Molecular Phylogenetics and Evolution 52 (2009) 268–272

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

MtZoa: A general mitochondrial amino acid substitutions model for animal evolutionary studies

Omar Rota-Stabelli, Ziheng Yang, Maximilian J. Telford *

Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 10 October 2008
Revised 22 December 2008
Accepted 23 January 2009
Available online 30 January 2009

Keywords:
Mitochondrial genome
Mitogenomics
Phylogeny
Metazoa
Model selection
Empirical
Mechanistic
Maximum likelihood
Bayesian inference

1. Introduction

Mitochondrial genome coded proteins are widely used as markers for the inference of phylogeny (mitogenomics). Their main advantages are unambiguous orthology, the richness of available sampling among eukaryotes and the relative ease of generating new data. On the other hand, mitochondrial sequences have been reported to suffer from an accelerated substitution rate and among-lineages compositional heterogeneity (Foster et al., 1997; Rota-Stabelli and Telford, 2008). These characteristics, if shared by phylogenetically unrelated species, may be responsible for convergent evolution (homoplasy) and promote the dilution of the true phylogenetic signal. Furthermore, the mitochondrial genetic code varies to different degrees between different metazoan lineages. In the light of this, mitogenomic studies are in need of realistic models of evolution that best represent the evolutionary process and reduce systematic bias.

The majority of deep level mitogenomic analyses are carried out at the amino acid level as nucleotide sequences are more susceptible to substitutional saturation and codon-based phylogenies may be complicated by differences in the genetic code. Amino acid datasets can be analyzed using the mechanistic GTR (general time reversible) model (Yang et al., 1998), allowing all the parameters of the model to be estimated from the dataset during the inference of phylogeny. A clear problem in this procedure is the large size of the amino acid alphabet, which makes the estimation of all the parameters a demanding computational task. Additionally, reliable estimation of the amino acid replacement rates needs a significant amount of substitutional information from the dataset and the small datasets typically used in phylogenetic analyses may not contain sufficient information. Consequently, amino acid alignments are generally analyzed using empirical amino acid replacement matrices, which have been pre-estimated from a large dataset and are represented in fixed matrices.

A current problem with existing empirical models is that they are based on the comparison of restricted datasets; MtREV (Adachi and Hasegawa, 1996) or MtMamm (Yang et al., 1998) are dominated by mammalian sequences and the recently released MtArt (Abascal et al., 2007) and MtPan (Carapelli et al., 2007) are both based on the analysis of arthropod-only datasets (Fig. 1). These matrices consequently reflect the substitution processes of either mammals or arthropods only and may be not appropriate for the analysis of other metazoan lineages, in particular lophotrochozoans and non-mammalian deuterostomes, for which many mitogenomic datasets are available, but few analyses have been conducted (Waeschenbach et al., 2006).

* Corresponding author. Fax: +44 2076797096.
E-mail address: m.telford@ucl.ac.uk (M.J. Telford).

1055-7903/$ - see front matter
doi:10.1016/j.ympev.2009.01.011

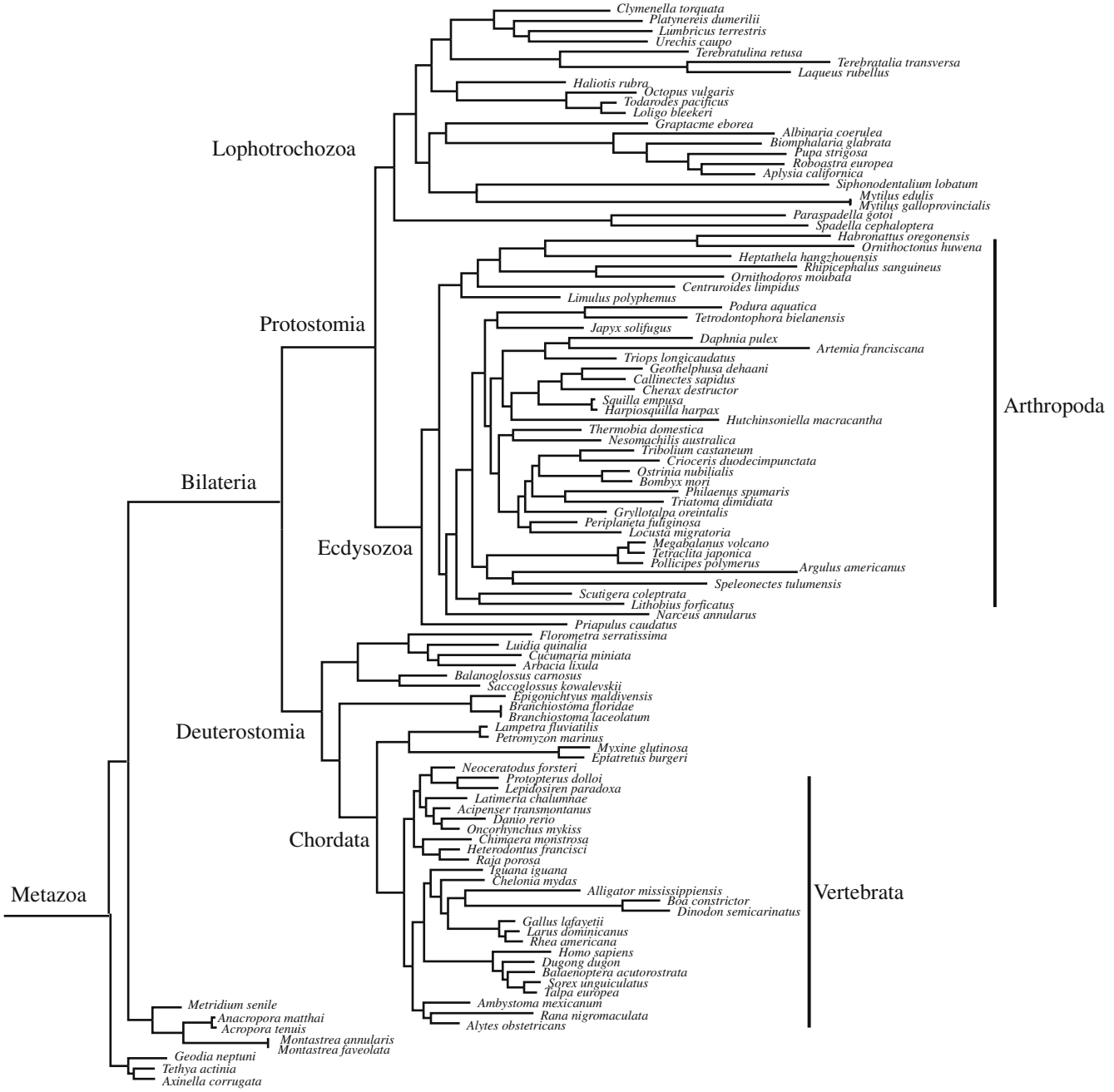

Fig. 1. Phylogenetic tree of the 108 metazoan species used to infer the MtZoa model. Note that commonly used empirical models such as MtREV (which is derived from vertebrates, indicated by vertical bar) and MtArt (derived from arthropods, indicated by vertical bar) are based on the comparison of restricted datasets. MtZoa is based on a larger and broader dataset, including lophotrochozoans, non-chordate deuterostomes and diploblastic metazoans. The topology was inferred using MrBayes under the MtREV model and some nodes have been constrained to reflect current knowledge of metazoan relationships; branch lengths have been estimated by PAML, during the inference of the MtZoa model.

1.1. Synopsis

In order to generate an empirical model that is more representative of the whole animal kingdom we estimated an empirical transition probability matrix, called MtZoa (Fig. 2A), based on the general reversible model and an alignment of 13 concatenated mitochondrial proteins from more than 100 phylogenetically diverse metazoan species. We tested how MtZoa and other models fit different metazoan datasets and show that our model is particularly indicated for the analysis of diverse metazoan, lophotrochozoan and deuterostome datasets.

2. Materials and methods

We assembled an alignment of the 13 mitochondrial proteins from 108 metazoan species, consisting of 39 deuterostomes, 22 lophotrochozoans 39 ecdysozoans and eight non-bilaterians. We constructed the corresponding tree (in Fig. 1) using MrBayes and the MtREV model and constraining some major nodes in order to reflect current knowledge of metazoan relationships and the so called "new animal phylogeny" (Webster et al., 2006; Telford et al., 2008; Dunn et al., 2008). We excluded lineages characterized by extremely accelerated substitution rate, such as urochordates