Molecular Phylogenetics and Evolution 65 (2012) 871–882

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogenetic analysis of genome rearrangements among five mammalian orders

Haiwei Luo[a], William Arndt[b], Yiwei Zhang[b], Guanqun Shi[c], Max A. Alekseyev[b], Jijun Tang[b], Austin L. Hughes[a,*], Robert Friedman[a]

[a] *Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA*
[b] *Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA*
[c] *Department of Computer Science, University of California, Riverside, CA 92521, USA*

ARTICLE INFO

Article history:
Received 30 March 2012
Revised 11 August 2012
Accepted 13 August 2012
Available online 21 August 2012

Keywords:
Phylogeny
Genome rearrangement
Breakpoint
Gene order
Placental mammal

ABSTRACT

Evolutionary relationships among placental mammalian orders have been controversial. Whole genome sequencing and new computational methods offer opportunities to resolve the relationships among 10 genomes belonging to the mammalian orders Primates, Rodentia, Carnivora, Perissodactyla and Artiodactyla. By application of the *double cut and join* distance metric, where gene order is the phylogenetic character, we computed genomic distances among the sampled mammalian genomes. With a marsupial outgroup, the gene order tree supported a topology in which Rodentia fell outside the cluster of Primates, Carnivora, Perissodactyla, and Artiodactyla. Results of breakpoint reuse rate and synteny block length analyses were consistent with the prediction of random breakage model, which provided a diagnostic test to support use of gene order as an appropriate phylogenetic character in this study. We discussed the influence of rate differences among lineages and other factors that may contribute to different resolutions of mammalian ordinal relationships by different methods of phylogenetic reconstruction.

1. Introduction

A well-resolved mammalian tree is essential for annotation of genetic features in their genomes and sequence evolution within this taxonomic class. However, the phylogenetic relationships of the 18 extant placental mammalian orders are highly contentious (Cannarozzi et al., 2007; Cao et al., 1998; Hallstrom and Janke, 2008; Kullberg et al., 2007; Li et al., 1990; Springer and de Jong, 2001; Wildman et al., 2007). Two major alternative hypotheses have been proposed regarding the evolutionary relationship within placental mammals. One maintains that Rodentia is more closely related to Primates than Perissodactyla, Artiodactyla and Carnivora (three orders of Laurasiatheria), (Fig 1A), while the alternative supports Rodentia as an outgroup to the other four orders (Fig 1B).

Earlier studies investigated the placental mammalian relationship with single genes of mitochondrial (Cao et al., 1994, 1998) and nuclear genomes (Easteal, 1988, 1990; Goodman et al., 1985; Li et al., 1990). Single gene analyses frequently created discrepancies in branching order, and it was thought that large molecular data sets of concatenated alignments have the potential to resolve these issues (Madsen et al., 2001; Murphy et al., 2001a,b). Whole genomic sequences of mammalian mitochondria, each ranging about 16,500–17,000 bases (Penny and Hasegawa, 1997), provided more base pairs than single genes. Nevertheless, evidence from mitochondrial genomic analyses is inconsistent. Although a majority of these mitochondrial genomic analyses supported the Primates–Laurasiatheria clade which excluded Rodentia (Arnason et al., 1997, 1999; Janke et al., 1994, 1997; Mouchaty et al., 2000; Pumo et al., 1998; Reyes et al., 2000; Springer et al., 1997), mitochondrial genomic evidence for Euarchontoglires (Primates–Rodentia clade) also existed (Arnason et al., 2002, 2008; Reyes et al., 2004). There are other controversies regarding the use of mitochondrial genomic data. A tree derived from first and second codon positions in mitochondrial genes supported the Euarchontoglires hypothesis, whereas another tree of amino acid sequence data suggested a Primates–Laurasiatheria clade (Arnason and Janke, 2002). Further studies showed that complete mitochondrial genomic sequences were valuable for resolving relationships within placental orders, but they appeared inadequate in resolving between-order relationships (Corneli, 2002).

Phylogenomic studies of concatenated alignments of nuclear and mitochondrial genes resolved the 18 extant placental orders into four superordinal groups: Xenarthra, Afrotheria, Laurasiatheria, and Euarchontoglires (Hallstrom et al., 2007; Madsen et al., 2001; Murphy et al., 2001a,b). Since Artiodactyla, Perissodactyla and Carnivora are three orders of Laurasiatheria and Primates and Rodentia are within Euarchontoglires, these studies favored the Primates and Rodentia as a superordinal clade while excluding Artiodactyla, Perissodactyla and Carnivora (Fig 1B). However, the basal position of Rodentia was again underscored in a recent study

---

* Corresponding author. Address: Department of Biological Sciences, University of South Carolina, Columbia, 715 Sumter Street, Columbia, SC 29208, USA.

E-mail address: austin@biol.sc.edu (A.L. Hughes).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.08.008

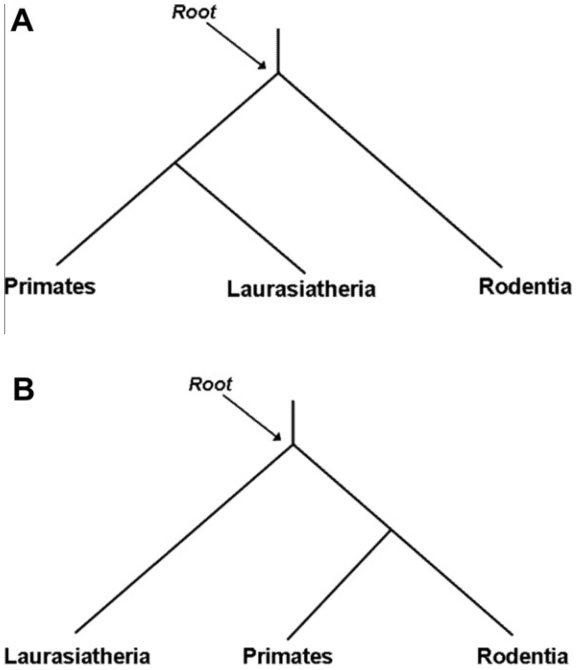

Fig. 1. Two alternative hypotheses for the evolutionary relationship among Primates, Laurasiatheria (Artiodactyla, Carnivora, Perissodactyla) and Rodentia. (A) Primates closer to Rodentia than Laurasiatheria. (B) Primates closer to Laurasiatheria than Rodentia.

of eight housekeeping genes across 22 placental mammals and three marsupials (Kullberg et al., 2006). In addition, another phylogenomic study using distance-, parsimony-, and likelihood-based methods yielded overwhelming support for a Primates–Carnivora clade while excluding Rodentia (Cannarozzi et al., 2007).

The controversial relationships among these mammalian orders have been highlighted by several recent studies. Murphy and others (2001b) claimed that the deep mammalian relationships were resolved by Bayesian-based phylogenetics. Afterwards, Misawa and Nei (2003) showed that Murphy and others' data (2001b) can lead to two different Bayesian trees, both of which were supported by high posterior probabilities.

Kolaczkowski and Thornton (2004) pointed out that when the substitution rate at a single base or amino acid position varies over evolutionary time (referred to as heterotachy), both Bayesian- and likelihood-based methods are statistically inconsistent, leading to an incorrect partition as the amount of data grows (Nishihara et al., 2007; Wildman et al., 2007). Likewise, Hughes and Friedman (2007) found that genes with different substitution rates yielded a different branching order between Carnivora and Rodentia, but concatenation of these genes yielded a tree supporting Rodentia branching first regardless of the method or model of sequence evolution.

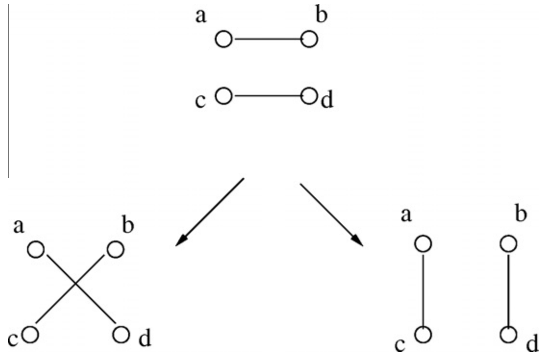

Fig. 2. A DCJ operation on adjacencies (a, b) and (c, d) can create two possible results: (a, c), (b, d) or (a, d), (b, c).

Since several factors can lead to a misleading topology, such as nucleotide or amino acid compositional bias (Nishihara et al., 2007), phylogenetic method (Kullberg et al., 2007), long-branch attraction, and heterotachy, it is desirable to seek whole genome-based phylogenetic characters to resolve the basal relationships in the mammalian tree.

Gene order is a type of rare genomic change, which provides independent ways to evaluate conflicting molecular sequence phylogenies (Rokas and Holland, 2000). It has been demonstrated as a useful phylogenetic character in resolving both shallow and deep prokaryotic relationships (Belda et al., 2005; Luo et al., 2008, 2009). Genome rearrangements include inversion, transposition, block exchange, circularization and linearization, all of which act on a single chromosome, and translocation, fusion, and fission, which act on two chromosomes. All of these operations are subsumed in the Double-Cut-and-Join (DCJ) model, which has formed the basis for much of the algorithmic research on rearrangements over the last few years.

A DCJ operation consists of cutting two adjacencies in the first genome, and rejoining the resulting four unconnected vertices in two new pairs. As a result, it swaps two gene ends in two different vertices of the same genome in the breakpoint graph. Fig. 2 shows an example of a DCJ operation on two adjacencies (a, b) and (c, d). It splits (a, b), (c, d), and can create new adjacencies (a, c), (b, d) or (a, d), (b, c). The DCJ distances between two permutations is defined as the number of minimal DCJ operations needed to transform one permutation into another. If two unichromosomal linear genomes are identical, we can see that there are N cycles in the breakpoint

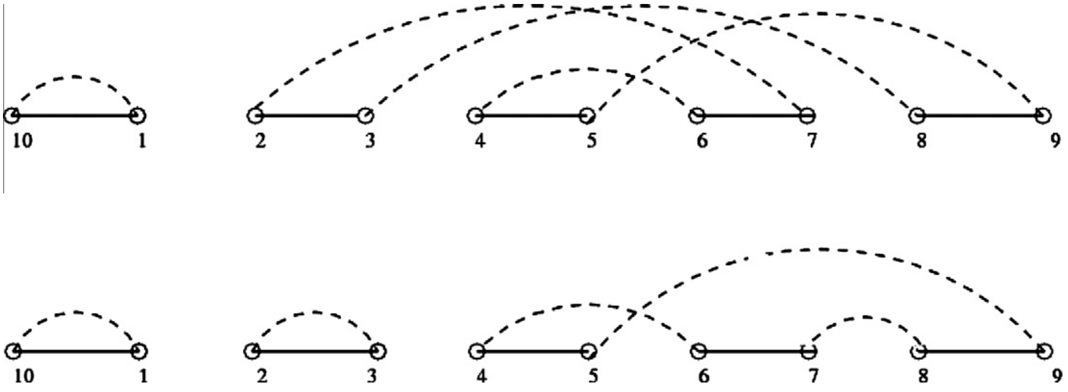

Fig. 3. On the top is the breakpoint graph of genome $G = 1, 2, -3, 4, -5$, with respect to the identity genome $G = 1, 2, 3, 4, 5$. We can see there are two cycles in the graph. An optimal DCJ operation removes adjacencies (2, 7), (3, 8) and creates adjacencies (2, 3), (7, 8) so that the number of cycles increases by one.