrus* spp. with the complete mitogenome available, we sequenced 99.2% of the mtDNA of *Ancistrus* sp. #1, 92.5% of the mtDNA of *Ancistrus* sp. #2a, and 94.7% of the mtDNA of *Ancistrus* sp. #2b (Table 1). The mitogenomes are deposited in GenBank under the GIs: KP960569, KP960568, KP960567, respectively. *Ancistrus* sp. #2a and *Ancistrus* sp. #2b are from the same species, as they share 99.8% identical nucleotides at their mitogenomes. *Ancistrus* sp. #1 is probably a different species, as its mitogenome differs from the other two fish on 6% of the nucleotide positions.

In terms of features number, the mitogenome of *Ancistrus* sp. #2b is the most complete of three *Ancistrus* spp. Having a single gap, which prevented its circularization, we sequenced the 13 protein-coding genes, the two ribosomal RNAs and 21 of the 22 tRNAs in *Ancistrus* sp. #2b (Fig. 1). Its single gap contains the sequence of the displacement loop (D-loop) and the $tRNA_{phe}$. In the mitogenome of *Ancistrus* sp. #1 just three tRNAs are missing: $tRNA_{leu2}$, $tRNA_{his}$, $tRNA_{ser1}$ (Fig. 1). The mtDNA of *Ancistrus* sp. #2a has the lowest coverage, missing the D-loop, and four tRNAs; $tRNA_{phe}$, $tRNA_{leu2}$, $tRNA_{pro}$, $tRNA_{thr}$ (Fig. 1). Supplemental information about the mitogenome features are provided elsewhere (see Table 3 (Daniel et al., submitted for publication)).

Interestingly, the number of supporting reads of each nucleotide varied greatly according to the position in the mitogenome (Fig. 1). More than 84,000 reads were found to support the sequences of

Table 1
Summary of the transcriptome and mitogenome data for the three fish (*Ancistrus* spp.).

	Ancistrus sp. #1	*Ancistrus* sp. #2a	*Ancistrus* sp. #2b
RNA Integrity Number — RIN	8.2	7.4	>7.00
Library insert size (bp)	230–800	268–792	285–370
Reads after QC	43,502,597	53,961,751	60,170,745
Transcripts			
Total	67,098	63,847	67,883
With BLASTX hit	35,710	31,886	33,953
For mitogenome	13	12	7
mtRNA reads (%)	2.6	1.8	0.8
Mitogenome coverage (%)	99.2	92.5	94.7
Heteroplasmic sites	10	8	6

Download English Version:

https://daneshyari.com/en/article/2815246

Download Persian Version:

https://daneshyari.com/article/2815246

Daneshyari.com