Investigating Difficult Nodes in the Placental Mammal Tree with Expanded Taxon Sampling and Thousands of Ultraconserved Elements

Jacob A. Esselstyn*,†, Carl H. Oliveros†, Mark T. Swanson, and Brant C. Faircloth

Museum of Natural Science and Department of Biological Sciences, Louisiana State University, Baton Rouge

†These authors contributed equally to this work.

*Corresponding author: E-mail: esselstyn@lsu.edu.

Accepted: August 25, 2017

Abstract

The phylogeny of eutherian mammals contains some of the most recalcitrant nodes in the tetrapod tree of life. We combined comprehensive taxon and character sampling to explore three of the most debated interordinal relationships among placental mammals. We performed in silico extraction of ultraconserved element loci from 72 published genomes and in vitro enrichment and sequencing of ultraconserved elements from 28 additional mammals, resulting in alignments of 3,787 loci. We analyzed these data using concatenated and multispecies coalescent phylogenetic approaches, topological tests, and exploration of support among individual loci to identify the root of Eutheria and the sister groups of tree shrews (Scandentia) and horses (Perissodactyla). Individual loci provided weak, but often consistent support for topological hypotheses. Although many gene trees lacked accepted species-tree relationships, summary coalescent topologies were largely consistent with inferences from concatenation. At the root of Eutheria, we identified consistent support for a sister relationship between Xenarthra and Afrotheria (i.e., Atlantogenata). At the other nodes of interest, support was less consistent. We suggest Scandentia is the sister of Primatomorpha (Euarchonta), but we failed to reject a sister relationship between Scandentia and Glires. Similarly, we suggest Perissodactyla is sister to Cetartiodactyla (Euungulata), but a sister relationship between Perissodactyla and Chiroptera remains plausible.

Key words: Atlantogenata, Eutheria, Perissodactyla, phylogenomics, Scandentia.

Introduction

Since the dawn of phylogenetics, biologists have sought to resolve the evolutionary relationships among placental mammals (i.e., Eutheria; Matthew and Simpson 1943; Simpson 1959; Novacek 1980; Miyamoto and Goodman 1986; Murphy et al. 2001a, 2001b). The emergence of molecules as characters beginning in the 1980s freed systematists from the burden of morphological homoplasy, and brought to the forefront novel topological hypotheses (e.g., Afrotheria; Porter et al. 1996; Stanhope et al. 1996; Springer et al. 2004). As technology improved from single-locus to genomic-scale data sets, the research community progressively resolved eutherian relationships (Miyamoto and Goodman 1986; Cao et al. 1994; Springer et al. 2004). Nevertheless, several difficult relationships remain, and different data sets and analytical approaches suggest a sampling of all possible evolutionary scenarios at some of these nodes.

One particularly challenging relationship, the earliest divergence within Eutheria, has garnered the most attention (Teeling and Hedges 2013). Although recent phylogenetic studies agree that placental mammals comprise the Xenarthra (sloths and armadillos), Afrotheria (elephants, dugongs, tenrecs, and hyraxes), and Boreoeutheria (Euarchontoglires + Laurasiatheria), investigators have routinely disagreed on the relationships among these three clades (Hallström et al. 2007; Hallström and Janke 2010). The three hypotheses for the earliest placental relationships are known as Atlantogenata (Afrotheria + Xenarthra), Epitheria (Boreoeutheria + Afrotheria), and Exafroplacentalia

Table 1
Summary of Topological Results from a Selection of Recent Phylogenetic Investigations of Mammals

Reference	No. Taxa	No. Loci	Eutheria	Scandentia	Perissodactyla
Hallström et al. (2007)	11	2,840 protein coding	Atlantogenata	NA	NA
Hallström and Janke (2010)	31	3,364 cDNA	Exafroplacentalia	Glires	Carnivora
McCormack et al. (2012)	29, 19	183 and 917 UCEs	Exafroplacentalia	Primates	variable
Meredith et al. (2011)	138	26 DNA and AA	Atlantogenata	Glires	Cetartiodactyla
Morgan et al. (2013)	64, 39	11 and 27 DNA	Atlantogenata	Primates	Cetartiodactyla
Murphy et al. (2001a)	64	16 protein coding	Exafroplacentalia	Dermoptera	Cetartiodactyla
Murphy et al. (2007)	44	20 DNA, coding indels, retroposons	Atlantogenata	Dermoptera	Chiroptera
Nishihara et al. (2006)	25	192 retroposons	Unresolved	Unresolved	Carnivora
Romiguier et al. (2013)	35	1,640 and 172 protein coding	Exafroplacentalia	Glires	Carnivora
Scornavacca and Galtier (2017)	39	5,299 exons	Exafroplacentalia	(Primates, Glires)	(Chiroptera, Carnivora, Cetartiodacyla)
Song et al. (2012)	37	447 DNA	Atlantogenata	Primates	Carnivora
Springer et al. (2003)	42	20 DNA	Exafroplacentalia	Dermoptera	Ferae
Tarver et al. (2016)	33	14,631 protein coding	Atlantogenata	Glires	Cetartiodactyla
Wildman et al. (2007)	11	1,698 protein coding	Atlantogenata	NA	NA

Note.—For each reference, we show the number of eutherian species sampled, the number and type of loci, and the relationships inferred. See text for definition of earliest placental hypotheses. For Scandentia and Perissodactyla, we list the inferred sister group.

(Boreoeutheria + Xenarthra). Although morphological studies often supported Epitheria (O'Leary et al. 2013), even if they did not yet recognize Afrotheria (McKenna 1975), most molecular studies supported Atlantogenata or Exafroplacentalia (Springer et al. 2004; Meredith et al. 2011; Romiguier et al. 2013; Foley et al. 2016; Tarver et al. 2016; summarized in table 1). Fundamentally, the disagreements among molecular studies arise from ancient speciation events that occurred in rapid succession (Murphy et al. 2001a; Wildman et al. 2007; Hallström and Janke 2010). Trees derived from rapid radiations contain short internal branches, which almost certainly contribute to high levels of incomplete lineage sorting (ILS) and gene tree topologies that are poorly supported, unresolved, or both (Jakob and Blattner 2006; Degnan and Rosenberg 2009; Koblmüller et al. 2010).

After the earliest divergence among placental mammals, two of the most difficult topological challenges within Eutheria are the placement of tree shrews (Scandentia) within Euarchontoglires (primates, tree shrews, colugos, rodents, and rabbits) and the position of horses and their relatives (Perissodactyla) within Laurasiatheria. Tree shrews were once considered the sister group of primates (Novacek 1992) and, as such, have been used extensively as medical models (Cao et al. 2003; Yang et al. 2013). However, several early molecular investigations based on a few loci placed colugos (Dermoptera) as the sister to Primates (Porter et al. 1996; Perelman et al. 2011), with tree shrews sister to Dermoptera + Primates (Primatomorpha; e.g., Murphy et al. 2001a). We refer to this arrangement (Scandentia, (Primates, Dermoptera)) as the Euarchonta hypothesis. Some more recent phylogenomic studies have again placed tree shrews sister to primates, but they did not sample Dermoptera (McCormack et al. 2012; Song et al. 2012). In contrast, other phylogenomic studies have placed tree shrews as sister to Glires (Rodentia + Lagomorpha; Hallström and Janke 2010; Romiguier et al. 2013; Mason et al. 2016; Tarver et al. 2016), or as sister to Glires + Primates (Scornavacca and Galtier 2017; table 1). As such, the placement of tree shrews relative to Primates, Dermoptera, and Glires remains open to debate. Again, these conflicts result, at least partially, from the very rapid radiation of Euarchontoglires, which is echoed in the short branches that record interordinal relationships (Mason et al. 2016). However, taxon sampling may also play a role in the inconsistent placement of tree shrews. Thus far, most phylogenomic studies that sampled tree shrews have included only members of the Tupaiidae. However, the monotypic tree shrew family Ptilocercidae is a distant relative of Tupaiidae (estimated ca. 60 Ma in Roberts et al. 2011); including Ptilocercidae in phylogenetic estimates may reduce conflict by breaking up the long tree-shrew branch. Mason et al. (2016) did include Ptilocercidae in their phylogenetic estimates and placed tree shrews as sister to Glires in their analysis of 631 concatenated protein-coding loci, but their coalescent analyses weakly favored tree shrew placement as the sister to all other members of Euarchontoglires.

Relationships within Laurasiatheria, the clade comprising bats, carnivores, pangolins, shrews, artiodactyls, cetaceans, and perissodactyls, have been particularly challenging to resolve. Within this clade, the most consistently supported interordinal relationships are that Eulipotyphla (shrews, moles, and hedgehogs) is sister to all other laurasiatherians and that Carnivora + Pholidota (Ferae) is often well supported

(Nishihara et al. 2006; Zhou et al. 2012). Relationships among the remaining members of Laurasiatheria have proven much more difficult to discern, and some authors have even referred to this challenge as a "speciation network" (Doronina et al. 2017). For instance, Perissodactyla (horses, tapirs, and rhinoceroses) has been placed variably as the sister to: Cetartiodactyla (whales and even-toed ungulates; Murphy et al. 2001a; Meredith et al. 2011; Zhou et al. 2012; Tarver et al. 2016); Carnivora (Xu et al. 1996; Nishihara et al. 2006; McCormack et al. 2012; Romiguier et al. 2013; Song et al. 2012) or Ferae (Murphy et al. 2001b; Springer et al. 2003; dos Reis et al. 2012); or to a clade containing Carnivora, Chiroptera, and Cetartiodactyla (Scornavacca and Galtier 2017; table 1). These relationships may be the most challenging of all to resolve among mammals. They appear to be analogous to the problem observed within Neoaves, where eight lineages diverged practically simultaneously relative to rates of substitution in various classes of genetic loci (Suh 2016). At the base of Laurasiatheria, the nearly contemporaneous divergence of Eulipotyphla, Chiroptera, Ferae, Perissodactyla, and Cetartiodactyla (Zhou et al. 2012) certainly leads to extensive nonphylogenetic signal relative to phylogenetic signal in large genomic character sets (Philippe et al. 2011).

Although tree shapes (i.e., the distribution of branch lengths) associated with rapid diversification clearly present a challenge to efforts at phylogenetic resolution, gene sampling schemes, taxon sampling schemes, and variation in species' natural history traits may also render some lineages difficult to place in a consistently supported phylogenetic hypothesis (Jeffroy et al. 2006). Short branches between speciation events are documented by few substitutions and may be obscured by extensive ILS. Variation in substitution rates, variation in generation times, or poor taxon sampling can produce long branch attraction, leading to the inference of spurious relationships between distant relatives (Felsenstein 1978; Hendy and Penny 1989; Huelsenbeck 1995). Sequences can become saturated with multiple substitutions when inferring ancient relationships, increasing the frequency of homoplasy and reducing the fit of sequence evolution models (Yang 1998; Lartillot et al. 2007). Hybridization can generate nonbinary relationships, which are not accommodated by traditional phylogenetic inference (Huson and Bryant 2006). Some of these issues can be overcome by explicit modeling (i.e., coalescent analyses that accommodate ILS) or improved sampling (greater taxon sampling to break up long branches and reduce homoplasy; Hendy and Penny 1989; Heath et al. 2008; Lanier and Knowles 2015). But others, particularly when speciation occurred rapidly, remain challenging (Song et al. 2012; Giarla and Esselstyn 2015). Although many investigators of mammalian phylogeny have embraced coalescent modeling for its ability to accommodate ILS and potentially escape the anomaly zone (Song et al. 2012; Edwards et al. 2016), others have criticized this approach, particularly for using unreliable gene trees (Gatesy and Springer 2014). Still others have concluded that ILS is a relatively minor issue in the deeper levels of mammalian phylogenetics (Scornavacca and Galtier 2017). Compounding these issues, researchers have also shown that a very small number of loci or sites can dominate signal in phylogenomic studies, leading to high support values that should be interpreted carefully (Brown and Thomson 2016; Shen et al. 2017). Summary coalescent methods, which have been criticized for not taking account of the strength of support for individual gene trees (Gatesy and Springer 2014), may in fact provide a remedy for this issue. Because individual gene trees are equally influential, this may prevent a few "rogue" loci from dominating signal.

Although many studies of higher-level mammalian relationships have employed a large suite of analytical techniques, most recent studies have emphasized obtaining larger genomic data sets, often at the expense of taxon sampling (table 1). When taxon sampling is limited, inferences of topology and branch lengths can be affected (Zwickl et al. 2002; Hedtke et al. 2006), and the value of comparing results among studies is greatly reduced. For example, limited taxon sampling has hindered our understanding of tree shrew placement because many studies do not include a representative of Dermoptera or Ptilocercidae, rendering potential comparisons among studies moot, and potentially leading to long branch attraction.

In addition to designing an appropriate taxon-sampling scheme, researchers hoping to resolve recalcitrant nodes are faced with choosing which loci are most appropriate. This seemingly simple decision obscures many potentially important variables, both practical (e.g., financial constraints) and theoretical (e.g., the relationship between GC content and the probability of recombination; Romiguier et al. 2013). Ultraconserved elements (UCEs) represent a potentially wise choice for resolving deep mammalian relationships because they evolve slowly and are therefore less affected by saturation than other loci; they are unambiguously homologous; they can be extracted bioinformatically from published genomes and easily collected in the lab using a standardized molecular approach; their lengths are generally sufficient to contain many parsimony informative sites at deep time scales, yet short enough to avoid recombination; and their GC content is typically low, reflecting a lower rate of recombination and reducing conflict among recovered gene trees (Chen et al. 2007; McCormack et al. 2012; Faircloth et al. 2012; Romiguier et al. 2013). However, the slow substitution rate of UCEs, while providing several advantages, may yield little phylogenetic signal on short internal branches.

Here, we attempt to resolve three of the most difficult nodes in the placental mammal tree using a novel combination of several thousand UCE loci, expanded taxon sampling that includes most mammalian families, and a suite of analytical approaches. We employ coalescent-based and

concatenation approaches to phylogeny estimation, we include formal tests of alternative topologies, and we explore in detail the support of individual loci for various topological hypotheses. Our goal is to provide further certainty with regard to eutherian relationships, while acknowledging the considerable noise around difficult nodes in phylogenomic data sets.

Materials and Methods

We collected UCE loci from 72 species using in silico alignment of a 5,060 UCE probe set (Faircloth et al. 2012) to genomes available from NCBI, followed by extraction of the matched region +300 bp of flanking nucleotides on each side (supplementary appendix S1, Supplementary Material online).

To improve taxon sampling, we enriched and sequenced UCEs in 28 additional species (supplementary appendix S1, Supplementary Material online). We extracted genomic DNA from tissue samples using a Qiagen DNeasy Blood and Tissue Kit. After quantifying DNA in the extracts with a Qubit 2.0 fluorometer, we cleaned 1,000 ng aliquots with 3× the volume of Sera-Mag Carboxylate-modified SpeedBeads (Rohland and Reich 2012) and eluted DNA into 30 µl of TE buffer. We mechanically sheared DNA in 2.5-min increments at 17 mA with an Epigentek Episonic sonicator until the average fragment size was ~500 bp, as assessed by eye on an electrophoretic gel.

We prepared DNA libraries using the Kapa Biosystems Hyper Prep Kit for Illumina platforms with dual indexed iTru adapters (Glenn et al. 2016). We used one-fourth of the manufacturer's recommended reagent volume, performing a 1× postligation bead cleanup, and increased the library amplification extension time to 1 min. We combined the resulting DNA libraries in equimolar pools of eight samples and enriched each pool for 5,060 UCE loci using the Tetrapods-UCE-5Kv1 probe set (Faircloth et al. 2012) sold by MYcroarray (Ann Arbor, MI). We followed the manufacturer's instructions for enrichment. We determined the size distribution of enriched libraries with an Agilent Bioanalyzer and removed remaining adapter dimer from pools where it was present using a 0.8× bead cleanup. We then quantified the enriched libraries with a Qubit 2.0 fluorometer and pooled them in equimolar ratios. We sequenced the enriched libraries on an Illumina HiSeq 3000 PE 150 lane at the Oklahoma Medical Research Foundation (Oklahoma City, OK).

The Oklahoma Medical Research Foundation demultiplexed raw reads from BCL files using bcl2fastq2 ver. 2.17.1.14 (Illumina Inc.) and returned FASTQ-formatted files to us. We trimmed low-quality bases and adapter sequences from reads using illumiprocessor ver. 2 (https://github.com/faircloth-lab/illumiprocessor; last accessed September 1, 2017), which incorporates trimmomatic (Bolger et al. 2014). We used the Python package PHYLUCE (Faircloth 2016) for subsequent data processing. We assembled cleaned reads into contigs using SPAdes ver 3.9.0 (Bankevich 2012) and extracted contigs for each taxon that matched UCE loci. We assembled an incomplete data set containing UCE loci that were present in at least 70 of the 100 taxa. We aligned each locus with MAFFT (Katoh and Standley 2013), and we trimmed resulting alignments to allow missing nucleotides at the flanks of each alignment only if at least 65% of taxa contained data, which is the default in PHYLUCE. We further trimmed uncertain alignment regions using Gblocks (Castresana 2000) with default parameters except for the minimum number of sequences for a flank position, which we set at 65% of taxa. We created a concatenated file of all loci using PHYLUCE. Our alignments contain a platypus and seven marsupial families as outgroups, with 92 eutherian species representing 80 families.

We performed maximum likelihood (ML) inference on the concatenated data set using ExaML ver. 3.0.15 (Kozlov et al. 2015) assuming a general time reversible model with Γ-distributed rates among sites. We performed 20 ML searches and evaluated node support with 100 bootstrap replicates. Bootstrapped data sets were generated with RAxML ver. 8.2.8 (Stamatakis 2014), but searches were conducted with ExaML.

We employed three coalescent-based approaches to estimate a species tree. First, we used SVDquartets (Chifman and Kubatko 2014, 2015), as implemented in PAUP* ver. 4.0a150 (Swofford 2003) to generate quartets of species. We then assembled a species tree from the quartets using Quartet MaxCut ver. 3.0 (Snir and Rao 2012). We performed SVDquartets analyses on the same 100 bootstrap replicates that we generated for concatenated ML analysis. We also employed two species-tree methods that use unrooted gene trees as input. We used PHYLUCE to generate 100 multi-locus bootstraps of the data (Seo 2008) and we performed gene-tree inference on these replicates in RAxML (Stamatakis 2014). For each of the 100 replicate data sets, we performed species tree estimation with ASTRAL ver. 4.10.11 (Mirarab et al. 2014; Mirarab and Warnow 2015) and ASTRID ver. 1.4 (Vachaspati and Warnow 2015).

For all analyses, we tested for convergence of bootstrap replicates a posteriori using the "autoMRE" option in RAxML (Pattengale et al. 2010; Stamatakis 2014). We summarized replicates using sumtrees.py (Sukumaran and Holder 2010). To test whether improved taxon sampling might explain any differences between our inferences and those of McCormack et al. (2012), we repeated phylogenetic estimation as above using ExaML, ASTRAL, ASTRID, and SVDquartets on a data set we reduced to 23 tips (supplementary appendix S1, Supplementary Material online) selected to mimic McCormack et al.'s (2012) alignments.

To explore the topological consistency of individual gene trees with hypothesized species relationships, we summarized the proportion of gene trees containing particular bipartitions and the average bootstrap support values (BS) of these

Table 2
Results of the Approximately Unbiased (AU) Tree Selection Test

Hypothesis	AU	*P* value
Atlantogenata	**1.0**	**1.0**
Epitheria	<0.0001	<0.0001
Exafroplacentalia	<0.0001	<0.0001
(Scandentia, Primatomorpha)	**0.922**	**0.926**
(Scandentia, Primates)	<0.0001	<0.0001
(Scandentia, Glires)	**0.078**	**0.073**
(Scandentia, (Glires, Primatomorpha))	<0.0001	<0.0001
(Perissodactyla, Cetartiodactyla)	**0.802**	**0.765**
(Perissodactyla, Chiroptera)	**0.266**	**0.219**
(Perissodactyla, Carnivora)	<0.0001	<0.0001
(Perissodactyla, Ferae)	<0.0001	<0.0001
(Perissodactyla, (Cetartiodactyla, (Chiroptera, Ferae)))	0.032	0.015

NOTE.—Bolded rows are those hypotheses that could not be rejected at $\alpha = 0.05$.

bipartitions under both single- and multi-condition filters using a custom Python script (http://github.com/carloliveros/mammals; last accessed September 1, 2017). For instance, to count the number of gene trees consistent with Atlantogenata (Afrotheria + Xenarthra), we filtered gene trees that contained Afrotheria and Xenarthra in a single clade (single-condition filter) and then we ran the same filter, but with the added conditions that both Afrotheria and Xenarthra were monophyletic. We ran similar, relevant filters for all plausible relationships at the deepest node among placental mammals, and for the relationships of Scandentia and Perissodactyla. To explore whether the proportion of gene trees and their average BS provided consistent signal, we tested for a positive correlation (Pearson product-moment) between these values in R ver. 3.0.2 (R Core Team 2013).

To explore the strength of support of individual loci for alternative topologies, we compared site likelihoods given certain topological constraints, following the general approach of Shen et al. (2017). We made these comparisons for each locus between our ML topology and several plausible alternatives (table 2). We used topologies from our ExaML search described earlier, and from the best of five ExaML searches given each of nine alternative topologies. We calculated site likelihoods for each topology (ML and alternatives) using RAxML. We then calculated a site-wise ΔlnL by subtracting the value obtained for the alternative hypothesis from the value for the ML hypothesis. We summed site-wise likelihood differences over the length of individual loci to obtain a locus-specific ΔlnL for each comparison between the ML solution and an alternative topology. As a result, ΔlnL is positive when it supports the ML hypothesis, but negative when it favors the alternative topology. To provide some perspective on how we should interpret the strength of support from ΔlnL values, we also analyzed the well-established sister relationship between rodents and lagomorphs (Glires). In this comparison, we tested Glires versus Rodentia + Euarchonta and Glires versus Lagomorpha + Euarchonta. To examine whether loci that strongly supported a particular hypothesis were potentially linked, we plotted the position on the human genome of loci with ΔlnL values >0.15 and <−0.15 for two hypotheses. These were the comparison of Atlantogenata with Epitheria, and Glires with Rodentia + Euarchonta.

Finally, to provide a more traditional frequentist approach to testing alternative topologies, we used the approximately unbiased test (AU; Shimodaira 2002) to compare ML solutions with alternate hypotheses for our three nodes of interest (i.e., excluding the "control" Glires comparisons). The AU test was conducted in CONSEL ver. 0.20 (Shimodaira and Hasegawa 2001).

Results

We collected UCE loci from 72 published genomes available from NCBI and we enriched and sequenced the same loci in 28 additional species (supplementary appendix S1, Supplementary Material online). Our sequencing resulted in 1.1–8.9 ($\bar{x} = 3.67$) million reads per sample, which we assembled to 5,267–120,835 ($\bar{x} = 24{,}675$) contigs per sample. On an average, we obtained more and longer loci from published genomes ($\bar{x} = 4{,}165$ loci, $\bar{x} = 712$ bp) than from our own enrichment of UCEs ($\bar{x} = 2{,}805$ loci, $\bar{x} = 599$ bp). The alignments we analyzed, which allowed a maximum of 30 missing taxa, contained: 100 taxa, including 92 eutherians representing 80 families; 3,787 loci; 101–1,178 ($\bar{x} = 680$) bp per locus; and 39–665 ($\bar{x} = 332$) variable sites, 23–507 ($\bar{x} = 241$) parsimony informative sites, 5% missing characters, and 39% GC content.

Our concatenated analysis produced strong support across most relationships, and includes many standard inferences for uncontroversial clades (fig. 1). This analysis supported the Atlantogenata (Afrotheria + Xenarthra) and Euarchonta (Scandentia + Primatomorpha) hypotheses, each with a bootstrap value (BS) of 100. The concatenated analysis placed Perissodactyla as the sister to Cetartiodactyla (Euungulata), but this was not well-supported (BS = 54) and should be viewed as unresolved. Branch lengths near the roots of Eutheria, Scandentia, and Perissodactyla are all extremely short (fig. 1).

Our coalescent species-tree inferences (fig. 2) were largely consistent with the concatenated topology, but generally received less support where short internal branches are evident in the concatenated tree (fig. 1). Support values across the ASTRAL and ASTRID trees were generally greater than those from SVDquartets. Consistent with the concatenated topology, ASTRAL and ASTRID supported the Atlantogenata hypothesis with BS of 100, whereas SVDquartets weakly favored (BS = 53) the Exafroplacentalia hypothesis (Xenarthra + Boreoeutheria). In contrast to the concatenated tree, both ASTRAL and ASTRID placed Scandentia as the sister to all other Euarchontoglires (i.e., Glires + Primatomorpha),

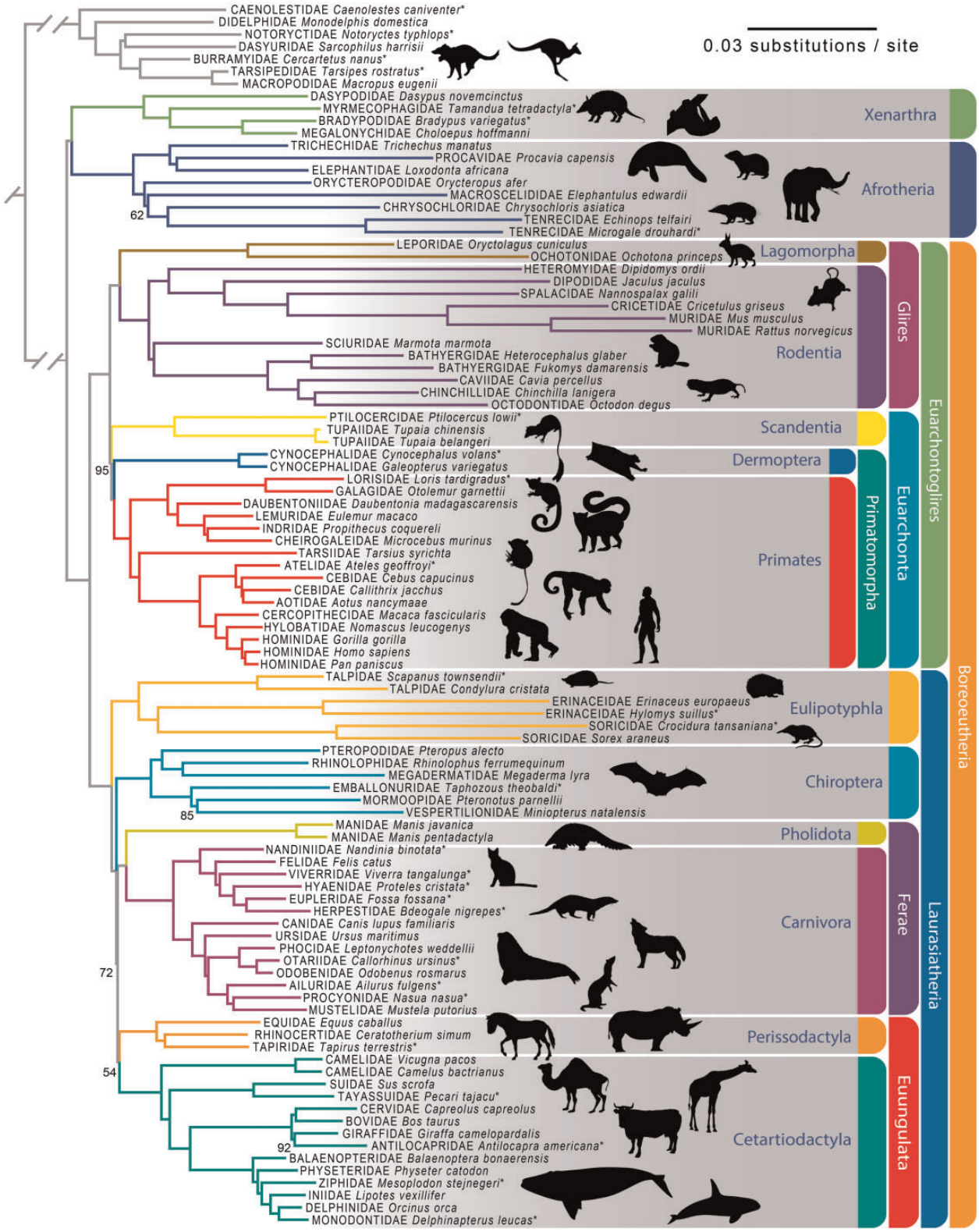

Fig. 1.—Maximum likelihood estimate of mammalian phylogeny from 3,787 concatenated ultraconserved elements. Clades discussed in the text are labeled. Branches between Metatheria (marsupials), Eutheria (placentals), and Prototheria (platypus; not shown) were truncated for ease of presentation. Bootstrap support values are 100 unless otherwise noted. Taxa with an asterisk after the specific epithet are those we enriched and sequenced.

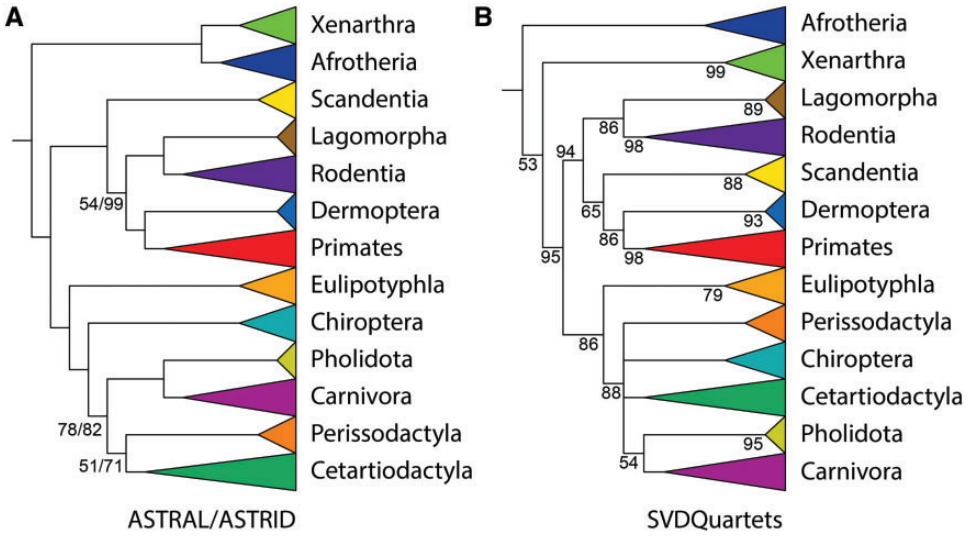

Fig. 2.—Species tree estimates of eutherian relationships derived from analysis of 3,787 ultraconserved elements in (*A*) ASTRAL and ASTRID, and (*B*) SVDquartets. The topology is collapsed to the ordinal level or higher. Bootstrap support values are 100 unless otherwise noted.

but with variable support (54 and 99, respectively; fig. 2*A*). SVDquartets, however, placed Scandentia as sister to Primatomorpha (BS = 65; fig. 2*B*), as in the concatenated tree. Perissodactyla was placed as the sister of Cetartiodactyla, but with BSs of only 51 (ASTRAL; fig. 2*A*) and 71 (ASTRID; fig. 2*A*). The SVDquartets tree (fig. 2*B*) was less resolved, and placed Perissodactyla in a polytomy with Cetartiodactyla, Ferae, and Chiroptera. This unresolved clade was sister to Eulipotyphla, as in our other inferences.

Because we anticipated that improved taxon sampling might explain any differences between our inferences and those of McCormack et al. (2012), who also analyzed UCEs, we repeated phylogenetic estimation on a data set we reduced to 23 tips (supplementary appendix S1, Supplementary Material online) selected to mimic McCormack et al.'s (2012) taxon sampling. Our analyses of this reduced data set were mostly, but not entirely consistent with our inferences from the full data set. As in the full data set, three of four phylogenetic estimates, supported Atlantogenata with BS values of 100 (supplementary fig. S2, Supplementary Material online). SVDquartets also supported Atlantogenata, but with BS of only 51. In the reduced data set, Scandentia was placed as sister to Primates (Dermoptera was not included) in all four analyses, with strong support (BS > 95) in coalescent analyses, but not in the concatenated tree (BS = 63; supplementary fig. S2, Supplementary Material online). Perissodactyla was not confidently placed in any position in analyses of the reduced data set. It was found as sister to Chiroptera (BS = 72) by ExaML, sister to Carnivora + Cetartiodactyla (BSs = 72 and 53) by ASTRAL, and in essentially a polytomy (BSs < 60) with Chiroptera, Carnivora, and Cetartiodactyla by ASTRID and SVDquartets (supplementary fig. S2, Supplementary Material online).

In the following sections, we consider evidence in support of three topological hypotheses for the earliest divergence in Eutheria, four hypotheses for the sister group of Scandentia, and five hypotheses for the sister group of Perissodactyla. The percentage of gene trees from the full data set consistent with individual topological hypotheses (table 2) was generally low (fig. 3), even for relationships that are beyond debate (e.g., primate monophyly = 40%; fig. 3*B*). Only 59% of gene trees contained one of our three hypotheses for the earliest divergence among placental mammals (fig. 3*A*). When we added constraints that included the monophyly of the descendent clades (e.g., for Atlantogenata, we added the requirement that Afrotheria and Xenarthra were each monophyletic), these numbers dropped further, with fewer than half of the gene trees presenting topologies consistent with any viable hypothesis. However, among these, more gene trees were consistent with Atlantogenata (24.7% of all gene trees) than the two alternatives (12% each; fig. 3*A*). Percentages were much lower for the other nodes of interest. For instance, only 4% of gene trees grouped Scandentia and Primates in our multi-constraint tally, with <3% consistent with each of the other hypotheses (fig. 3*B*). Our single-constraint tallies provided virtually no further evidence regarding the best topology, with nearly identical percentages for three of the four possible relationships of tree shrews (fig. 3*B*). Individual gene trees that placed Perissodactyla and Cetartiodactyla in a clade (fig. 3*C*) were slightly more common than alternative topologies, with 5.2% captured by the multi-constraint filter and 6% by the single-constraint filter. Among the alternative hypotheses, Perissodactyla + Chiroptera was most common, with 4.1% (multi-constraint) and 4.7% (single-constraint) of gene trees containing this topology (fig. 3*C*).

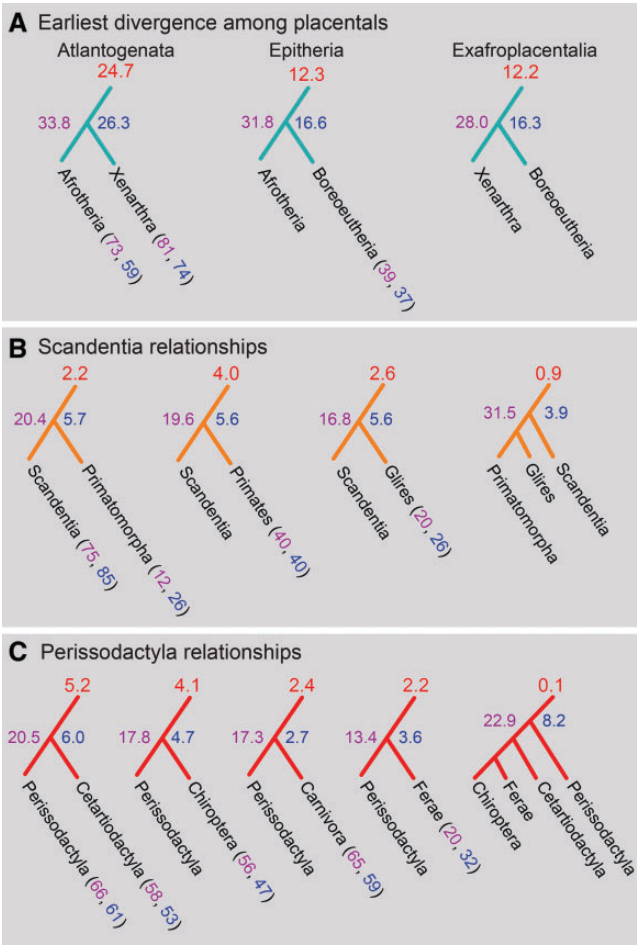

Fig. 3.—Summary of the consistency of individual gene trees with competing topological hypotheses for species relationships. (*A*) Hypotheses regarding earliest divergence among placentals, (*B*) relationships of Scandentia, and (*C*) relationships of Perissodactyla. The percentage of gene trees consistent with a multi-constraint topological filter is shown above each topology (e.g., 24.7% of gene trees contain monophyletic Afrotheria, Xenarthra, and Afrotheria + Xenarthra). The percentage of gene trees consistent with a single-constraint topological filter is shown to the right of the relevant node (e.g., 26.3% contain a monophyletic Afrotheria + Xenarthra). Average bootstrap support (BS) at a particular bifurcation are shown to the left of each node (e.g., 33.8% mean BS for Afrotheria + Xenarthra). In each panel, relevant tip labels include the percent of gene trees containing that clade, followed by their average BS at that node among those gene trees (e.g., Afrotheria found in 73% of gene trees with mean of 59% BS).

Bootstrap values on individual gene trees were generally low (fig. 3), but positively correlated ($t = 15.7$, df = 22, $P \ll 0.0001$) with the proportion of gene trees (single-constraint filters) supporting that same hypothesis (supplementary fig. S1, Supplementary Material online). Even well-established relationships, such as the monophyly of Xenarthra, Afrotheria, and Boreoeutheria were supported by only 74, 59, and 37 average BS, respectively. More debatable nodes received lower values. For instance, the Atlantogenata hypothesis was supported by an average BS of 34 (fig. 3*A*). Nevertheless, BS values for the two alternative hypotheses were lower still. For Scandentia, BSs were very similar for a sister relationship to Primatomorpha (20.4) as they were for a sister relationship to Primates (19.6), but the value was somewhat lower for a sister relationship to Glires (16.8). For Perissodactyla, mean BSs were highest for placing it with Cetartiodactyla (20.5), with a maximum of 17.8 supporting alternative placements.

Per gene ΔlnL values, which quantify the support of individual loci for paired hypotheses and are positive when favoring the primary hypothesis but negative when they support the alternative hypothesis (Shen et al. 2017), were small, even for our "control" tests of Glires. This indicates that very few genes provide strong information regarding each hypothesis (fig. 4). For the earliest divergence among placental mammals, >80% of loci had positive values for each comparison, suggesting weak, but common support for Atlantogenata. However, among these 80% of loci, only six loci had a ΔlnL value >0.5 in each comparison (fig. 4). Loci supporting Atlantogenata and Glires with ΔlnL values >0.15 are scattered across the human genome (supplementary fig. S3, Supplementary Material online). At the opposite end of the scale, only one locus had a ΔlnL value <–0.5. For the placement of Scandentia sister to Primatomorpha versus Primates, ΔlnL values were mostly (84%) positive, but no values exceeded an absolute value of 1 (fig. 4). For our comparisons with the alternative hypothesis that Glires is the sister group, two loci had positive values >0.25 favoring the primary hypothesis, but most loci (66%) had values just below zero (fig. 4). For the alternative hypothesis that Scandentia is sister to Glires + Primatomorpha, most (83%) values were positive but near zero (fig. 4). For the relationships of Perissodactyla, most ΔlnL values were again near zero, but surprisingly in the comparison between our maximum likelihood hypothesis (Euungulata) and Perissodactyla + Chiroptera, 62% of values were negative, suggesting weak support for this alternative hypothesis (fig. 4). The loci with the greatest ΔlnL magnitudes were also negative in this comparison (fig. 4). In contrast, Euungulata was more strongly favored over Perissodactyla + Carnivora, with 12 ΔlnL values >1.0 (fig. 4). However, this distinction disappeared with comparison to Perissodactyla + Ferae; although most values were positive (90%) in this comparison, magnitudes were exceedingly low (fig. 4). For the alternate hypothesis that Perissodactyla is the sister to all Laurasiatheria except Eulipotyphla, ΔlnL values were again flat, but mostly (85%) positive. Although we did not test for a correlation, the number of parsimony informative sites in a locus was not clearly related to the magnitude of ΔlnL values in any of our comparisons (fig. 4).

Approximately unbiased tests (AU; Shimodaira 2002), which also employ site likelihoods, statistically favored Atlantogenata over both alternates (table 2). For Scandentia,

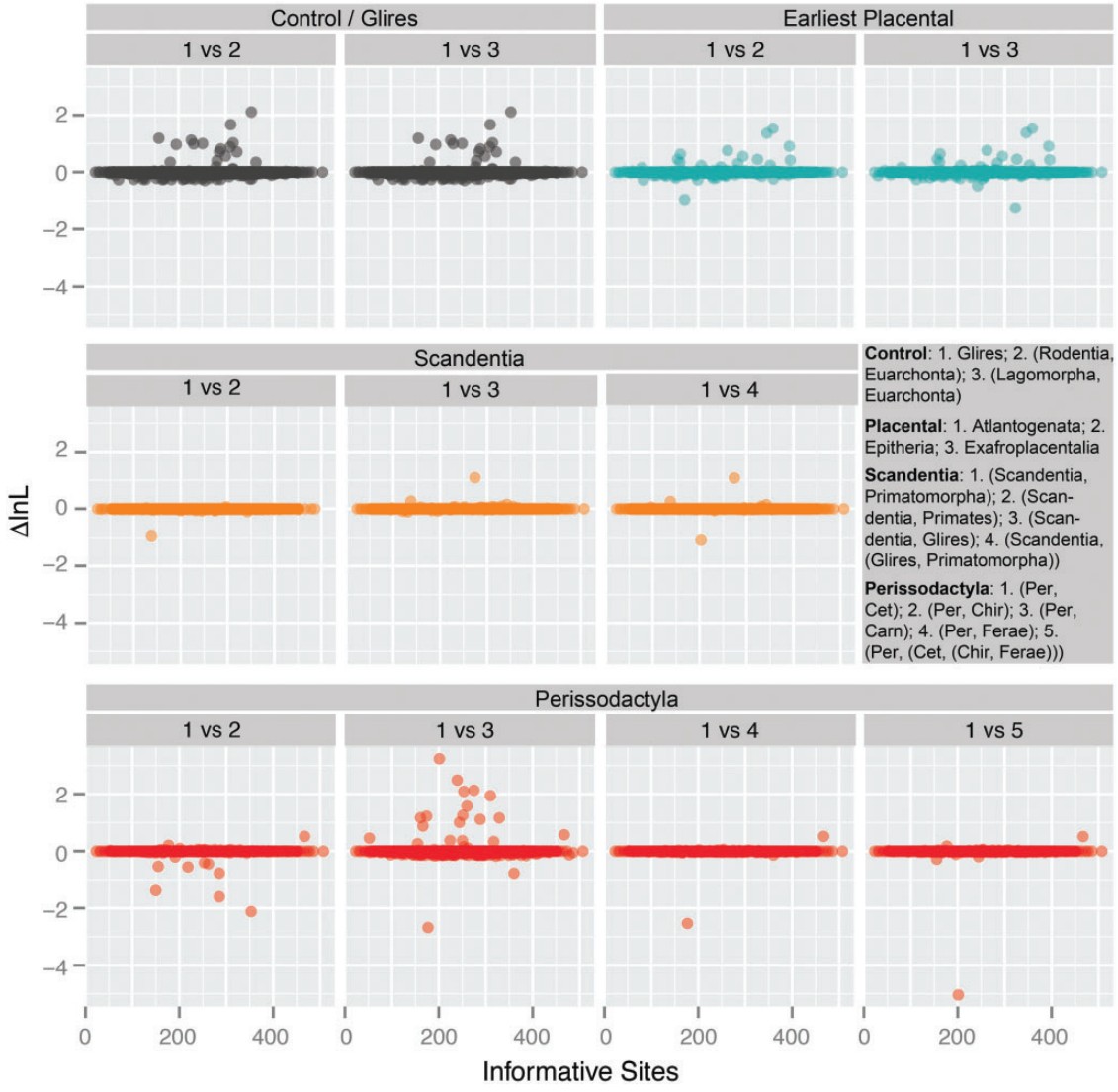

Fig. 4.—Plots of per gene ΔlnL against the number of parsimony informative sites for our maximum likelihood (ML) tree (Hypothesis 1) versus 2–4 alternative hypotheses (Hypotheses 2–5). Each topological hypothesis is defined in the gray box at center right. The following orders are abbreviated: Perissodactyla (Per); Cetartiodactyla (Cet); Chiroptera (Chir); and Carnivora (Carn).

the AU test rejected a sister relationship to Primates and to Glires + Primatomorpha, but failed to reject a sister relationship to Glires (table 2). Similarly, the AU test rejected three plausible alternate hypotheses for Perissodactyla, but failed to reject Perissodactyla + Chiroptera (table 2).

Discussion

Resolving the most challenging nodes in the placental mammal tree remains difficult. Our combination of expanded taxon sampling and thousands of loci represents a balanced approach relative to past studies that emphasized either genomic sampling or taxon sampling, but not both (Bininda-Emonds et al. 2007; Tarver et al. 2016). Our decision to use UCEs for this project provided a ready source of thousands of loci with clear homology and low probability of saturation or recombination, all obstacles that would otherwise make inferences difficult. However, the relatively low substitution rate of UCEs meant that although our alignments contained an average of 241 parsimony informative sites per locus, these sites provided only limited resolution of the difficult questions we sought to answer. We employed a variety of analytical techniques, including concatenation and a suite of coalescent species tree approaches. Each analysis has strengths and weaknesses, and as such, consistency between these results should be viewed as better support for a hypothesis than BS values from a single analysis (Suh 2016). Although our results are not perfectly consistent across analyses, they are largely congruent and, where conflict is evident, the conflicting relationships are not strongly supported in all but one case.

The resolution we gained is strongest for the earliest divergence among placentals, and somewhat less clear for the relationships of Scandentia and Perissodactyla, where we eliminated some, but not all alternative hypotheses. In general, BSs were high across the concatenated tree, with more modest values on our coalescent topologies. This may be due to coalescent methods better reflecting uncertainty (Edwards et al. 2016), to individual genes containing limited phylogenetic signal, or to the way we quantified support. We bootstrapped sites for concatenation and SVDquartets, but we bootstrapped across both loci and sites for ASTRAL and ASTRID analyses. The latter bootstrapping approach seems likely to reduce support values in data sets with high variance in gene tree topologies.

Because the relationships we attempted to resolve are so recalcitrant, and because our taxon sampling was relatively dense, it is not surprising that relatively few individual gene trees were consistent with accepted species relationships. Similarly, BSs were generally low at our nodes of interest on these gene trees. Although concatenation may reveal hidden support for relationships by combining individual genes with low information content (Gatesy and Springer 2014), it can be misleading where ILS is common (Edwards et al. 2016). Coalescent methods are designed to accommodate ILS (Edwards et al. 2007; Degnan and Rosenberg 2009), but may be impeded by inaccurate estimates of gene trees (Gatesy and Springer 2014; Blom et al. 2017). Furthermore, the relative importance of these issues may vary over the depth of time across which analyses are conducted (Lanier and Knowles 2015). The relationships we attempted to resolve range from roughly 100 (crown Eutheria) to 80 Ma (crown Laurasiatheria; Meredith et al. 2011). Although there is extensive noise among our gene trees, many widely accepted relationships are highly supported by our coalescent analyses. For example, Glires (Rodentia + Lagomorpha) is supported by all three of our species tree analyses (fig. 2), despite the presence of this relationship in only 20% of gene trees (fig. 3). An earlier phylogenetic study that used UCEs (McCormack et al. 2012) was criticized for poor gene tree resolution because only 8 of 183 (4%) UCE loci contained Glires (Gatesy and Springer 2014). The proportionally greater gene tree consistency with species relationships in our study, despite our expanded taxon sampling, may be attributable to the longer UCEs we analyzed (680 vs. 517 bp in the 29-taxa data set from McCormack et al. 2012).

Another factor potentially affecting accuracy in phylogenomic studies is the fit of sequence evolution models. Tarver et al. (2016) suggested an appropriate model was key to successfully reconstructing the earliest divergence among placental mammals, and they endorsed usage of site-heterogeneous models (e.g., CAT-GTR; Lartillot and Philippe 2004) for their data. However, the computational tools needed to quantify model fit and implement locus-specific models in large data sets are not yet available. This forced Tarver et al. (2016) to subset their data: their preference for site-heterogeneous models was based on replicates of ~5,000 sites. Tarver et al. (2016) also removed their representative tree shrew and horse from some analyses with site-heterogeneous models because the models would not converge. Due to these limitations, we implemented a simpler site-homogeneous model (GTR + Γ). However, given that our inferences at the root of Eutheria are consistent with Tarver et al.'s (2016), model fit does not appear to be a significant issue in our UCE data.

Regarding the earliest split among placentals, our study provides strong support for the Atlantogenata hypothesis. We inferred Atlantogenata in our concatenated analysis and in two of three coalescent-based approaches. Only SVDquartets suggested an alternative, but this was essentially a polytomy with BS of 53 (fig. 2*B*). Approximately twice as many gene trees contained Atlantogenata as either of the alternative hypotheses, and the BS values of these relationships were higher, on average, than those of the alternative hypotheses (fig. 3). Interestingly, McCormack et al. (2012) supported Exafroplacentalia with a smaller UCE data set (fewer taxa, fewer loci, and shorter loci in some analyses). They noted that 73% of their gene trees resolved one of the plausible relationships for this node. Our percentage (49%) was much lower, and we attribute this difference to our expanded taxon sampling, especially within Xenarthra, where we sampled four families (only one in McCormack et al. 2012). Nevertheless, only 1% more gene trees in McCormack et al. (2012) were consistent with Exafroplacentalia than with Atlantogenata. Given this minor difference, and the fact that we still inferred Atlantogenata consistently on our reduced-taxa data set, we suspect Atlantogenata now emerges from UCE data as the favored hypothesis due to the larger number of loci and longer loci we analyzed. Per-gene ΔlnL values also showed weak, but consistent preference for Atlantogenata (fig. 4). The loci with the strongest support for this hypothesis are scattered across the human genome (supplementary fig. S3, Supplementary Material online), indicating they are not linked. Finally, our AU test of Atlantogenata rejected both alternative hypotheses (table 2). In summary, every analysis we conducted either favored Atlantogenata (concatenation, summary coalescent, ΔlnL, and AU test), or lacked signal (SVDquartets).

We considered four possible relationships for Scandentia (fig. 3). Many previous studies of interordinal relationships in mammals did not sample colugos (Dermoptera), rendering comparisons among these studies difficult. To improve taxon sampling and potentially distinguish between a larger number of alternative hypotheses, we included both genera of Dermoptera, and we added *Ptilocercus*, the sole member of the tree shrew family Ptilocercidae. Our concatenated and SVDquartets analyses supported the Euarchonta hypothesis (Scandentia + Primatomorpha), but summary coalescent analyses placed Scandentia as the sister to all other Euarchontoglires. ASTRID results strongly conflicted

(BS = 100; fig. 2A) with the concatenated topology, but the ASTRAL result received much lower support (BS = 54; fig. 2A). Surprisingly, our reduced data set consistently placed Scandentia as sister to Primates (Dermoptera not included). This suggests that our expanded sampling of taxa and loci relative to those of McCormack et al. (2012) may have increased the amount of noise relative to phylogenetic signal. In contrast to our summary coalescent results from the full data set, the AU test soundly rejected the placement of Scandentia as sister to all other Euarchontoglires (table 2). Other studies using a relatively large number of loci have often placed Scandentia as the sister to Glires (Romiguier et al. 2013; Tarver et al. 2016). Although this relationship did not arise in any of our phylogenetic inferences, our AU test failed to reject it (table 2). The percentages of gene trees consistent with a sister relationship between Scandentia and Primatomorpha, Primates, or Glires were essentially equal (5.6–5.7%; fig. 3) and per-gene ΔlnL values are nearly all very close to zero (fig. 4), suggesting little distinguishing capacity among individual loci in this data set. As such, the placement of tree shrews remains uncertain. Efforts to rigorously resolve this relationship will likely require an extraordinary amount of sequence data with just the right substitution rate (Lanier et al. 2014) combined with very careful exploration of phylogenetic signal and noise.

Within Laurasiatheria, our analyses consistently placed Eulipotyphla as the sister to all other members. However, among the remaining five orders, matters were more difficult. Our topological inferences consistently identified Euungulata (Perissodactyla + Cetartiodactyla), but this conclusion was strongly supported only by concatenation (fig. 1). At this node, ASTRAL and ASTRID had BSs of only 54 and 71, respectively (fig. 2A), whereas SVDquartets placed Perissodactyla in a polytomy with Chiroptera, Ferae, and Cetartiodactyla. Our reduced data set provided far less resolution regarding Perissodactyla (supplementary fig. S2, Supplementary Material online), suggesting that, at least in this case, expanded taxon sampling improved resolution. A few more gene trees were consistent with Euungulata than other hypotheses, and they had higher mean BS values (fig. 3). Consideration of ΔlnL values revealed that very few loci favored a sister relationship with Carnivora, Ferae, or a clade containing Cetartiodactyla, Ferae, and Chiroptera (fig. 4). However, these values also showed that no loci strongly favored Euungulata over Perissodactyla + Chiroptera, but a few loci provide modest support for Chiroptera as the sister group (fig. 4). It is somewhat surprising that our concatenated analysis supported Euungulata given that most loci had negative ΔlnL values for the comparison with Perissodactyla + Chiroptera. Perhaps these values explain how McCormack et al. (2012) placed Perissodactyla as sister to Chiroptera when using concatenation, but not in coalescent analyses. The few loci that stand out with negative ΔlnL values (i.e., favoring Perissodactyla + Chiroptera) in figure 4 may have exerted strong influence on McCormack et al.'s (2012) concatenated analysis, as has been noted in other phylogenomic data sets (Brown and Thomson 2016). Similar to our other analyses, the AU test rejected placements of Perissodactyla as sister to Carnivora, Ferae, and a clade containing Cetartiodactyla, Chiroptera, and Ferae. However, the AU test failed to reject a sister relationship to Chiroptera. Overall, our results clearly support a sister relationship of Perissodactyla to either Cetartiodactyla or Chiroptera, but not the other relationships that have been put forth (table 1). Between these two remaining plausible hypotheses, we favor Euungulata because it was identified by all phylogenetic inferences that produced a bifurcation (concatenation, ASTRAL, and ASTRID).

Although we cannot claim to have finally resolved the remaining recalcitrant interordinal relationships among placental mammals, our results have increased our confidence in the Atlantogenata hypothesis and narrowed the focus on the relationships of Scandentia and Perissodactyla. The influence of taxon sampling appears to be somewhat unpredictable. Improved taxon sampling had no obvious effect on our results for the earliest divergence among placentals, it reduced consistency among our results for Scandentia, and it improved consistency among analyses for Perissodactyla. This seemingly implies that all of the many publications that have argued that better taxon sampling is important (Zwickl et al. 2002), not necessary (Rosenberg and Kumar 2001), or a potential detriment (Rokas and Carroll 2005), are, in fact, correct in some circumstances.

The general agreement of our summary coalescent methods in producing largely the same topology as concatenation suggests that these methods can infer species trees over deep evolutionary time and on very short branches, at least as well as concatenation performs under these circumstances. Even when the underlying gene trees are highly variable and not strongly supported, our topological inferences mostly agreed. Although we do not have a means of quantifying how much gene tree variation is due to ILS versus estimation errors, it is reassuring that our approaches were so consistent. We identified only one case of strong conflict between concatenation and a summary coalescent analysis. Future analyses that fully use sequence data and coestimate species trees and gene trees in a multispecies coalescent framework (e.g., *BEAST) may stand a better chance than summary coalescent methods of resolving the most difficult nodes, but these approaches are currently intractable with large data sets. Now that DNA sequence alignments are routinely enormous, computational advances and, in some instances, greater taxon sampling may hold the real promise of refining the tree of life.

Data Availability

Data associated with this manuscript are available under BioProject PRJNA390442, available at http://www.ncbi.nlm.nih.gov/bioproject/390442, last accessed September 1, 2017.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank the following institutions and their staff for providing tissue samples: Museum of Texas Tech University, Field Museum of Natural History, University of Kansas Biodiversity Institute, South Australia Museum, University of Alaska Museum of the North, Museum of Southwestern Biology, Museum of Vertebrate Zoology, Sam Noble Museum, Louisiana State University Museum of Natural Science, University of Washington Burke Museum, and Australian National Wildlife Collection. This research was funded by the National Science Foundation (DEB-1441634, DEB-1343517, DEB-1242260) and startup funds from LSU to B.C.F. Portions of this research were conducted with high performance computing resources at LSU (http://www.hpc.lsu.edu, last accessed September 1, 2017). We thank D. Descouens, C. Huh, D. Liao, C. Rebler, V. Smith, S. Werning, and M. Yrayzoz for silhouette images from phylopic.org under a Creative Commons license and two anonymous reviewers for their helpful suggestions.

Literature Cited

Bankevich A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comp Biol. 19:455–477.

Bininda-Emonds ORP, et al. 2007. The delayed rise of present-day mammals. Nature 446:507–512.

Blom MPK, Bragg JG, Potter S, Moritz C. 2017. Accounting for uncertainty in gene tree estimation: summary-coalescent species tree inference in a challenging radiation of Australian lizards. Syst Biol. 66:352–366.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Brown JM, Thomson RC. 2016. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. Syst Biol. 66(4):517–530.

Cao J, Yang E-B, Su J-J, Li Y, Chow P. 2003. The tree shrews: adjuncts and alternatives to primates as models for biomedical research. J Med Primatol. 32(3):123–130.

Cao Y, Adachi J, Janke A, Pääbo S, Hasegawa M. 1994. Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. J Mol Evol. 39(5):519–527.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17(4):540–552.

Chen CTL, Wang JC, Cohen BA. 2007. The strength of selection on ultraconserved elements in the human genome. Am J Hum Genet. 80(4):692–704.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30(23):3317–3324.

Chifman J, Kubatko L. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J Theor Biol. 374:35–47.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol. 24(6):332–340.

Doronina L, et al. 2017. Speciation network in Laurasiatheria: retrophylogeneomic signals. Gen Res. 27(6):997–1003.

dos Reis M, et al. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc R Soc Lond [Biol] 279(1742):3491–3500.

Edwards SV, Liu L, Pearl DK. 2007. High-resolution species trees without concatenation. Proc Natl Acad Sci U S A. 104(14):5936–5941.

Edwards SV, et al. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol Phylogenet Evol. 94(Pt A):447–462.

Faircloth BC. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32(5):786–788.

Faircloth BC, et al. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol. 61(5):717–726.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool. 27(4):401–410.

Foley NM, Springer MS, Teeling EC. 2016. Mammal madness: is the mammal tree of life not yet resolved? Philos Trans R Soc B. 371(1699):20151040.

Gatesy J, Springer MS. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol Phylogenet Evol. 80:231–266.

Giarla TC, Esselstyn JA. 2015. The challenges of resolving recent, rapid radiations: empirical and simulated phylogenomics of Philippine shrews. Syst Biol. 64(5):727–740.

Glenn TC, et al. 2016. Adapterama I: universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). bioRxiv 049114.

Hallström BM, Janke A. 2010. Mammalian evolution may not be strictly bifurcating. Mol Biol Evol. 27:2804–2816.

Hallström BM, Kullberg M, Nilsson MA, Janke A. 2007. Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. Mol Biol Evol. 24:2059–2068.

Heath TA, Hedtke SM, Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J Syst Evol. 46:239–257.

Hedtke S, Townsend TM, Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst Biol. 55(3):522–529.

Hendy MD, Penny D. 1989. A framework for the quantitative study of evolutionary trees. Syst Zool. 38(4):297–309.

Huelsenbeck JP. 1995. The performance of phylogenetic methods in simulation. Syst Biol. 44(1):17–48.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23(2):254–267.

Jakob SS, Blattner FR. 2006. A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. Mol Biol Evol. 23(8):1602–1612.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Koblmüller S, Egger B, Sturmbauer C, Sefc KM. 2010. Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the

endemic Lake Tanganyika cichlid tribe Tropheini. Mol Phylogenet Evol. 55(1):318–334.
Kozlov AM, Aberer AJ, Stamatakis A. 2015. ExaML version 3: a tool for phylogenomic analyses on supercomputers. Bioinformatics 31(15):2577–2579.
Lanier HC, Huang H, Knowles LL. 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Mol Phylogenet Evol. 70:112–119.
Lanier HC, Knowles LL. 2015. Applying species-tree analyses to deep phylogenetic histories: challenges and potential suggested from a survey of empirical phylogenetic studies. Mol Phylogenet Evol. 83:191–199.
Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. 7(Suppl 1):S4.
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 21(6):1095–1109.
Mason VC, et al. 2016. Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates. Sci Adv. 2(8):e1600633.
Matthew WD, Simpson GG. 1943. Relationships of the orders of mammals. J Mamm. 24(3):304–311.
McCormack JE, et al. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 22:746–754.
McKenna MC. 1975. Toward a phylogenetic classification of the Mammalia. In: Luckett WP, Szalay FS, editors. Phylogeny of the primates. New York: Plenum Press. p. 21–46.
Meredith RW, et al. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334(6055):521–524.
Morgan CC, et al. 2013. Heterogeneous models place the root of the placental mammal phylogeny. Mol Biol Evol. 30(9):2145–2156.
Mirarab S, et al. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30(17):i541–i548.
Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31(12):i44–i52.
Miyamoto MM, Goodman M. 1986. Biomolecular systematics of eutherian mammals: phylogenetic patterns and classification. Syst Biol. 35(2):230–240.
Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny Gen Res. 17:413–421.
Murphy WJ, et al. 2001a. Molecular phylogenetics and the origins of placental mammals. Nature 409(6820):614–618.
Murphy WJ, et al. 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294:2348–2351.
Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. Proc Natl Acad Sci U S A. 103(26):9929–9934.
Novacek MJ. 1980. Cranioskeletal features in tupaiids and selected Eutheria as phylogenetic evidence. In: Luckett WP, editor. Comparative biology and evolutionary relationships of tree shrews. New York: Plenum Press. p. 35–93.
Novacek MJ. 1992. Mammalian phylogeny: shaking the tree. Nature 356(6365):121–125.
O'Leary MA, et al. 2013. The placental mammal ancestor and the post K-Pg radiation of placentals. Science 339:662–667.
Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2010. How many bootstrap replicates are necessary? J Comp Biol. 17(3):337–354.
Perelman P, et al. 2011. A molecular phylogeny of living primates. PLoS Genet. 7(3):e1001342.
Philippe H, et al. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9(3):e1000602.
Porter CA, Goodman M, Stanhope MJ. 1996. Evidence on mammalian phylogeny from sequences of exon 28 of the von Willebrand factor gene. Mol Phylogenet Evol. 5:89–101.
R Core Team. 2013. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
Roberts TE, Lanier HC, Sargis EJ, Olson LE. 2011. Molecular phylogeny of treeshrews (Mammalia: Scandentia) and the timescale of diversification in Southeast Asia. Mol Phylogenet Evol. 60(3):358–372.
Rohland N, Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22(5):939–946.
Rokas A, Carroll SB. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol Biol Evol. 22(5):1337–1344.
Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. Mol Biol Evol. 30(9):2134–2144.
Rosenberg MS, Kumar S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc Natl Acad Sci U S A. 98(19):10751–10756.
Scornavacca C, Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. Syst Biol. 66(1):112–120.
Seo TK. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol Biol Evol. 25(5):960–971.
Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat Ecol Evol. 1(5):126.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51(3):492–508.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17(12):1246–1247.
Simpson GG. 1959. Mesozoic mammals and the polyphyletic origin of mammals. Evolution 13(3):405–414.
Snir S, Rao S. 2012. Quartet MaxCut: a fast algorithm for amalgamating quartet trees. Mol Phylogenet Evol. 62(1):1–8.
Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and multispecies coalescent model. Proc Natl Acad Sci U S A. 109(37):14942–14947.
Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. 2003. Placental mammal diversification and the Cretaceous–Tertiary boundary. Proc Natl Acad Sci U S A. 100(3):1056–1061.
Springer MS, Stanhope MJ, Madsen O, De Jong WW. 2004. Molecules consolidate the placental mammal tree. Trends Ecol Evol. 19(8):430–438.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.
Stanhope MJ, et al. 1996. Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. J Mol Evol. 43(2):83–92.
Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. Zoolog Scripta 45:50–62.
Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26(12):1569–1571.
Swofford DL. 2003. PAUP*. Phylogenetic analysis using parsimony and other methods. Version 4. Sinauer Associates, Sunderland, Massachusetts.
Tarver JE, et al. 2016. The interrelationships of placental mammals and the limits of phylogenetic inference. Genome Biol Evol. 8(2):330–344.

Teeling EC, Hedges SB. 2013. Making the impossible possible: rooting the tree of placental mammals. Mol Biol Evol. 30:1999–2000.
Vachaspati P, Warnow T. 2015. ASTRID: Accurate Species TRees from Internode Distances. BMC Genomics 16(Suppl 10):S3.
Wildman DE, et al. 2007. Genomics, biogeography, and the diversification of placental mammals. Proc Natl Acad Sci U S A. 282:20141013.
Xu X, Janke A, Arnason U. 1996. The complete mitochondrial DNA sequence of the greater Indian rhinoceros, *Rhinceros unicornis*, and the phylogenetic relationship among Carnivora, Perissodactyla, and Artiodactyla (+ Cetacea). Mol Biol Evol. 13(9):1167–1173.
Yang Z. 1998. On the best evolutionary rate for phylogenetic analysis. Syst Biol. 47(1):125–133.
Yang Z-F, et al. 2013. The tree shrew provides a useful alternative model for the study of influenza H1N1 virus. Virol J. 10:111.
Zhou X, et al. 2012. Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the laurasiatherian mammals. Syst Biol. 61(1) 150–164.
Zwickl DJ, Hillis DM, Crandall K. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst Biol. 51(4):588–598.

Associate editor: Davide Pisani