Gene Reports 5 (2016) 10–17

Contents lists available at ScienceDirect

Gene Reports

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

Complete mitochondrial genomes are not necessarily more informative than individual mitochondrial genes to recover a well-established annelid phylogeny

Victor Corrêa Seixas [a], Paulo Cesar Paiva [b], Claudia Augusta de Moraes Russo [a,*]

[a] Universidade Federal do Rio de Janeiro, CCS, Instituto de Biologia, Departamento de Genética, Av. Carlos Chagas Filho 373 – Sala A2-97 – Bloco A – Ilha do Fundão, Rio de Janeiro, RJ, 21941-570, Brazil
[b] Universidade Federal do Rio de Janeiro, CCS, Instituto de Biologia, Departamento de Zoologia, Av. Carlos Chagas Filho 373 – Sala A0-108 – Bloco A – Ilha do Fundão, Rio de Janeiro, RJ, 21941-902, Brazil

ARTICLE INFO

Article history:
Received 1 July 2016
Received in revised form 21 July 2016
Accepted 29 July 2016
Available online 05 August 2016

Keywords:
Mitogenome
Phylogenetics
Efficiency of molecular markers
Next generation sequencing
Invertebrates

ABSTRACT

Here, we describe the complete mitogenome of the polychaete *Laeonereis culveri*, and we use this genome along with eight other annelid species to test the performance of mitochondrial markers and tree-building methods to recover a reliable phylogeny of Annelida. We also compared the efficiency of tree-building methods (Neighbour-joining, maximum likelihood and Bayesian inference) and models. A total of 420 phylogenies were reconstructed and revealed that amino acids greatly outperformed the nucleotide data set. Bayesian inference was more susceptible to type I error, due to high rates of incorrect clades with high statistical support. Conversely, maximum likelihood and Neighbour-joining were more prone to type II error, as many accepted clades were poorly supported. Our results clearly support the idea that the choice of the method and model was not as important as the choice of the gene. Six genes recovered the reference topology at least once. *COB* had the best performance followed by *CO1*, *CO3*, *ND1*, *ND6* and *ND4L*. Even for these genes, the support of the deepest clades was low in many cases. The concatenation of genes only improved the performance for maximum likelihood. These results showed that mitochondrial genes must be selected very carefully, mainly when used for deeper relationships.
© 2016 Published by Elsevier Inc.

1. Introduction

Phylogenetic relationships among major vertebrate groups are, in general, more reliable than those in protostomes. The reasons for this include a more complete genomic sampling for vertebrates and the fact that morphological-based phylogenies are less reliable for invertebrates due to the scarcity of comparable characteristics (Dunn et al., 2014).

As a consequence, well-established vertebrate phylogenies have been useful to ascertain the performance of mitochondrial and nuclear genes in recovering well established topologies, that is, topologies that are very consistent when analysed by various types of data (Russo et al., 1996; Meiklejohn et al., 2014). For vertebrates, it has been suggested that the *ND5* gene tends to outperform other mitochondrial genes, but the *ND4*, *ND2*, *CO1* and *COB* genes are also informative in phylogenetic reconstructions using vertebrate sequences (Russo et al., 1996; Zardoya and Meyer, 1996; Miya and Nishida, 2000; Mueller, 2006; Meiklejohn et al., 2014).

On the other hand, for invertebrate groups, it remains unclear whether the genes most often used, such as *CO1* and *CO2*, are indeed the most informative genes for answering phylogenetic questions (Caterino et al., 2000; Halanych and Janosik, 2006). Even for invertebrates, it is possible to carefully select invertebrate groups for which the phylogenetic relationships are well established.

Therefore, in this paper, we carefully selected nine annelid species to assemble a well-established phylogeny to address the following questions. (1) How efficient are mitochondrial genes at recovering a well-established Annelida phylogeny? (2) Is it more effective to use a few carefully selected genes or is the use of more genes better? (3) Is gene performance associated with alignment statistics, such as gene length and variability levels? (4) Is gene performance affected by tree-building methods and substitution models? (5) Does gene performance vary for shallow and deeper nodes of the phylogeny? Our rationale is that, if we are able to clarify the aspects of mitochondrial markers in recovering this well-established phylogeny, metazoan phylogenetic studies would greatly benefit from these cheap and widely available mitochondrial markers.

2. Materials and Methods

2.1. Taxon sampling

There are approximately 50 complete mitochondrial genomes of Annelida available in public databases. To assess the efficiency of mitochondrial genes for recovering a well-established topology, we selected mitogenomes from species for which the phylogenetic relationship is

Abbreviations: BI, Bayesian inference; M, Maximum Likelihood; NJ, neighbor-joining; CO, subunit of cytochrome oxidase; Cytb, cytochrome b; NAD, NADH dehydrogenase.

* Corresponding author.
E-mail address: claudiaamrusso@gmail.com (C.A.M. Russo).

http://dx.doi.org/10.1016/j.genrep.2016.07.011
2452-0144/© 2016 Published by Elsevier Inc.

very well supported by morphological and molecular data (Rouse and Fauchald, 1997; Erséus and Källersjö, 2004; Rousset et al., 2007; Struck et al., 2007; Zrzavý et al., 2009; Weigert et al., 2014, 2016; Andrade et al., 2015; Laumer et al., 2015; Struck et al., 2015). The organisms selected were *Erpobdella octoculata* (Family: Erpobdellidae; GenBank: NC_023927), *Whitmania pigra* (Haemopidae; NC_013569), *Amynthas aspergillus* (Megascolecidae; NC_025292), *Lumbricus terrestris* (Lumbricidae; NC_001673), *Pista cristata* (Terebellidae; NC_011011), *Terebellides stroemii* (Trichobranchidae; NC_011014), *Marphysa sanguinea* (Eunicidae; NC_023124), *Perinereis nuntia* (Nereididae; NC_020609) and *Laeonereis culveri* (Nereididae; KU992689 – sequenced for this work). Our selection covers different taxonomic levels that represent six monophyletic groups within Annelida (Fig. 1).

2.2. Brief characterization of the complete mitochondrial genome of *Laeonereis culveri*

Laeonereis culveri genomic DNA was extracted using the DNeasy 96 Blood and Tissue kit (Qiagen Inc.), and sequences were generated using a Roche 454 platform (proceeded in Laboratório Nacional de Ciências da Computação - Brazil). Reads were first assembled with GS Data Analysis (Life Science, Roche) and used as references with MITObim software (Hahn et al., 2013). The final contig was inspected using the Bowtie2 (Langmead and Salzberg, 2012) and SAMtools (Li et al., 2009) packages, and the final annotation was performed in the MITOS webserver (Bernt et al., 2013a). Start and stop codons were manually checked.

The mitochondrial genome of *L. culveri* is 14,918 base pairs (bp) in length and comprises 2 rRNA, 22 tRNA, 13 protein-coding genes and 1 control region (Supp. Table S1). Two tRNA genes were identified for the leucine and serine amino acids. The control region is 459 bp in length, the 12S rRNA is 815 bp in length and the 16S rRNA is 1,007 bp in length. tRNA sequences varied in length from 56 to 68 bp, and protein-coding genes varied in length from 159 (*ATP8*) to 1,641 (*ND5*) bp.

All regions were found in the forward strand, except for that of the tRNA-Gly. Protein-coding genes using three start codons ATG (six protein-coding genes), ATT (four protein-coding genes) and ATA (three protein-coding genes). Stop codons are complete for all coding genes, and most of them end with a TAA codon, with the exception of *COII* and *COB* (which end with a TAG codon). The mitochondrial genome nucleotide composition was 38.3% for T (U), 35.5% for A, 14.4% for C and 11.8% for G. The A + T bias was 73.8%, while G + T was 26.2%. The gene order is in agreement with that observed in other nereidids, including *Platynereis dumerilii* (AF178678), *Perineris nuntia* (NC_020609) and *P. aibuhitensis* (NC_023943).

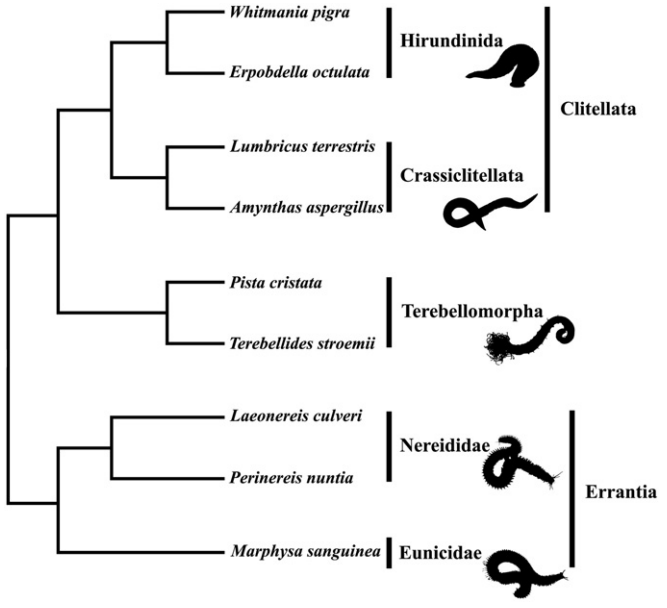

Fig. 1. Annelida phylogeny used to represent the reference topology to evaluate the efficiency of mitochondrial genes in constructing this phylogeny.

2.3. Phylogenetic analysis

The phylogenetic analyses were conducted separately with 13 mitochondrial protein-coding genes and a concatenated matrix using nucleotide (1) and amino acid (2) sequences. Amino acid sequences greatly outperformed nucleotide sequences; therefore, the latter were excluded from the following analysis (Supp. Figs. S1 and S2). Each gene sequence was translated into the amino acid sequence (invertebrate mitochondrial code) and aligned using Muscle 3.8 (Edgar, 2004). Single gene alignments were assembled in a concatenated matrix using FASconCAT (Kück and Meusemann, 2010).

The best substitution models for individual genes and the concatenated matrix were estimated in jModelTest 2 (Darriba et al., 2012) for nucleotides and in ProtTest 2.4 (Abascal et al., 2005) for amino acids, according to the Bayesian information criterion (BIC) (Supp. Table S2). Apart from the model test suggestion, we also included simpler models to diminish errors in parameter estimation. For amino acid sequences, *p*-distance and Poisson, JTT, Blosum62, MtZoa and GTR models were used.

Topologies were reconstructed for each gene alignment and for the concatenated matrix for three tree-building methods: Neighbour-joining (NJ), Maximum Likelihood (ML) and Bayesian Inference (BI). For the NJ tree, MEGA 5.2 was used to build the tree, and we evaluated the statistical confidence of the clades using the bootstrap test with 1,000 pseudo-replicates (Tamura et al., 2011). The ML tree for amino acid sequences was constructed using RAxML 8.2 (Stamatakis, 2014), and the best ML tree was selected from 20 independent runs. In this case, all selected substitution models included the Gamma distribution. A rapid bootstrap algorithm was implemented with 1,000 pseudo-replicates.

The BI was conducted in MrBayes 3.2 (Ronquist et al., 2012), sampling every 100 generations with a total of 500,000 (amino acid) or 1,000,000 (nucleotide) trees and a final burn in of 25% generations. Two independent runs and four chains were implemented. The final results were verified according to the standard deviations of split frequencies (<0.01) and the effective sampling size (ESS>200) via Tracer 1.5 as recommended (Rambaut and Drummond, 2007). The statistical test applied in BI phylogenies was the posterior probability (PP). RAxML 8.2 and MrBayes 3.2 analyses were conducted in CIPRES Science Gateway (Miller et al., 2010).

2.4. Distance metrics between topologies

Topological comparisons between trees were used to estimate the accuracy of each marker, model and tree-building method for recovery of the well-established Annelida phylogeny (considering different data sets, methods and models) using two metrics. The first metric uses the internal branches of each topology to define bipartitions of taxa sets. If topologies are perfectly bifurcating, final topological distance is roughly defined as twice the number of different bipartitions, (Penny and Hendy, 1985). This measure was calculated in the R package ape (Paradis et al., 2004) and the formula was as follows:

$$d_T = 2[Min(q_1, q_2) - p] + |q_1 - q_2|$$

where q_1 and q_2 are the number of partitions for trees 1 and 2, respectively, and p is the number of identical partitions between the topologies (Rhzetsky and Nei, 1992).

Download English Version:

https://daneshyari.com/en/article/2820460

Download Persian Version:

https://daneshyari.com/article/2820460

Daneshyari.com