Molecular Phylogenetics and Evolution 66 (2013) 748–756

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Incorporating indels as phylogenetic characters: Impact for interfamilial relationships within Arctoidea (Mammalia: Carnivora)

Peng-tao Luan [a,b], Oliver A. Ryder [c], Heidi Davis [c], Ya-ping Zhang [a,d,*], Li Yu [a,b,*]

[a] Laboratory for Conservation and Utilization of Bio-Resources and Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming 650091, China
[b] Key Laboratory for Animal Genetic Diversity and Evolution of High Education in Yunnan Province, Yunnan University, Kunming, Yunnan 650091, China
[c] San Diego Zoo Institute for Conservation Research, P.O. Box 15600, San Diego, CA 92027, USA
[d] State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

ARTICLE INFO

Article history:
Received 21 March 2012
Revised 27 October 2012
Accepted 29 October 2012
Available online 10 November 2012

Keywords:
Insertion/deletion events
Arctoidea
Indel coding
Phylogenetic analyses

ABSTRACT

Insertion and deletion events (indels) provide a suite of markers with enormous potential for molecular phylogenetics. Using many more indel characters than those in previous studies, we here for the first time address the impact of indel inclusion on the phylogenetic inferences of Arctoidea (Mammalia: Carnivora). Based on 6843 indel characters from 22 nuclear intron loci of 16 species of Arctoidea, our analyses demonstrate that when the indels were not taken into consideration, the monophyly of Ursidae and Pinnipedia tree and the monophyly of Pinnipedia and Musteloidea tree were both recovered, whereas inclusion of indels by using three different indel coding schemes give identical phylogenetic tree topologies supporting the monophyly of Ursidae and Pinnipedia. Our work brings new perspectives on the previously controversial placements among Arctoidea families, and provides another example demonstrating the importance of identifying and incorporating indels in the phylogenetic analyses of introns. In addition, comparison of indel incorporation methods revealed that the three indel coding methods are all advantageous over treating indels as missing data, given that incorporating indels produces consistent results across methods. This is the first report of the impact of different indel coding schemes on phylogenetic reconstruction at the family level in Carnivora, which indicates that indels should be taken into account in the future phylogenetic analyses.

1. Introduction

Insertion and deletion events (indels) provide a suite of markers complementary to nucleotide substitutions with enormous potential for molecular phylogenetics, including lack of functional constraints, a high substitution rate and less homoplasy (Giribet and Wheeler, 1999; Rokas and Holland, 2000; Simmons et al., 2001; Dessimoz and Gil, 2010; Yu et al., 2011). However, in earlier studies, indels were seldom mentioned or included, mainly because the indels are notorious for adding difficulties in alignment and analysis (Giribet and Wheeler, 1999; Creer et al., 2006; Creer, 2007; Yu et al., 2011). Until recently, identification and incorporation of indels in phylogenetic framework have received considerable attention because an increasing number of studies have shown that indels hold considerable phylogenetic signal and can improve branch supports or can even change a tree topology (Simmons and Ochoterena, 2000; Simmons et al., 2001, 2007; Ogden and Rosenberg, 2007; Egan and Crandall, 2008; Dessimoz and Gil, 2010). For example, Simmons et al. (2001) provided an analysis of 38 DNA regions (including 5 rDNA, 5 ITS, 6 introns, and 22 exons) and found that including gap characters resulted in a change in the amount of resolution or topology of the tree in 28 datasets; Egan and Crandall (2008) used 8 DNA regions (including 2 nuclear and 6 chloroplast genes) to address the ambiguous relationships among the North American Psoraleeae (one of the subdivisions of the plant family Fabaceae), suggesting that the utility of indels could greatly increase phylogenetic resolution and yield different tree topologies in 3 DNA regions.

Most studies, however, largely reached their conclusions by analyzing computer simulation data or a small number of genes from experimental work (Simmons et al., 2007; Ogden and Rosenberg, 2007; Egan and Crandall, 2008; Dwivedi and Gadagkar, 2009). There is, therefore, a need for performing additional empir-

* Corresponding authors. Addresses: State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. Fax: +86 871 5195430 (Y.-p. Zhang). Laboratory for Conservation and Utilization of Bio-Resources and Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming 650091, China. Fax: +86 871 5033362 (L. Yu).

E-mail addresses: zhangyp@mail.kiz.ac.cn (Y.-p. Zhang), yuli1220@yahoo.com.cn (L. Yu).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.10.023

ical studies using larger datasets of the impact of indel inclusion on phylogenetic inferences. In the present study, we are the first to explore the potential effect of indel characters on Arctoidea phylogeny.

The suborder Arctoidea (Mammalia: Carnivora) comprises eight families and is usually organized into three groups: family Ursidae (bears), Pinnipedia (families Odobenidae, walrus; Otariidae, sea lions; and Phocidae, true seals), and Musteloidea (families Procyonidae, raccoons; Mustelidae, weasels; Ailuridae, red panda; and Mephitidae, skunks) (e.g., Flynn et al., 2005; Fulton and Strobeck, 2006; Sato et al., 2006, 2009; Arnason et al., 2007). The precise relationships among Ursidae, Pinnipedia and Musteloidea have been subjects of intense recent controversies (e.g., Yu et al., 2004, 2008, 2011; Delisle and Strobeck, 2005; Fulton and Strobeck, 2006, 2007; Sato et al., 2006, 2009; Yu and Zhang, 2006; Arnason et al., 2007; Peng et al., 2007; Agnarsson et al., 2010; Eizirik et al., 2010). Most previous studies provided either an inconsistent view on the issue or weak statistical support for discriminating alternative hypotheses. Many studies, including morphological studies (e.g., Flower, 1869; Wyss and Flynn, 1993; Hunt and Barnes, 1994), cytochrome b sequences analysis of 243 taxa in Carnivora (Agnarsson et al., 2010) and neighbor-joining analysis of 12 mt protein-coding genes in 15 carnivores (Peng et al., 2007), as well as ML analyses of 22 introns (Yu et al., 2011), connected Ursidae and Pinnipedia to the exclusion of Musteloidea. Other analyses, however, either grouped Pinnipedia and Musteloidea together (Wolsan, 1993; Bininda-Emonds and Russell, 1996; Yu et al., 2004, 2008; Delisle and Strobeck, 2005; Flynn et al., 2005; Fulton and Strobeck, 2006, 2007; Yu and Zhang, 2006; Sato et al., 2006, 2009; Arnason et al., 2007; Peng et al., 2007; Schröder et al., 2009; Eizirik et al., 2010), or associated Ursidae with Musteloidea (Flynn and Nedbal, 1998; Delisle and Strobeck, 2005; Yu and Zhang, 2006).

In sum, the higher-level phylogenetic relationships of Arctoidea have been extensively investigated with mitochondrial (mt) DNA and more recently with nuclear genes (Yu et al., 2004, 2008, 2011; Delisle and Strobeck, 2005; Fulton and Strobeck, 2006; Sato et al., 2006, 2009; Yu and Zhang, 2006; Arnason et al., 2007; Peng et al., 2007; Agnarsson et al., 2010; Eizirik et al., 2010). In these studies, phylogenetic information comes from the analyses of nucleotide substitution characters. To better understand the Arctoidea phylogeny, it is interesting to exploit the information of indels besides nucleotide substitutions, given the rich indels harbored in the intron sequences (Creer, 2007).

In a recent large-scale nuclear intron study by us (22 introns) to investigate the interfamilial relationships of Arctoidea (Yu et al., 2011), we performed phylogenetic analyses using nucleotide substitutions characters in introns. Indels in these introns were either completely excluded or coded as missing data. In that study, the monophyly of Ursidae and Pinnipedia tree and the monophyly of Pinnipedia and Musteloidea tree were both recovered, depending on the tree-building methods used. It is therefore interesting to know whether the incorporation of those indels that were discarded in our previous analyses would provide further insights into the trichotomy among Ursidae, Pinnipedia and Musteloidea.

We tested the impact of indels as phylogenetic characters on phylogeny of Arctoidea here by combining indel characters with the nucleotide substitution characters from the 22 nuclear introns of our previous study (Table 1, Yu et al., 2011). In the present study, indel characters were incorporated into the phylogenetic analyses by using three different coding schemes: the fifth state, simple indel coding (Simmons and Ochoterena, 2000) and modified complex indel coding (Müller, 2006). In addition, indels are examined under both parsimony and Bayesian tree-building methods. Our objectives are (1) to reconstruct phylogenetic relationships within Arctoidea with an emphasis on those among Ursidae, Pinnipedia and Musteloidea; (2) to assess the impact of incorporating indels as phylogenetic characters within Arctoidea.

2. Materials and methods

2.1. Sequence data and alignments

Twenty-two nuclear intron gene sequences of 16 Arctoidea species (Tables 1 and 2; Yu et al., 2011; GenBank ID: FJ692614–FJ693013) were used for the analyses.

First, each intron gene region was aligned using the probabilistic alignment kit (PRANK; Löytynoja and Goldman, 2005, 2008) algorithm with the option "+F" ($PRANK_{+F}$). PRANK algorithm (Löytynoja and Goldman, 2005) implements Markov models and probabilistic score schemes to handle multi-nucleotide indel events and distinguishes insertions from deletions, a step that is fundamental in the context of intron sequence alignment (Löytynoja and Goldman, 2005, 2008; Benavides et al., 2007). In comparison, in most other programs, indels are penalized relative to nucleotide changes, and arbitrarily chosen penalties (Cline et al., 2002; Morrison, 2006; Benavides et al., 2007). In PRANK, the default setting with the guide tree generated by the software was used because the true phylogeny of Arctoidea is unknown. Then, all 22 alignments were concatenated and optimized by locating and removing ambiguous indel positions by the program Gblocks 0.91b (Castresana, 2000) (allowed gap = all).

2.2. Indel treatments and phylogenetic analysis

To test the potential impact of indels on phylogenetic inferences, the concatenated alignment was divided into two data sets by (1) removing all indel characters (referred as dataset 1); all positions with indels and the adjacent non-conserved positions are eliminated using option "allowed gap = none" in Gblocks 0.91b (Castresana, 2000); (2) including all indel characters (referred as dataset 2); positions with indels are treated in the same way as other positions using option "allowed gap = all" in Gblocks 0.91b (Castresana, 2000). The quality of the two datasets was assessed using the head-or-tails (HoT) method (Landan and Graur, 2007). These alignments have been submitted to TreeBASE (TreeBASE ID: S12201).

The sequence characterizations were estimated with program MEGA 5 (Tamura et al., 2011) for different data sets. The resultant numbers of parsimony-informative sites were also used as one of the measures to evaluate the indel coding schemes.

Phylogenetic analyses of dataset 1 were performed using PAUP*4.0b10 (Swofford, 2002) for maximum parsimony (MP) analyses, using RAxML 7.3.0 (Stamatakis, 2006; Silvestro and Michalak, 2010) for maximum likelihood (ML) analyses and using MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) for partitioned Bayesian analyses (PBA). In MP analysis, a heuristic search algorithm was adopted with the TBR branch swapping, random addition of taxa, and 1000 replicates per search. Nodal supports were evaluated using bootstrap analysis with 1000 bootstrap replicates. In ML analysis, each of 22 nuclear intron genes was considered as a different data partition. The "rapid bootstrap" option was run from starting random seeds to generate 100 nonparametric bootstrap replicates. The GTR-CAT model was used because it is faster and less memory intensive than GTR-GAMMA model. The other parameters were set as default. In PBA analysis, each of 22 nuclear intron genes was also considered as a different data partition, which allowed different optimal parameters for each partition (Ronquist and Huelsenbeck, 2003). The best-fit models of 22 intron genes, which were used in the priors of Bayesian