The gene tree delusion ☆

Mark S. Springer *, John Gatesy *

Department of Biology, University of California, Riverside, CA 92521, USA

ARTICLE INFO

ABSTRACT

Higher-level relationships among placental mammals are mostly resolved, but several polytomies remain contentious. Song et al. (2012) claimed to have resolved three of these using shortcut coalescence methods (MP-EST, STAR) and further concluded that these methods, which assume no within-locus recombination, are required to unravel deep-level phylogenetic problems that have stymied concatenation. Here, we reanalyze Song et al.'s (2012) data and leverage these re-analyses to explore key issues in systematics including the recombination ratchet, gene tree stoichiometry, the proportion of gene tree incongruence that results from deep coalescence versus other factors, and simulations that compare the performance of coalescence and concatenation methods in species tree estimation. Song et al. (2012) reported an average locus length of 3.1 kb for the 447 protein-coding genes in their phylogenomic dataset, but the true mean length of these loci (start codon to stop codon) is 139.6 kb. Empirical estimates of recombination breakpoints in primates, coupled with consideration of the recombination ratchet, suggest that individual coalescence genes (c-genes) approach ∼12 bp or less for Song et al.'s (2012) dataset, three to four orders of magnitude shorter than the c-genes reported by these authors. This result has general implications for the application of coalescence methods in species tree estimation. We contend that it is illogical to apply coalescence methods to complete protein-coding sequences. Such analyses amalgamate c-genes with different evolutionary histories (i.e., exons separated by >100,000 bp), distort true gene tree stoichiometry that is required for accurate species tree inference, and contradict the central rationale for applying coalescence methods to difficult phylogenetic problems. In addition, Song et al.'s (2012) dataset of 447 genes includes 21 loci with switched taxonomic names, eight duplicated loci, 26 loci with non-homologous sequences that are grossly misaligned, and numerous loci with >50% missing data for taxa that are misplaced in their gene trees. These problems were compounded by inadequate tree searches with nearest neighbor interchange branch swapping and inadvertent application of substitution models that did not account for among-site rate heterogeneity. Sixty-six gene trees imply unrealistic deep coalescences that exceed 100 million years (MY). Gene trees that were obtained with better justified models and search parameters show large increases in both likelihood scores and congruence. Coalescence analyses based on a curated set of 413 improved gene trees and a superior coalescence method (ASTRAL) support a Scandentia (treeshrews) + Glires (rabbits, rodents) clade, contradicting one of the three primary systematic conclusions of Song et al. (2012). Robust support for a Perissodactyla + Carnivora clade within Laurasiatheria is also lost, contradicting a second major conclusion of this study. Song et al.'s (2012) MP-EST species tree provided the basis for circular simulations that led these authors to conclude that the multispecies coalescent accounts for 77% of the gene tree conflicts in their dataset, but many internal branches of their MP-EST tree are stunted by an order of magnitude or more due to wholesale gene tree reconstruction errors. An independent assessment of branch lengths suggests the multispecies coalescent accounts for ⩽15% of the conflicts among Song et al.'s (2012) 447

gene trees. Unfortunately, Song et al.'s (2012) flawed phylogenomic dataset has been used as a model for additional simulation work that suggests the superiority of shortcut coalescence methods relative to concatenation. Investigator error was passed on to the subsequent simulation studies, which also incorporated further logical errors that should be avoided in future simulation studies. Illegitimate branch length switches in the simulation routines unfairly protected coalescence methods from their Achilles' heel, high gene tree reconstruction error at short internodes. These simulations therefore provide no evidence that shortcut coalescence methods out-compete concatenation at deep timescales. In summary, the long c-genes that are required for accurate reconstruction of species trees using shortcut coalescence methods do not exist and are a delusion. Coalescence approaches based on SNPs that are widely spaced in the genome avoid problems with the recombination ratchet and merit further pursuit in both empirical systematic research and simulations.

1. Introduction

Interordinal relationships among placental mammals have been the topic of debate for more than a century. Variations of this debate have focused on morphology versus molecules, the importance of fossils, taxon sampling and long branches, as well as differences between trees based on nuclear versus mitochondrial genes (D'Erchia et al., 1996; Sullivan and Swofford, 1997; Shoshani and McKenna, 1998; Madsen et al., 2001; Murphy et al., 2001a,b; Scally et al., 2001; Arnason et al., 2002, 2008; Lin et al., 2002; Asher et al., 2003; Reyes et al., 2004; Springer et al., 2004, 2005, 2007; Springer and Murphy, 2007; Meredith et al., 2011; O'Leary et al., 2013). Most camps are now in agreement that there are four major clades of placental mammals: Afrotheria, Xenarthra, Euarchontoglires, and Laurasiatheria (Murphy et al., 2001b; Springer et al., 2003; Bininda-Emonds et al., 2007; Wildman et al., 2007; Meredith et al., 2011; dos Reis et al., 2012; O'Leary et al., 2013). However, several polytomies remain contentious including the placental root, the paenungulate trichotomy, the position of treeshrews (Scandentia) within Euarchontoglires, and relationships among the laurasiatherian orders Carnivora, Perissodactyla, Cetartiodactyla, and Chiroptera (Kriegs et al., 2006; Nishihara et al., 2006, 2009; Janečka et al., 2007; Murphy et al., 2007; Hallström and Janke, 2010; Hallström et al., 2011; Meredith et al., 2011; dos Reis et al., 2012; McCormack et al., 2012; Nery et al., 2012; Xu et al., 2012; Morgan et al., 2013; O'Leary et al., 2013; Romiguier et al., 2013) (Fig. 1). Recently, advocates of shortcut coalescence methods such as STAR (Liu et al., 2009a,b) and MP-EST (Liu et al., 2010) have claimed that these methods provide robust support for several of these longstanding problems in placental phylogenetics (McCormack et al., 2012; Song et al., 2012; Kumar et al., 2013). If these claims are correct, then shortcut coalescence methods have fundamentally advanced our understanding of placental mammal phylogeny. However, it would be premature to accept these claims without critical analysis from other workers in the field, as occurred with the community's general acceptance of the four major clades of placental mammals.

Along these lines, Gatesy and Springer (2014) provided a detailed critique of McCormack et al.'s (2012) study that called attention to fundamental problems with their data and analyses including a paucity of informative sites at most UCE (ultraconserved element) loci, highly inaccurate gene trees with arbitrary bifurcations, and unrealistic deep coalescences that exceed 100 MY in over 43% of the gene trees. Gatesy and Springer (2014) also highlighted problems with the application of shortcut coalescence methods to 'deep' phylogenetic problems. However, an in depth critical analysis of Song et al.'s (2012) claims has lagged behind. Song et al.'s (2012) study is one of the driving forces for the application of shortcut coalescence methods to deep phylogenetic problems in the Tree of Life, and their dataset provides the basis for recent simulations that suggest coalescence methods are sometimes more accurate than concatenation for species tree reconstruction at deep timescales (Mirarab et al., 2014a,b,c; Zimmermann et al., 2014). Roch and Warnow (2015) reiterated this conclusion based on Mirarab et al.'s (2014a,b,c) simulations. It is therefore imperative to critically evaluate Song et al.'s (2012) data, analyses, and conclusions.

Song et al. (2012) analyzed a phylogenomic dataset for mammals (447 loci, 37 taxa) with both concatenation and coalescence methods, and claimed that Bayesian and maximum likelihood (ML) analyses of the concatenated data yielded strongly incongruent results with subsampling of loci and taxa, whereas analyses with two shortcut coalescence methods, STAR (Liu et al., 2009a,b) and MP-EST (Liu et al., 2010), yielded a phylogenetic hypothesis that is robust to variable gene and taxon sampling. Song et al. (2012) also reanalyzed Meredith et al.'s (2011) family-level dataset for Mammalia (26 genes, 169 taxa) with coalescence methods and concluded that 400 genes, not 26, would be required to achieve a species tree dominated by high bootstrap support for this large sampling of taxa. Based on their results, Song et al. (2012) argued that directly accounting for incomplete lineage sorting (ILS) is of critical importance for resolving ancient radiations and claimed that their results "firmly place the eutherian root between Atlantogenata and Boreoeutheria and support ungulate polyphyly and a sister-group relationship between Scandentia and Primates" (p. 14942). These authors further claimed their study "demonstrates that the incongruence introduced by concatenation methods is a major cause of long-standing uncertainty in the phylogeny of placental mammals" (p. 14942), and suggested that this problem can be solved with phylogenomic data and shortcut coalescence methods.

Here, we examine ramifications of the recombination ratchet for c-gene size and develop the concept of gene tree stoichiometry that has been ignored in most of the systematics literature (Section 3.1), characterize fundamental errors in Song et al.'s (2012) phylogenomic data set (447 genes) and analyses that undercut all of their major conclusions (Section 3.2), highlight critical flaws of recent simulation studies (Mirarab et al., 2014a,b,c; Zimmermann et al., 2014; Roch and Warnow, 2015) that are underpinned by Song et al.'s (2012) dataset (Section 3.3), discuss the gene tree delusion that results from the application of coalescence methods to unrealistically long, "pseudo" c-genes (Section 3.4), and offer a prospectus on the future of coalescence and concatenation methods for species tree construction at deep divergences in the Tree of Life (Section 3.5). Song et al.'s (2012) reanalysis of Meredith et al.'s (2011) dataset is addressed elsewhere (Gatesy, Meredith, Springer, unpublished).
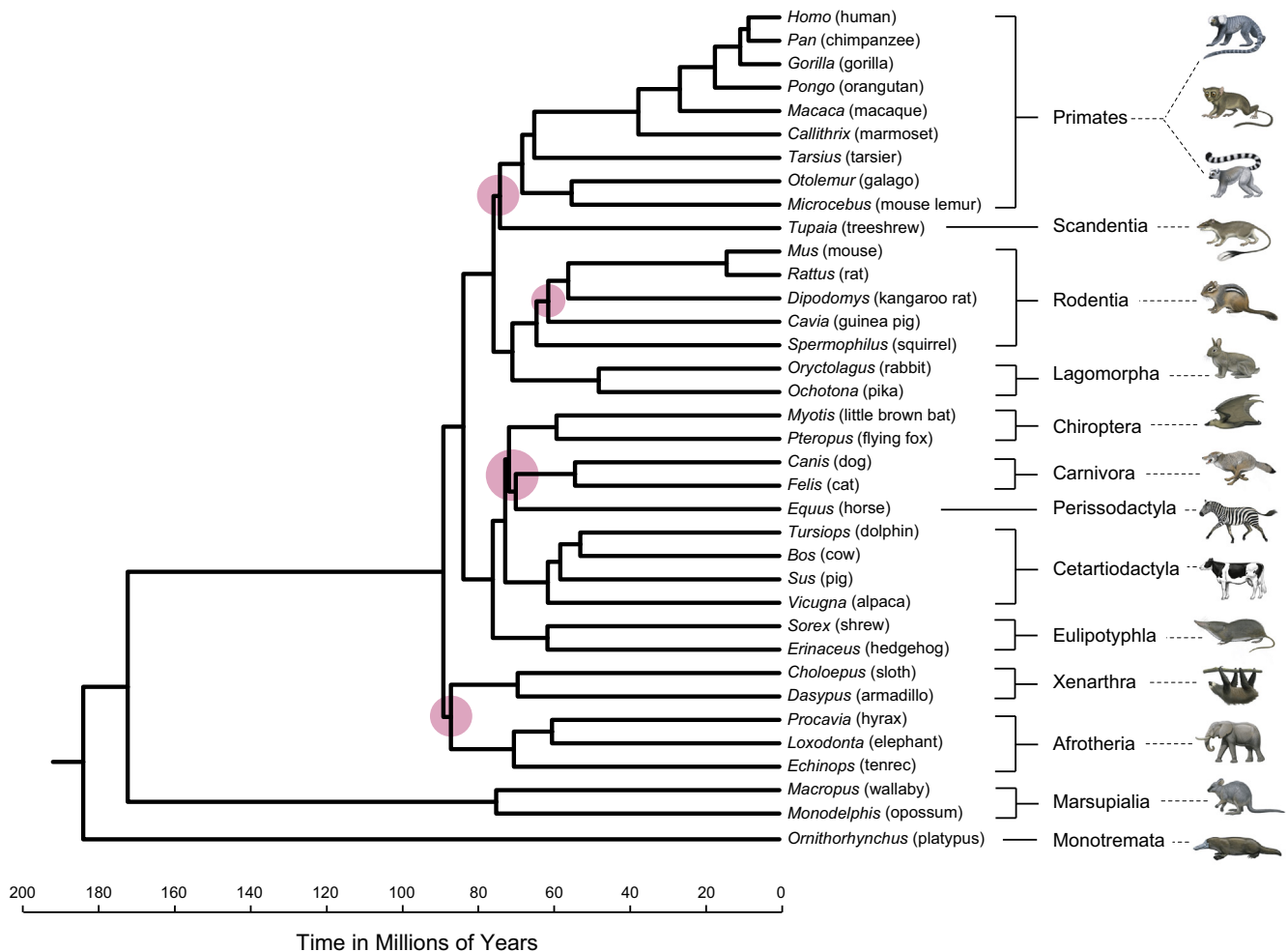
Fig. 1. Timetree from dos Reis et al. (2012) with branches in millions of years. Higher-level taxa are delimited to the right by brackets. Controversial nodes on this tree are highlighted with pink circles. Paintings of mammals are by C. Buell. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Materials and methods

2.1. Calculation of c-gene size for primate data

Hobolth et al. (2007) estimated the mean length of fragments that support four different resolutions of the human–chimp–gorilla trichotomy for four autosomal regions of the genome that have different recombination rates. The four resolutions of the trichotomy are human + chimp without deep coalescence (HC1), human + chimp with deep coalescence (HC2), chimp + gorilla (CG), and human + gorilla (HG). Hobolth et al. (2007) also estimated the proportion of each autosomal region that resides in each of the four states (i.e., HC1, HC2, CG, HG). Based on this information we calculated the total number of fragments and the mean length of these fragments for each of Hobolth et al.'s (2007) targeted regions. These calculations were extrapolated to regions of the genome that are 139.6 kb, which is the mean length of Song et al.'s (2012) 447 loci from start codon to stop codon. We also extended these calculations to illustrate the effect of the recombination ratchet on c-gene size. All calculations and formulas are provided in Table 1.

2.2. Phylogenetic analyses

PhyML searches with different search parameters (nearest neighbor interchange [NNI] branch swapping, subtree pruning and regrafting [SPR] branch swapping, best of NNI + SPR branch swapping) and different models of sequence evolution (e.g., HKY, GTR, GTR + Γ) were employed to replicate Song et al.'s (2012) 447 gene trees. As in Song et al. (2012), a single tree-search was executed in each analysis (i.e., multiple starting points for branch swapping were not utilized). We used both PhyML 2.4.4 (Guindon and Gascuel, 2003) and PhyML 3.0 (Guindon et al., 2010). PhyML 3.0 allows for NNI, SPR, or best of NNI + SPR searches, whereas PhyML 2.4.4 only implements NNI branch swapping. RAxML version 8.1.11 searches were performed with 200 bootstrap replications and a search for the best ML tree all in the same run on CIPRES (Miller et al., 2010). We employed a GTR + Γ model of sequence evolution with GTRGAMMA in RAxML for both the bootstrapping phase and the search for the best tree.

2.3. Shortcut coalescence analyses

Shortcut coalescence analyses were performed with 447 gene trees from Song et al. (2012). Shortcut coalescence analyses also were run on three additional sets of gene trees that were based on the 447 gene alignment from Song et al. (2012): 424 RAxML gene trees from Mirarab et al. (2014a,b), 413 gene trees estimated here (PhyML) following correction of various errors in the dataset of Song et al. (2012), and 413 gene trees (RAxML) that account for these same errors. MP-EST analyses were performed with MP-EST 1.4 (Liu et al., 2010) and ten searches for the optimal MP-EST tree

Table 1
Calculation of the minimum number of genomic fragments with different histories (and recombination boundaries) that are expected to occur in representative regions of the genome that are 139.6 kb in length, which is the average size of Song et al.'s (2012) "pseudo c-genes" from start codon to stop codon.

A. Target #	B. Size of target (base-pairs)[a]	C. Fraction in HC1[a]	D. Fraction in HC2, CG, or HG[a]	E. Base-pairs in HC1 = column B × column C	F. Base-pairs in HC2, CG, or HG = column B × column D	G. Mean size of HC1 fragments (base-pairs)[a]	H. Mean size of HC2, CG, or HG fragments (base-pairs)[a]	I. Total number of HC1 fragments = column E/column G	J. Total number of HC2, CG, or HG fragments = column F/column H	K. Total number of fragments = column I + (3 × column J)	L. Mean size for all fragments (base-pairs) = column B/column K	M. Number of fragments for 139.6 kb region = 139,000/column L
1	1,255,492	0.48	0.17	602636.16	213433.64	1684	65	357.8599525	3283.594462	10208.64334	122.9832367	1135.114051
106	257,871	0.51	0.16	131514.21	41259.36	2710	41	48.52922878	1006.325854	3067.50679	84.0653396	1660.61305
121	230,666	0.23	0.26	53053.18	59973.16	2469	81	21.48771972	740.4093827	2242.715868	102.8511919	1357.300752
122	92,240	0.47	0.18	43352.8	16603.2	532	65	81.49022556	255.4338462	847.791764	108.800302	1283.084673 Mean of four regions = 1359.0 fragments per 139.6 kb

Calculations are based on primate data from Hobolth et al. (2007) for a three-taxon problem (human = H, chimp = C, gorilla = G) and ignore branch length heterogeneity except for HC1 versus HC2 (e.g., branch length heterogeneity within HC1 is ignored). The mean number of fragments for a 139.6 kb fragment will increase with more taxa owing to the recombination ratchet. The number of recombination boundaries is equal to the number of fragments − 1, e.g., 1359 − 1 = 1358 recombination boundaries for a 139.6 kb region of the genome using the mean number of fragments from Hobolth et al.'s (2007) four targets. Given eight additional internodes on Song et al.'s (2012) species tree that are similar to the human–chimp–gorilla trichotomy, the recombination ratchet predicts that there will be approximately 9 × 1358 = 12,222 recombination events across the mammal tree for a 139.6 kb region, although some of these recombination events will be superimposed at the same recombination boundary. A 139,600 bp region contains 139,599 potential recombination sites. The probability that a potential recombination site will be hit by one or more of 12,222 independent recombination events is = 1 − (139,598/139,599)^{12,222} = 0.0842, and the number of recombination boundaries will therefore be equal to 139,599 × 0.0842 = 11,760. The mean size of segments that are separated by recombination boundaries will be equal to 139,600/11,760 = 11.9 bp.

[a] Data from Hobolth et al. (2007).

after recompiling MP-EST 1.4 with more robust search parameters than in the default version ('MAX_ROUNDS' changed from 10,000,000 to 20,000,000; 'NUM_NOCHANGE' changed from 20,000 to 100,000). All gene trees were rooted with *Gallus gallus* (chicken). We also performed MP-EST bootstrap analyses (100 pseudoreplicates) with PhyML gene trees for 447 loci from Song et al. (2012) and RAxML gene trees for our curated set of 413 loci. RAxML gene trees were taken from Mirarab et al. (2014b) except for 21 gene trees with switched taxon names that were excluded from these authors' 424 gene-tree dataset. For these 21 genes, we corrected the taxon names in Song et al.'s (2012) original dataset and ran RAxML analyses as described above to obtain optimal and bootstrap trees. Each bootstrap pseudoreplicate was analyzed with MP-EST 1.4 with the same 'MAX_ROUNDS' and 'NUM_NOCHANGE' settings as above with one search per replicate. The majority rule consensus of these 100 MP-EST bootstrap trees was then calculated with PAUP* 4.0a136 for Macintosh (X86) (Swofford, 2002). STAR analyses (Liu et al., 2009a,b) were conducted on the STRAW website (Shaw et al., 2013; http://bioinformatics.publichealth.uga.edu/SpeciesTreeAnalysis/) with the same 447, 424, and 413 gene trees as above. We also performed bootstrap analyses with Song et al.'s (2012) 100 pseudoreplicate dataset for 447 genes and our 100 pseudoreplicate dataset for 413 genes (RAxML) by running STAR on each replicate and summarizing bootstrap trees with a majority-rule consensus tree in PAUP* as above. ASTRAL analyses were performed with ASTRAL 4.7.0 (Mirarab et al., 2014a). We performed searches for the optimal trees with 447, 424, and 413 gene trees and also performed bootstrap analyses with Song et al.'s (2012) 100 pseudoreplicate dataset for 447 genes and our 100 pseudoreplicate dataset for 413 genes (RAxML) by running ASTRAL on each pseudoreplicate and summarizing bootstrap trees with a majority-rule consensus tree in PAUP*.

2.4. Tree comparison metrics, deep coalescence calculations, and simulations

RF distances (Robinson and Foulds, 1981) were obtained directly from the STRAW site for STAR analyses or with HashRF 1.0 (Sul and Williams, 2008) for MP-EST and ASTRAL. Maximum deep coalescences were calculated for each gene tree as in Gatesy and Springer (2014) using dos Reis et al.'s (2012) timetree as the species tree estimate for Mammalia. Deep coalescence is "…the failure of ancestral copies to coalesce (looking backwards in time) into a common ancestral copy until deeper than previous speciation events" (Maddison, 1997; p. 523). Deep coalescence implies the retention of ancestral polymorphism across multiple speciation events. Gene trees may show evidence for deep coalescence through topological disagreements with the species tree. If these conflicts are assumed to result from retention of ancestral polymorphisms, the various conflicting nodes in the gene tree can be used to measure the duration of deep coalescence by comparing the gene tree to the dated species tree. The maximum deep coalescence for a given gene tree is the maximum discrepancy (in years) between the age of a particular crown clade on the species tree and the age of the least inclusive clade on the gene tree that includes all of the same taxa in that crown clade. Coalescence debt is the sum of all deep coalescences, in years, on a gene tree, and was calculated by comparing all nodes on a gene tree to the species tree. We used DendroPy v3.12.0 (Sukumaran and Holder, 2010) and a script from Mirarab et al. (2014b) to simulate 1000 gene trees for each of four different species trees. The species trees were based on dos Reis et al.'s (2012) timetree for Mammalia with four different conversions of time in millions of years to coalescent units (CUs) following Patel et al. (2013), Rannala and Yang (2003), Carstens and Dewey (2010), and Hobolth et al. (2007). HashRF 1.0 was used to calculate RF distances for each set of gene

trees relative to the dos Reis et al. (2012) species tree. RF distances reported by HashRF 1.0 (Sul and Williams, 2008) are normalized and these values were doubled to make them compatible with RF distances reported by STRAW (Shaw et al., 2013).

3. Results and discussion

3.1. The incredible shrinking c-gene, the recombination ratchet, and gene tree stoichiometry

Song et al.'s (2012) sequence data were obtained from the OrthoMam database for protein-coding sequences (Ranwez et al., 2007). Song et al. (2012) reported an average locus length of 3.1 kilobases (kb) for their 447 loci. Song et al. (2012) acknowledged that intra-locus recombination "would constitute a violation of the multispecies coalescent model" that may mislead phylogenetic analysis, and tested for recombination by plotting the consistency index (CI) of each locus, which is a measure of homoplasy, against the length of each locus. The rationale for this test is that longer loci have more opportunities for recombination, and therefore increased homoplasy, relative to shorter loci. Song et al. (2012) did not find a correlation between CI and locus length and concluded that recombination is not a confounding factor for their dataset. However, Song et al.'s (2012) protein-coding sequences have been trimmed of introns and span much longer regions of the genome (Fig. 2). Song et al. (2012) reported locus lengths of 2040, 1800, and 2220 base pairs (bp) for *PRKG1*, *GALNTL6*, and *IL1RPL1*, respectively, whereas the true genic lengths for these coding sequences (with introns) are 1.30, 1.23, and 1.17 million bp. Moreover, the true mean length of Song et al.'s (2012) 447 loci is ~139.6 kb, which is 45× higher than the mean value of 3.1 kb reported by Song et al. (2012) (Fig. 2). Song et al.'s (2012) mistaken gene lengths are gross underestimates of actual gene lengths, and their conclusion that recombination is not a systematically confounding factor in their dataset is invalid.

Gatesy and Springer (2013) previously noted that coalescence analysis with distantly linked, concatenated exons is a hybrid method ("concatalescence") that ignores the basic logic of coalescence methods for inferring species trees. More specifically, concatalescence, also known as binning (Bayzid and Warnow, 2013; Mirarab et al., 2014a,b,c; Roch and Warnow, 2015), is an amalgam of supermatrix and coalescence approaches that concatenates c-genes with different histories prior to inferring a species tree with coalescence methods (Gatesy and Springer, 2013, 2014; Springer and Gatesy, 2014). Concatalescence has emerged as a popular approach for species tree inference with both plants (Xi et al., 2013, 2014; Zhong et al., 2013; Wickett et al., 2014) and animals (Hallström et al., 2011; Chiari et al., 2012; Shaw et al., 2012; Liang et al., 2013; Tsagkogeorga et al., 2013; Mirarab et al., 2014a,b,c; Zimmermann et al., 2014). Further, some authors have promoted binning methods that apply concatalescence to DNA segments from different chromosomes (Mirarab et al., 2014a,b,c).

By contrast with Song et al.'s (2012) "pseudo" c-genes that commonly span >100,000 bp (Fig. 2), empirical estimates for the size of adjacent fragments of the genome that exhibit different histories are much smaller (Fig. 3). C-genes may be even tinier if adjacent segments are topologically identical but exhibit branch length heterogeneity (Fig. 4; Edwards, 2009). For mammalian autosomal chromosomes, conflicting fragments of the genome that disagree with the inferred species tree have a mean length of <100 bp based on two empirical studies of primates (Table 1; Hobolth et al., 2007, 2011), all of which (*Homo*, *Pan*, *Gorilla*, *Pongo*, *Macaca*) were sampled in Song et al.'s (2012) dataset. These short fragments are more than three orders of magnitude (>1000×) shorter than the average locus length in Song et al.'s (2012) study (Fig. 2).
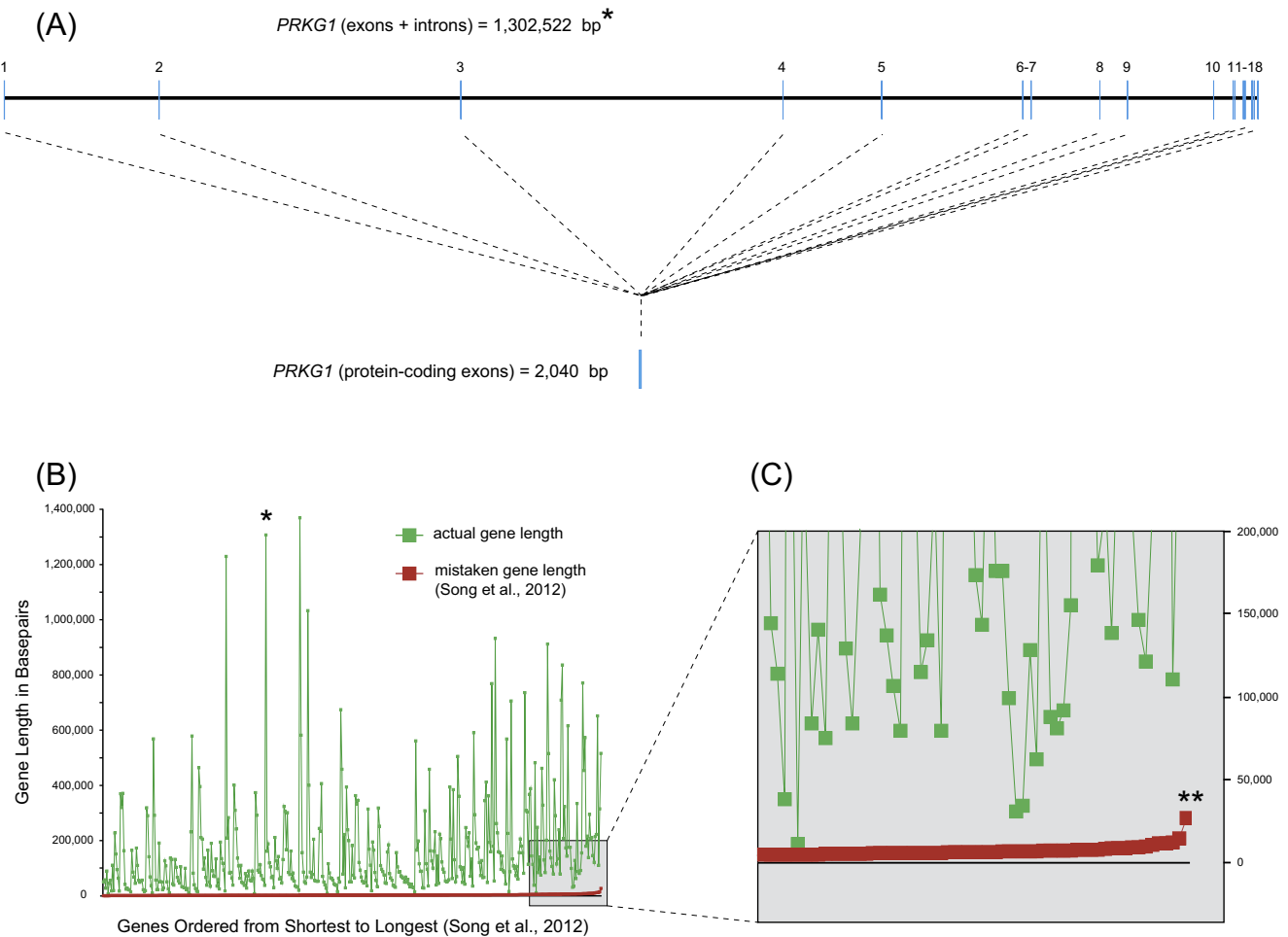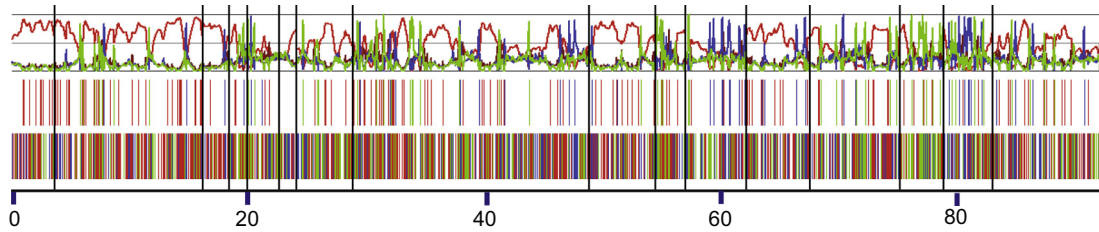
Fig. 2. Gene lengths reported by Song et al. (2012) versus true gene lengths inclusive of introns. A. *PRKG1* includes 18 short protein-coding exons (total of 2040 bp) and spans more than 1.3 megabases from start codon to stop codon in *Homo*. B. Actual (green) and mistaken (red) gene lengths for 447 genes. Song et al.'s (2012) mistaken gene lengths are arranged from shortest to longest. Single asterisk denotes the *PRKG1* gene. C. 6× enlargement of Song et al.'s (2012) longest genes. The double asterisks denote *SYNE1*, which Song et al. (2012) reported as their longest locus at 27.3 kb. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Hobolth et al. (2007) examined the human–chimp–gorilla trichotomy and showed the mean lengths of fragments that support human + chimp (without deep coalescence, HC1 of Hobolth et al., 2007) are 532, 1648, 2469, and 2710 bp for four autosomal regions of the genome that have different recombination rates (mean = 1839.75 bp) (Table 1). By contrast, the mean lengths of fragments that support alternative c-gene histories (human + gorilla [HG], chimp + gorilla [CG], human + chimp with coalescence that is deeper than the cladogenic separation of human + chimp and gorilla [HC2]) are only 65, 65, 81, and 41 bp for these same genomic regions, respectively (Fig. 3, Table 1). According to these estimates, conflicting fragments of the genome are clearly very small with only a few taxa. Hobolth et al. (2007) also estimated the percentage of each autosomal region that supports each topology. It is therefore possible to calculate the minimum number of genomic fragments with different histories (and recombination boundaries) that are expected to occur in a representative region of the genome (Fig. 3, Table 1). The average number of conflicting fragments for a 139.6 kb fragment (i.e., Song et al.'s mean gene length from start codon to stop codon; Fig. 2B) ranges from 1095 to 1634 for the four autosomal regions examined (mean of 1308). For these same autosomal regions, the mean number of fragments that support HC1 and the inferred species tree is ~51.

Thus, based on this dataset, the average number of fragments that are separated by recombination boundaries in a 139.6 kb region is 1308 + 51 = 1359 even without accounting for branch length heterogeneity within HC1, HC2, HG, and CG fragments (Table 1).

The sizes of c-gene fragments can only decrease as the number of taxa increases owing to the recombination ratchet that dictates an inverse relationship between c-gene size and the number of taxa (Fig. 3). As more taxa are added to a phylogenetic analysis, the number of recombination breakpoints can only increase. At the same time, the number of nodes that must be resolved by the inexorably shrinking c-gene ratchets upward (Gatesy and Springer, 2014). The species tree for Song et al.'s (2012) 36 mammalian taxa includes numerous short internodes that are broadly similar to the human–chimp–gorilla problem. For example, eight additional internodes for Song et al.'s (2012) topology are in the range of 1.0–3.3 MY (dos Reis et al., 2012) and are only slightly shorter or slightly longer than dos Reis et al.'s (2012) estimate for the ancestral branch of *Homo* + *Pan* (2.2 MY) (Fig. 1). If these short internal branches contribute similar levels of recombination and ILS as the human–chimp–gorilla problem, then there will be 11,760 recombination boundaries and 11,761 fragments of DNA that are separated by recombination boundaries in a 139.6 kb region of DNA when multiple hits at recombination sites are taken

(A) Estimate of Mean C-gene Size for Human, Chimp, and Gorilla : ~109 bp

(B) Recombination Ratchet for 9 Taxa and 3 Short Internal Branches

Recombination break points for A, B, C : Mean c-gene size = ~109 bp

Recombination break points for A, B, C, D, E, F: Mean c-gene size = ~55 bp

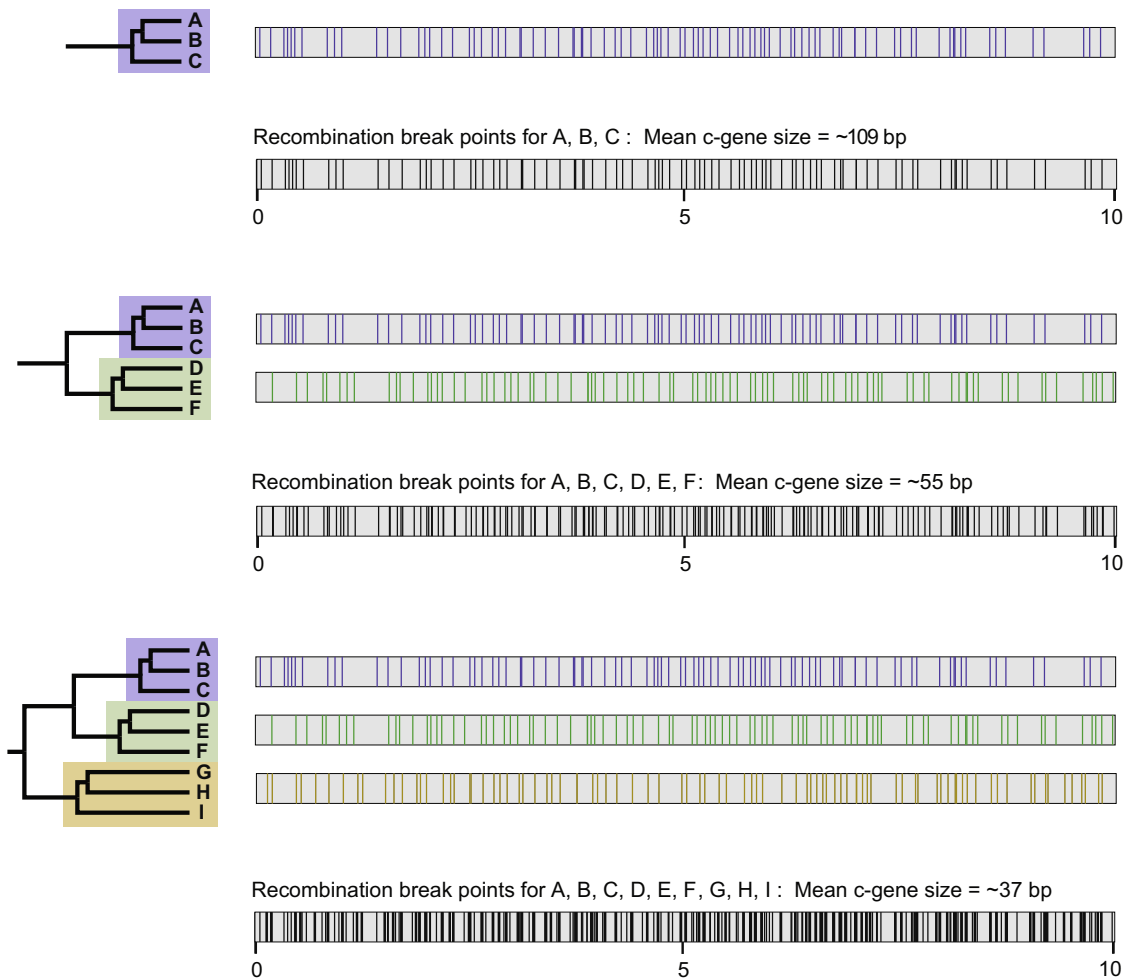Recombination break points for A, B, C, D, E, F, G, H, I : Mean c-gene size = ~37 bp

Fig. 3. Empirical and inferred estimates of c-gene sizes based on Hobolth et al.'s (2007) analysis of Target 122 (~92 kb region of human chromosome 20) for three primates (human, chimp, gorilla) and an outgroup. Estimates of c-gene size do not fully account for branch length heterogeneity, which becomes more common on long branches where there is no deep coalescence (Edwards, 2009). A. Hobolth et al.'s (2007) analysis of Target 122 for three primates (human, chimp, gorilla): site information without the outgroup (bottom), site information with the outgroup (middle), and posterior probabilities of topological histories (top). The bottom panel shows sites that are shared by human and chimp in red, sites that are shared by human and gorilla in blue, and sites that are shared by chimp and gorilla in green. In the middle panel, sites strongly supporting states HC1 and HC2 are shown in red, sites strongly supporting HG are shown in blue, and sites strongly supporting CG are shown in green. The top panel shows posterior probabilities for the HC1 (red), HC2 (dark red), HG (blue), and CG (green) topologies. B. The recombination ratchet for a topology with nine taxa and three short internal branches (also see Table 1). Recombination break points for each group of three taxa (A, B, and C; D, E, and F; G, H, and I) are modeled after Target 122 region of human, chimp, and gorilla (Hobolth et al., 2007). Mean c-gene size is reduced from ~109 bp to ~55 bp when recombination break points for ABC and DEF are mapped onto the same 10 kb segment. Mean c-gene size is reduced from ~109 bp to ~37 bp when recombination break points for ABC, DEF, and GHI are mapped onto the same 10 kb segment. The image in panel A was kindly provided by A. Hobolth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

into account (Table 1). The mean size of these fragments is ~12 bp (i.e., 139.6 kb/11,760 recombination boundaries = 11.9 bp) (Table 1), approaching the size of the basic unit of analysis in the concatenation approach, 1 bp (Gatesy and Springer, 2014). Fig. 3 illustrates this recombination ratchet for c-genes that are modeled after Hobolth et al.'s (2007) autosomal region #122 and a
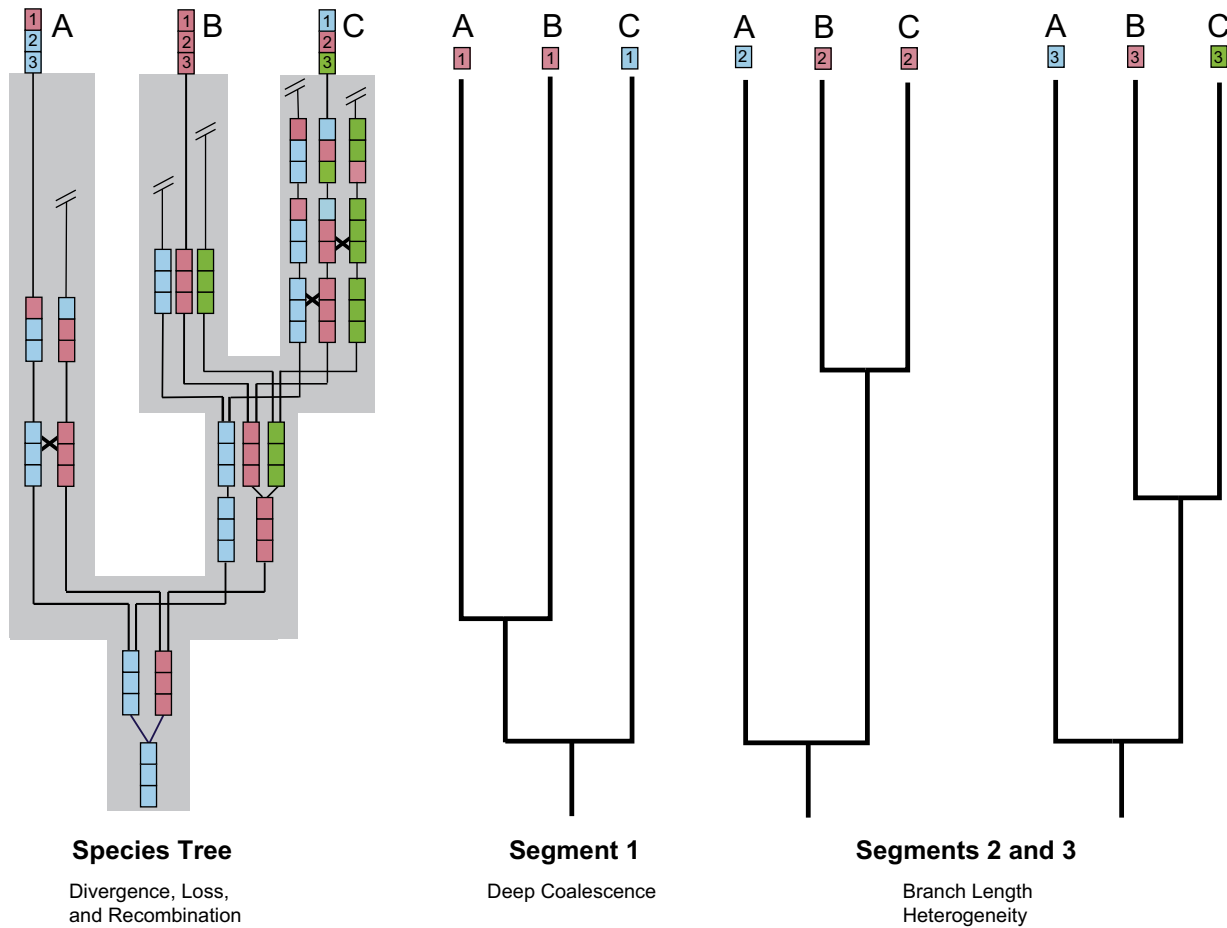
Fig. 4. Deep coalescence and branch length heterogeneity as two distinct types of gene tree heterogeneity (Edwards, 2009), both of which must be taken into account to achieve balanced gene tree stoichiometry that is required for coalescence methods. The species tree for taxa A, B, and C is shown on the left along with allelic divergence (⌇), recombination (✗), and allelic loss (⌇) at a single gene locus. Distinct histories for DNA segments 1, 2, and 3 are shown to the right. Segment 1 is topologically different from the species tree and illustrates deep coalescence. Segments 2 and 3 exhibit the same topology as the species tree, but are genealogically distinct from each other because of branch length heterogeneity.

nine-taxon phylogeny with three short internal branches. Even with just nine taxa and three short internodes, average c-gene size is only ∼37 bp.

As noted above, these calculations largely ignore branch length heterogeneity, which must be taken into account for accurate estimates of c-gene size. Branch length heterogeneity occurs when gene trees are topologically identical but have branch lengths that differ in CUs (Edwards, 2009) (Fig. 4). Unlike deep coalescence, which becomes more likely when an internal branch length on a species tree is short, branch length heterogeneity becomes more common when a branch length on a species tree is long. Edwards (2009, p. 4) admonished systematists that branch length heterogeneity is ubiquitous and is "probably the most common of all causes of gene tree heterogeneity". He further warned that branch length heterogeneity, like deep coalescence, may be a "potent source of phylogenetic inconsistency in real datasets". When internal branch lengths are sufficiently long, deep coalescence (and topological incongruence) disappears and all gene tree heterogeneity is manifested through branch length heterogeneity. Importantly, individual c-genes that reflect both types of gene tree heterogeneity (topological incongruence, branch length heterogeneity) are delimited by recombination break points (Figs. 4 and 5).

Deep coalescence and branch length heterogeneity reside at opposite ends of the branch length spectrum (Edwards, 2009). On long internal branches, deep coalescence is translated into

branch length heterogeneity. Hobolth et al.'s (2007) analysis of four autosomal regions for human, chimp, and gorilla suggests that ILS leads to topological incongruence that differs from the species tree for ∼38.5% of the genome. By contrast, almost all of this topological incongruence is instead manifested as branch length heterogeneity for human, chimp, and orangutan comparisons where only 1.4% of the genome conflicts with the species tree (Hobolth et al., 2011). The average size of conflicting fragments that support human + orangutan or chimp + orangutan is <100 bp (Hobolth et al., 2011), as is the case for conflicting fragments of the genome that support human + gorilla or chimp + gorilla (Hobolth et al., 2007). However, the translation of recombination breakpoints into branch length heterogeneity results in longer fragments that support human + chimp to the exclusion of orangutan (mean fragment length = 6.3 kb) than human + chimp to the exclusion of gorilla (mean fragment length = 1.8 kb). This does not imply that c-genes supporting human + chimp to the exclusion of orangutan are longer than c-genes supporting human + chimp to the exclusion of gorilla. Instead, individual c-genes with distinct recombination break points may be equally tiny in both cases, but these break points are not immediately obvious when adjacent c-genes exhibit branch length heterogeneity but not topological incongruence. Internal branches that are even longer, such as stem Placentalia (Fig. 1; ∼84 MY), essentially convert all deep coalescence into branch length heterogeneity, because retention of ancestral polymorphism for such long time spans is not tenable
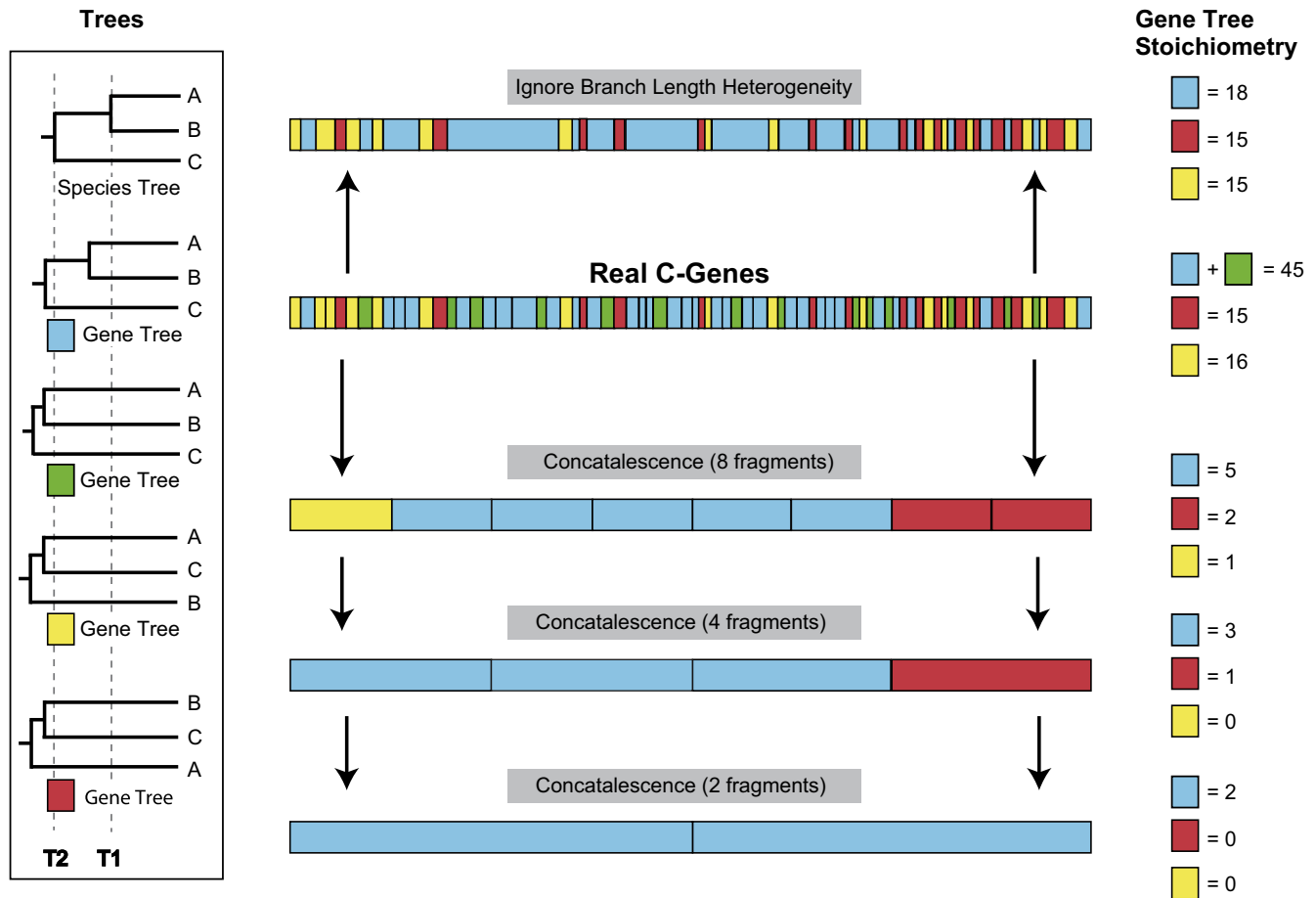
Fig. 5. Hypothetical case that illustrates true gene tree stoichiometry versus distorted gene tree stoichiometry. Balanced gene tree stoichiometry assumes a one-to-one correspondence between real c-genes and gene trees that are inferred from these c-genes. Balanced gene tree stoichiometry is distorted when branch length heterogeneity is ignored, when individual c-genes are concatenated into "pseudo" c-genes that erase the history of individual c-genes, and when gene trees (and their branch lengths) are reconstructed inaccurately. A hypothetical scenario is presented to illustrate these concepts. T1 and T2 are speciation times between A and B and A + B and C, respectively. Boxes in the "Real C-genes" panel correspond to individual c-genes that are separated by recombination boundaries. These c-genes are grouped into four different categories. Blue and green c-genes agree with the species tree, but differ in their coalescence times relative to T2; yellow and red boxes depict c-genes that are topologically incongruent with the species tree. When branch length heterogeneity is ignored, adjacent c-genes that support the same topology are merged together (green and blue fragments both support A + B and are shown in their merged states with blue boxes). The net result of ignoring branch length heterogeneity is a decrease in the relative proportion of fragments that support the species tree. Concatalescence results in an increase in the relative proportion of fragments that support the species tree, but an extreme reduction in the overall number of fragments. The often abundant errors in gene tree reconstruction that are expected for tiny c-genes are ignored in the example shown, but would lead to further distortions in true gene tree stoichiometry. Additional distortions would result from natural selection on certain loci, but not others, because natural selection violates the complete neutrality assumption of the multispecies coalescent. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

unless extreme diversifying selection is invoked (Gatesy and Springer, 2014).

Coalescence methods that give primacy to gene tree heterogeneity should recognize individual c-genes that exhibit "branch length heterogeneity" (*sensu* Edwards, 2009) as well as topological incongruence. Adjacent c-genes that are characterized by branch length heterogeneity but not topological incongruence should not be treated as a single c-gene. The recognition of individual c-genes that exhibit each type of gene tree heterogeneity, topological incongruence and branch length heterogeneity, is critical for the success of coalescence methods that depend on balanced gene tree "stoichiometry," i.e., gene trees are sampled in proportion to their true frequencies (Fig. 5). Theoretical arguments for the superiority of coalescence methods are compromised if gene tree stoichiometry is distorted and gene trees are not sampled in proportion to their true frequencies. Biases in gene tree stoichiometry are illustrated in Fig. 5, which shows recombination break points between c-genes that exhibit topological incongruence as well as branch length heterogeneity for three taxa. When branch length heterogeneity is ignored, the frequency of the most common gene tree decreases relative to the two alternate topologies.

Concatalescence, by contrast, increases the frequency of the most common gene tree relative to the alternate topologies. For a three-taxon problem that is not in the anomaly zone, the correct species tree should still be recovered even with these distortions. However, the ramifications of biased gene tree stoichiometry are less clear for more complex species trees with consecutive, short internal branches that may be in or near the anomaly zone. In the anomaly zone, concatalescence, like concatenation, would be expected to amplify the signal of the most probable gene tree, which does not match the species tree, and therefore yield an inaccurate representation of the species tree (Degnan and Rosenberg, 2006). Valid shortcut coalescence analysis of actual c-genes, which may be tiny, would be expected to yield highly inaccurate gene trees, and at deep timescales, resolution of splits in the anomaly zone presumably would be challenging as well. C-genes are the basic unit of analysis for species tree inference with coalescence methods (Maddison, 1997), but advocates of coalescence methods

for species tree construction have largely ignored this fundamental issue.

Our calculations based on empirical data suggest that recombination is rampant even for short stretches of DNA. "Pseudo" c-genes as employed by Song et al. (2012) and others (Hallström et al., 2011; Shaw et al., 2012; Kumar et al., 2013; Tsagkogeorga et al., 2013; Xi et al., 2013, 2014; Zhong et al., 2013; Mirarab et al., 2014a,b,c; Wickett et al., 2014) are concatenations of many different c-genes, especially for longer loci whose exons span hundreds of kilobases or more than a megabase of the genome (Figs. 2 and 3). This extreme concatalescence approach effectively blends together signals from conflicting segments of the genome, erasing the coalescence pattern. If the objective of coalescence methods is to decipher a species tree from a population of gene trees that represent different histories, then it is essential to use real c-genes rather than "pseudo" c-genes that lump together many different histories. If data that have been analyzed from mammalian genomes thus far are indicative (Hobolth et al., 2007, 2011), real c-genes will lack sufficient phylogenetic signal to recover accurate gene trees. Unfortunately, concatenating many c-genes to accrue greater "signal" of the coalescence pattern is nonsensical because the signal is in the correct stoichiometry of c-genes, not muddled concatenations of various c-genes with different histories (Fig. 5). The sizes of real c-genes are determined by recombination and not by the preferences of investigators who have chosen extremely long c-genes (that do not exist) that save current shortcut coalescence methods from failure relative to concatenation.

3.2. Insufficiency of the data and analyses of Song et al. (2012) for resolving the phylogeny of placental mammals

3.2.1. Vanishing support for two of Song et al.'s (2012) robustly supported clades

Problems with Song et al.'s (2012) data and analyses are not limited to the recombination ratchet and c-gene size. Even if the gigantic stretches of DNA that Song et al. (2012) utilized in their analysis (Fig. 2) do correspond to actual c-genes, their coalescence analyses would remain invalid due to a series of additional problems that are outlined below. Song et al. (2012) reported bootstrap trees for both MP-EST and STAR in their Fig. S1 for 447 genes, but did not show optimal trees based on these methods. We performed shortcut coalescence analyses with the 447 gene trees from Song et al. (2012) to determine the optimal species tree(s) with MP-EST, STAR, and also ASTRAL that is less prone to misrooting
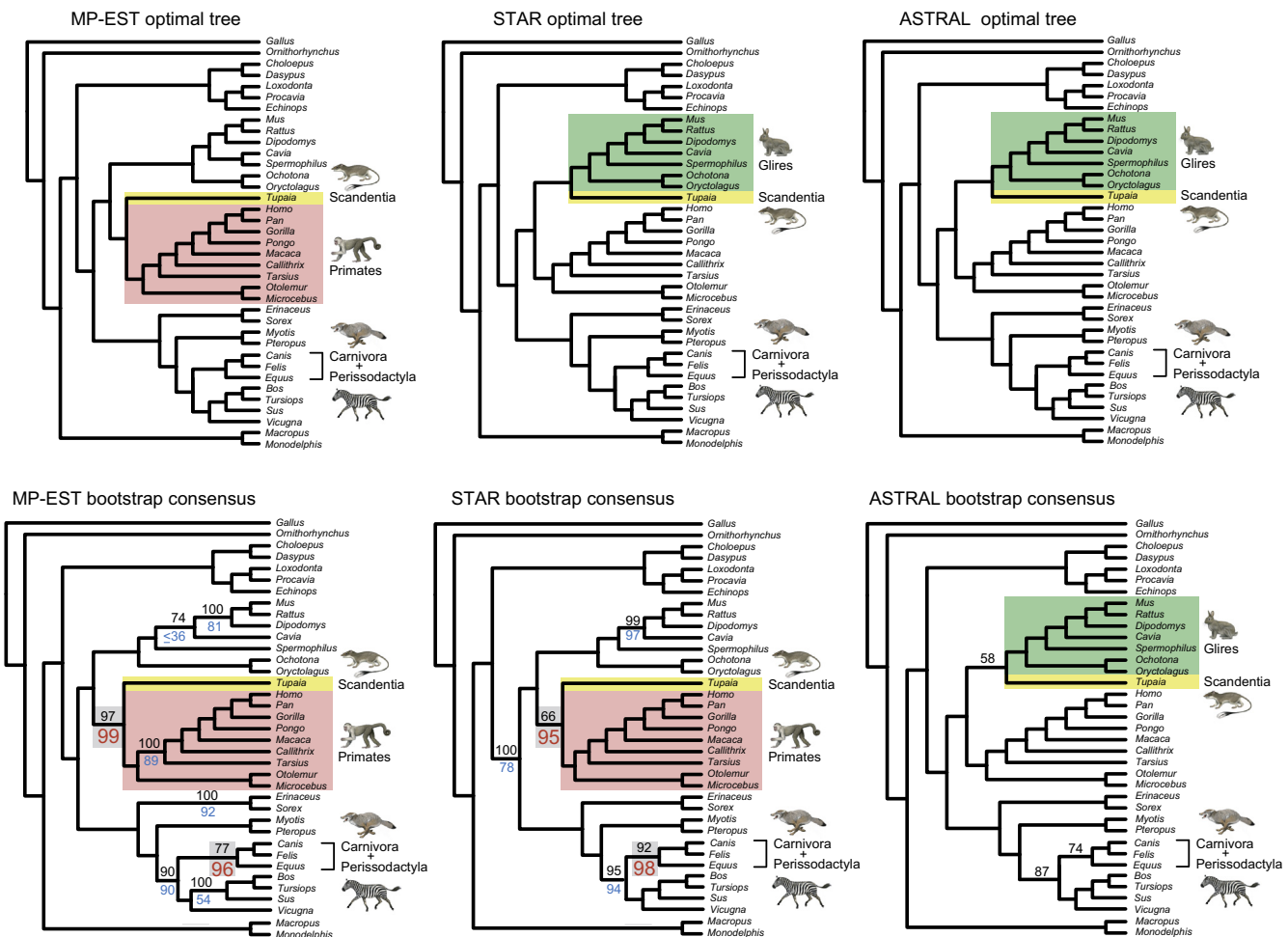

Fig. 6. Optimal and bootstrap consensus trees (MP-EST, STAR, ASTRAL) based on Song et al.'s (2012) 447 gene trees (top) and bootstrapped gene trees with 100 pseudoreplicates (bottom). Optimal trees based on STAR and ASTRAL both position *Tupaia* (treeshrew) closer to Glires than to Primates and contradict Song et al.'s (2012) assertion that coalescence methods provide robust support for an association of treeshrews and primates. Bootstrap consensus trees show support percentages (black numbers above branches) for MP-EST, STAR, and ASTRAL analyses reported here (clades without support values are supported at 100% level). Song et al. (2012) reported higher (red numbers below branches) and lower (blue numbers below branches) support scores for some clades, including scores ≥95% for two controversial clades (*Tupaia* + Primates, Perissodactyla + Carnivora) that received lower, non-significant support in our re-analyses. The ASTRAL bootstrap consensus tree provides more support for *Tupaia* + Glires than *Tupaia* + Primates, but lower support than MP-EST and STAR at controversial laurasiatherian nodes (e.g., Perissodactyla + Carnivora). Paintings of mammals are by C. Buell. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

artifacts (Simmons and Gatesy, 2015). The optimal MP-EST tree (Fig. 6) is concordant with Song et al.'s (2012, Fig. S1) MP-EST bootstrap tree. The optimal STAR tree (Fig. 6) positions Scandentia (treeshrew) as the sister taxon to Glires (rodents + rabbits) and disagrees with Song et al.'s (2012, Fig. S1) STAR bootstrap tree that positions treeshrew as the sister taxon to Primates with robust bootstrap support (95%). The optimal ASTRAL tree (Fig. 6) also supports treeshrew plus Glires. If the STAR bootstrap support at this node is correct as reported by Song et al. (2012) (but see below), this would represent a rare and disturbing phylogenetic pattern where there is >95% bootstrap support for a clade that is not found in the optimal species tree for that method.

Given these differences in results, we performed coalescence bootstrap analyses using the same sets of bootstrap gene trees utilized by Song et al. (2012). In agreement with Song et al. (2012), we recovered strong bootstrap support (97% versus 99% in Song et al., 2012) for treeshrew + Primates with MP-EST 1.4 (Fig. 6). However, we recovered *Cavia* (guinea pig) as the sister taxon to *Rattus* + *Mus* + *Dipodomys* with 74% support whereas Song et al. (2012) recovered a sister group relationship between *Cavia* and *Spermophilus* (64%). We also recovered much lower support (77%) for Perissodactyla (*Equus*) + Carnivora (*Canis* + *Felis*) than Song et al. (96%) and much higher support (100%) for dolphin (*Tursiops*) + cow (*Bos*) + pig (*Sus*) than Song et al. (54%) (Fig. 6). These large differences in bootstrap support may reflect the increased rigor of our MP-EST searches relative to those executed by Song et al. (2012). Our STAR bootstrap tree (Fig. 6) also shows important differences with Song et al.'s (2012) STAR bootstrap tree including much lower support for treeshrew + primates (66% versus 95%) and much higher support for Boreoeutheria (100% versus 78%). Finally, ASTRAL recovered treeshrew + Glires with 58% bootstrap support, an unimpressive bootstrap score for Perissodactyla + Carnivora of 74%, and a bootstrap score for Perissodactyla + Carnivora + Chiroptera that, in contrast to Song et al. (2012), drops below 90% (Fig. 6).

Song et al. (2012, p. 14946) claimed that their "study confirms Scandentia (tree shrews) as the sister group of Primates, providing a context to study early character and genome evolution in the lineages leading to primates and humans." Song et al. (2012) reported bootstrap scores of 99% and 95% for treeshrew + Primates based on MP-EST and STAR, respectively. These scores are seemingly robust as they are both ⩾95% with one score approaching 100%, but our re-analyses recovered lower bootstrap support scores of 97% and 66% with MP-EST and STAR, respectively. Further, bootstrap support for treeshrew + Primates is only 42% with ASTRAL, which has so far outperformed MP-EST in simulations (e.g., Mirarab et al., 2014a). Optimal trees based on two of three coalescence methods (STAR and ASTRAL) fail to support treeshrew + primates and instead position treeshrew with Glires to the exclusion of Primates (Fig. 6).

This pattern of inflated bootstrap support is repeated for another controversial node that Song et al. (2012) claimed to resolve with convincing support. Specifically, Song et al. (2012) argued that shortcut coalescence methods provide robust support for Perissodactyla + Carnivora. Indeed, their reported bootstrap support scores for Perissodactyla + Carnivora are impressive, 96% and 98% for MP-EST and STAR, respectively. However, our re-analyses of the very same bootstrap trees that Song et al. (2012) utilized show drops below the 95% level for this clade (77% for MP-EST; 92% for STAR) and still lower support for ASTRAL (only 74%). So, two of the problematic nodes that Song et al. (2012) claimed to resolve with convincing support were not recovered with robust bootstrap support (>95%) in our re-analyses with Shaw et al.'s (2013) implementation of STAR, MP-EST 1.4 with more thorough searches than were performed by Song et al. (2012), and Mirarab et al.'s (2014a) ASTRAL program that outperforms MP-EST.

3.2.2. Monkeyroos

Song et al.'s (2012) primary dataset included 447 genes. We detected a variety of investigator errors in this matrix by visual inspection of alignments and aberrant gene trees. Sequences for an Old World monkey (*Macaca*) and a wallaby (*Macropus*) were switched for 21 loci as previously reported by Mirarab et al. (2014a). In each case, the resulting gene trees position an Old World monkey (*Macaca*) with an opossum (*Monodelphis*) and a wallaby within Primates, thus supporting two clades of primates + marsupials that are separately derived in the afflicted gene trees (Fig. 7A). We presume that these errors that severely distorted ∼5% of the gene trees in the dataset were related to the first three letters of the swapped genus names that are identical ("Mac").

3.2.3. Data duplications

Inspection of likelihood scores for gene trees that were generated from Song et al.'s (2012) original 447-gene dataset suggested that eight loci were mistakenly duplicated based on identical scores. Song et al.'s (2012) Supplementary Table S2 confirmed that these eight loci (*BRPF1*, *DNNC1I2*, *SBF2*, *FLT3*, *ARHGAP17*, *NEDD9*, *COL12A1*, *LYST*) were accidentally replicated in their dataset (Fig. 7B). When these inadvertent duplications are deleted, there are only 439 unique loci.

3.2.4. Misalignments and non-homologous sequences

We inspected Song et al.'s (2012) 447 alignments for gross alignment problems and identified 26 genes for which non-homologous exons (or introns) from predicted, alternative splice variants were aligned with each other (Table 2). For example, cow (*Bos*), macaque (*Macaca*), mouse (*Mus*), and platypus (*Ornithorhynchus*) have transcript variant X2 of gene 209 (*GLS*) whereas all of the remaining taxa have transcript variant 1. Song et al.'s (2012) tree for gene 209 puts *Bos* and *Macaca* in a clade with platypus (Fig. 8; Table 2). This implies multiple deep coalescences >100 MY and breaks up the monophyly of Primates and Cetartiodactyla, two well-established mammalian orders. Similarly, cow (*Bos*), guinea pig (*Cavia*), orangutan (*Pongo*), and mouse (*Mus*) have transcript variant 1 of gene 163 (*SUPT3H*), whereas other sequences with complete 5′ ends are transcript variant 2 of this gene. Song et al.'s (2012) tree for gene 163 has *Cavia* nested inside of hominoid primates as the sister taxon to *Pongo*. This error results in non-monophyletic Primates and Rodentia (Fig. 8).

3.2.5. Peculiar phylogenetic trees arising from missing data

We identified 54 gene trees that exhibit unconventional and even bizarre relationships that may be attributed to extensive missing data in one or more sequences (Fig. 9; Table 3). The amount of missing data for platypus (*Ornithorhynchus*) exceeds 50% for 15 genes where platypus is grossly misplaced (Table 3). For example, platypus is sister to armadillo (*Dasypus*) in gene tree 29 (Fig. 9), but 95.1% of the *Ornithorhynchus* sequence is missing. Similarly, platypus is 91.1% missing for gene 213 and is sister to pig (*Sus*) on the corresponding gene tree (Fig. 9). Problems with missing data are not limited to platypus. Pig and dog (*Canis*) have high percentages of missing data for gene 53 (pig = 93.1%, dog = 73.5%) and these taxa are sister groups on the corresponding gene tree, resulting in non-monophyly of both Carnivora and Cetartiodactyla (Table 3).

3.2.6. NNI branch swapping and a poorly fitting model of sequence evolution

Song et al. (2012) reported that individual gene trees were estimated with RAxML and a GTR + Γ model of sequence evolution. We were unable to reproduce their gene trees with RAxML and GTR + Γ. Instead, we replicated Song et al.'s (2012) gene trees
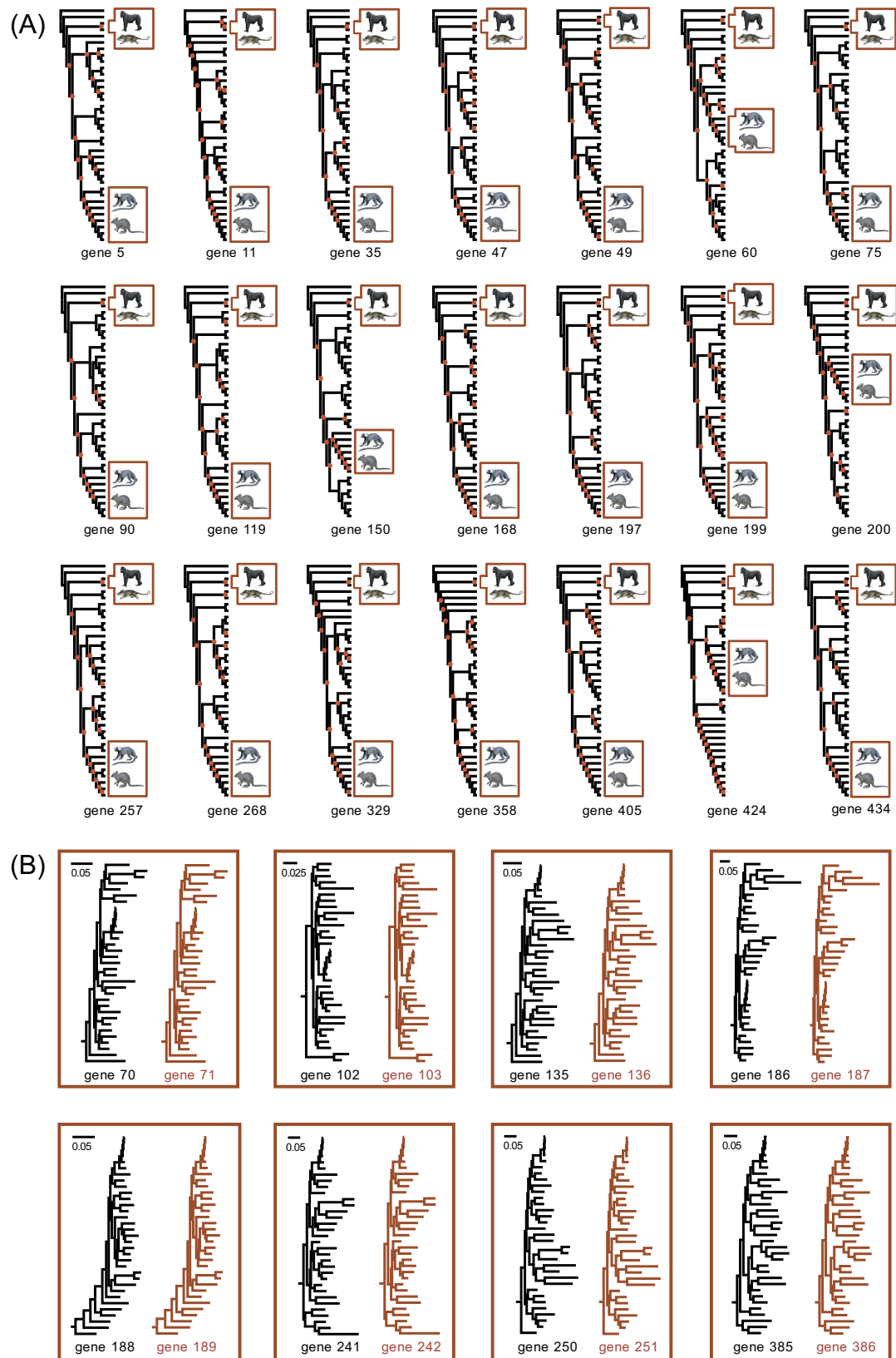
Fig. 7. Editing errors in Song et al.'s (2012) dataset of 447 genes. A. Twenty-one gene trees with switched taxonomic labels for a primate (*Macaca* = macaque) and a marsupial (*Macropus* = wallaby). Each gene tree implies polyphyly of Primates, polyphyly of Marsupialia, and deep coalescence in excess of 100 million years, which exceeds the age of crown Placentalia (dos Reis et al., 2012). Nodes that disagree with the species tree are indicated with filled red circles. B. Gene trees derived from eight genes that were inadvertently duplicated by Song et al. (2012). Paintings of mammals are by C. Buell. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Genes from Song et al. (2012) that include aligned, non-homologous regions.

Gene #	Name	Description of problem
26	RAPGEF5	5′ end of gene (see alignment positions 244–815) in *Echinops*, *Macaca*, and *Oryctolagus* is not homologous with sequences from other placental mammals and maps to a different region of *Homo sapiens* BAC clone CTA-252P22 (AC005079). Tree 26 clusters these three taxa together with *Pongo*, which is coded as "?" for this region of RAPGEF5
53	C12ORF63-like	5′ end of gene in *Canis*, *Mus*, and *Rattus* (see alignment positions 16–1060) is not homologous with sequences from other placental sequences and maps to a different region of *Homo sapiens* 12 BAC RP11-515D8 (AC008118). *Canis* clusters with *Sus* on Tree 53. *Sus* is represented by only 246 bp of sequence (length in *Homo* = 3591 bp) and is coded as "?" in the non-homologous region
62	MERTK	*Canis* and *Cavia* share 5′ ends (see alignment positions 1–75) that appear to be homologous with each other but are not homologous with other mammalian sequences. The first exon of *Homo* MERTK is on BAC clone RP11-592G13 (exon 1 is 61 bp) and is homologous to the predicted *Canis* MERTK mRNA (XM_005630437) on cont3.10904 (AAEX03010904), but not to positions 1–75 of *Canis* in Song et al.'s (2012) MERTK alignment
144	TRPM3	5′ end of gene (see alignment positions 1–215) in *Mus*, *Ochotona*, and *Pongo* is not homologous with corresponding region in other taxa. The 5′ sequence in *Pongo* is homologous to transcript variant X2 of TRPM3 in *Homo* (XM_005252218), whereas the sequence in other taxa is homologous to transcript variant 9 of *Homo* (NM_001007471). On Tree 144, *Ochotona* does not group with *Oryctolagus*. Also, *Pongo* groups with *Gorilla*, which is missing the first 215 bp
163	SUPT3H	5′ end of gene (see aligned positions 1–131) in *Bos*, *Cavia*, *Pongo*, and *Mus* is not homologous to other amniotes and maps to *Homo sapiens* scaffold107_800a (ADDF02111680), whereas sequences from other taxa map to *Homo sapiens* scaffold107_801 and scaffold107_802. *Cavia* is nested inside of Hominoidea on Tree 163
209	GLS	3′ end of gene (see aligned positions 1739–2104) in *Bos*, *Macaca*, *Mus*, and *Ornithorhynchus* is not homologous to other amniotes and maps to a different region of *Homo sapiens* CTG_1103279030326 (ABBA01062392). Tree 209 puts *Bos* and *Macaca* in clade with *Ornithorhynchus*
216	SOX5	*Sus* has block (see alignment positions 1781–2028) that is not homologous with other taxa. *Sus* is sister to *Canis* + *Gorilla* + *Sorex* on Tree 216
232	PITNC1	*Equus*, *Pan*, *Pongo*, *Homo*, and *Mus* share 3′ end of gene (see alignment positions 689–980) with *Gallus* that is not homologous with the corresponding region for other taxa in alignment. Sequences in former taxa map to different region of *Homo sapiens* CTG_1103279007186 (ABBA01028984) than other taxa. *Equus*, *Pongo*, *Pan*, and *Homo* are among the first taxa to branch off from other mammals on Tree 232, e.g., *Equus* is sister to all other mammals
235	LRCH1	*Mus*, *Rattus*, *Cavia*, *Macaca*, *Callithrix*, *Pongo*, *Pan*, *Bos*, and *Sus* have a 3′ end (see alignment positions 2368–2552) that is not homologous to the corresponding region in other mammals. The 3′ end in the former taxa is associated with transcript variant 3 whereas the 3′ end in other mammals is associated with transcript variants 1 and 2. The different 3′ ends map to different regions of *Homo sapiens* RefSeqGene ((NG_021335) on chromosome 13. *Homo* is excluded from a clade that contains other anthropoids on Tree 235. Also, *Bos* and *Sus* cluster to the exclusion of *Tursiops*
256	GRAMD3	*Pan*, *Pongo*, and *Canis* have 5′ ends that are not homologous with the corresponding region in other taxa that have a different splice site variant (see alignment positions 1–134). The different 5′ ends map to different regions of *Homo sapiens* CTG_1103276779381 (ABBA01068906). *Pongo* clusters with *Macaca* on Tree 256 instead of with other hominids
257	ADGRP	*Pan* and *Homo* share a 3′ end that is not homologous with *Pongo*, *Gorilla*, and *Macaca* (mislabeled as *Macropus*, see alignment positions 861–937). The different 3′ ends map to different regions of *Homo sapiens* clone RP3-413H6 (AL022724). *Pongo*, *Gorilla*, and *Macaca* cluster together on Tree 257
259	CDK19	Internal block in *Macaca* (see alignment positions 205–472) is not homologous with other sequences and maps to a different region of *Homo sapiens* utg7180000013841_quiver (JSAF02024017). *Macaca* and *Gorilla* are sister taxa on Tree 259. The latter taxon is scored as "?" for this middle region
262	LAMB1	*Homo* and *Gorilla* share a 5′ end (see alignment positions 120–226) that is not homologous with other sequences and maps to a different region of *Homo sapiens* utg7180000000092_quiver (JSAF02005899). *Homo* and *Gorilla* cluster on Tree 262
272	GRHL3	*Gorilla* and *Callithrix* share a 3′ end that is not homologous to that of *Pan*, *Homo*, and *Pongo* (see alignment positions 2044–2245) and maps to a different region of *Homo sapiens* clone RP3-462O23 (AL031431). *Gorilla* is outside of a clade that contains *Pan*, *Homo*, and *Pongo* on Tree 272
284	MCM4	*Bos* has an internal region (alignment positions 2180–2365) that is aligned with other sequences but is not homologous with other sequences. Comparison to the sheep *Ovis aries* (CBYI010028997) shows that the aligned region in *Bos* is upstream of the aligned region in other taxa. *Sus* also has internal region (alignment positions 412–730) that is not aligned with other sequences. *Bos* and *Sus* both have very long branches on Tree 284 and are in clade with *Tursiops* that is misplaced outside of both Laurasiatheria and Boreoeutheria
298	MCM8	*Pan* and *Pongo* have internal sequences that are not homologous to the corresponding internal sequences in *Homo* and *Gorilla* (see alignment positions 1354–1538). Sequences from this internal region in *Pan* and *Pongo* map to a different region of *Homo sapiens* utg7180000000189_quiver (JSAF02017883) than sequences from *Homo* and *Gorilla*. Further, *Callithrix* has a different non-homologous sequence in this region that maps to yet another region of *Homo sapiens* utg7180000000189_quiver (JSAF02017883). *Homo* and *Gorilla* cluster together on Tree 298, as do *Pan* and *Pongo*
309	KCNMA1	*Homo* and *Pongo* share the same splice site variant (e.g., XM_006717821) whereas *Pan* is associated with different variants (e.g., NM_001161352). *Homo* and *Pongo* are sister taxa to the exclusion of *Pan* on Tree 309
315	NDUFS3	The 5′ end of *Macaca* (see alignment positions 1–250) is not homologous to other sequences and maps to a different region of *Homo sapiens* utg7180000019116_quiver (JSAF02014306). *Macaca* is sister to *Vicugna* on Tree 315
317	SLC26A7	The 3′ end of *Homo* and *Pongo* (alignment positions 1959–2043) is not homologous to the corresponding region in *Pan* and maps to a different region of *Homo sapiens* utg7180000018351_quiver (JSAF02006426). *Homo* and *Pongo* also cluster together to the exclusion of *Pan* on Tree 317
332	PDE4B	*Pan*, *Macaca*, *Sus*, and *Rattus* have a different 5′ end that is not homologous to the corresponding region in other mammals (see alignment positions 1–654) and that maps to a different region of *Homo sapiens* utg7180000019175_quiver (JSAF02018171). *Pan* and *Macaca* are sister taxa on Tree 332
340	ERC1	*Equus* and *Macaca* share a 3′ end (see alignment positions 3090–3416) that is not homologous with the 3′ end of other mammals and maps to a different region of *Homo sapiens* utg7180000001238_quiver (JSAF02003641)
345	ACPP	*Homo*, *Pan*, and *Macaca* have a 3′ end (see alignment positions 1201–1342) that is not homologous to the corresponding region in other mammals and maps to a different region of *Homo sapiens* utg7180000025576_quiver (JSAF02014119). These three taxa also cluster together on Tree 345
368	GABRB3	*Homo* is transcript variant 1 of GABRB3 (also *Mus*, see alignment positions 16–166). The 5′ end of this sequence is not homologous to other mammalian sequences and maps to a different region of *Homo sapiens* utg7180000001082_quiver (JSAF02029183). *Pan* and *Gorilla* cluster to the exclusion of *Homo* on Tree 368

Table 2 (continued)

Gene #	Name	Description of problem
394	ANKRD42	*Mus*, *Rattus*, *Cavia*, *Bos*, *Callithrix*, *Sus*, *Canis*, *Equus*, and *Ornithorhynchus* have a 3′ end that is not homologous to the 3′ end in other mammals (see alignment positions 1207–1394) and maps to a different region of *Homo sapiens* utg7180000018169_quiver (JSAF02013883). Cetrtiodactyls are diphyletic on Tree 394 (*Bos* + *Sus* form a clade that is separate from *Tursiops* + *Vicugna*); *Rattus*, *Mus*, and *Cavia* are paraphyletic at the base of Placentalia
428	UNC5D	*Pongo*, *Pan*, and *Macaca* have a 5′ end that is not homologous to the 5′ end in other mammals (see alignment positions 1–98) and maps to a different region of *Homo sapiens* utg7180000000923_quiver (JSAF02025428). *Pan*, *Pongo*, and *Macaca* cluster together on Tree 428
429	CASP9	*Homo*, *Gorilla*, *Vicugna*, *Tursiops*, *Procavia*, and *Loxodonta* have a 3′ end (preprotein variant of *Homo*) that is not homologous to the 3′ end (transcript variant alpha of *Homo*) in other mammals (e.g., *Bos*, *Sus*, and *Pan*; see alignment positions 1556–1704) that maps to a different region of *Homo sapiens* utg7180000025419_quiver (JSAF02036961). *Homo* and *Gorilla* are sister taxa with a long stem branch on Tree 429; *Vicugna* and *Tursiops* are also sister taxa

and branch lengths using Phyml 2.4.4 with NNI branch swapping and a GTR model of sequence evolution without an allowance for among-site rate variation (Supplementary Text S1). NNI branch swapping is inferior to both SPR and tree bisection and reconnection (TBR) branch swapping and is prone to getting stuck in suboptimal islands of trees. Song et al. (2012, p. 14946) noted that, "It was suggested by the Akaike Information Criterion that GTR + Γ is the best-fit model for 364 genes…" and "GTR + Γ was selected as the substitution model used in the concatenation and coalescent methods for reconstructing the phylogeny of eutherian mammals." Instead, Song et al. (2012) implemented the simpler, more poorly fitting GTR model for all 447 gene-tree searches.

Given these issues it is not surprising that the majority of Song et al.'s (2012) individual gene trees are highly incongruent with accepted views of mammalian phylogeny (Table S1). Numerous trees exhibit basal splits in Placentalia between a single outlier species and all other placentals and may be accounted for by various combinations of swapped terminals (Fig. 7A), poor tree searches, misaligned nucleotides (Fig. 8), long-branch misplacement, and/or missing data (Fig. 9). Single species that are the sister to all other placental mammals on one or more trees include cow (*Bos*) (loci 133, 440), guinea pig (*Cavia*) (loci 16, 58, 173, 178, 381, 444), sloth (*Choloepus*) (loci 104, 123, 331), armadillo (*Dasypus*) (21, 291), kangaroo rat (*Dipodomys*) (loci 322, 361, 393, 443), tenrec (*Echinops*) (17 different loci), horse (*Equus*) (locus 97), hedgehog (*Erinaceus*) (26 different loci), mouse lemur (*Microcebus*) (loci 193, 270), mouse (*Mus*) (locus 275), pika (*Ochotona*) (locus 298), galago (*Otolemur*) (loci 92, 131, 332), flying fox (*Pteropus*) (locus 274), shrew (*Sorex*) (26 different loci), squirrel (*Spermophilus*) (loci 10, 125, 148, 297), tarsier (*Tarsius*) (loci 20, 41, 407), and treeshrew (*Tupaia*) (loci 84, 215, 252, 282, 293, 330, 356, 411) (Table S1).

In every case, these taxa should be nested much more deeply within Placentalia rather than sister to all remaining placentals. For example, cow belongs to the order Cetartiodactyla along with dolphin (*Tursiops*), pig (*Sus*), and alpaca (*Vicugna*) and is not the sister to a larger group that includes these three cetartiodactyl taxa plus all other placental mammals. In addition, Muridae (rat + mouse) is the sister to all other placentals on 33 gene trees. A basal or near-basal separation of murids from other placental mammals is the basis for the "guinea pig is not a rodent" hypothesis (D'Erchia et al., 1996) that was subsequently shown to result from long-branch misplacement (Bergsten, 2005). There are also gene trees that exhibit wildly divergent placements for the egg-laying monotreme, *Ornithorhynchus* (platypus), including sister-group relationships to *Tarsius* (locus 1 of 447), *Dipodomys* (loci 24, 216, 243, 260, 340), *Dasypus* (locus 29), *Vicugna* (locus 37), Carnivora (locus 48), *Choloepus* (locus 51), *Procavia* (locus 68), *Tupaia* (loci 74, 432), *Felis* (loci 80, 371), *Echinops* (locus 86), *Sus* (loci 121, 213), *Sorex* (loci 132, 217), *Myotis* (loci 147, 269), *Otolemur* (locus 153), *Loxodonta* (loci 163, 301), *Monodelphis* (locus 166), *Oryctolagus* (locus 199), *Bos* + *Macaca* (locus 209), *Cavia* (locus 319), and Eulipotyphla (locus 341) (Fig. 9; Table S1).

3.2.7. Reanalysis of Song et al.'s (2012) data following correction of investigator errors

We reanalyzed Song et al.'s (2012) 447 genes with PhyML 3.0, subtree pruning and regrafting (SPR) + NNI branching swapping, and a GTR + Γ model of sequence evolution. The resulting likelihood scores are ∼2022 log likelihood units better, on average, than likelihood scores for Song et al.'s (2012) gene trees based on NNI branch swapping and a GTR model of sequence evolution with PhyML 2.4.4 (Fig. 10; Table S2). In the case of gene #149, the log likelihood increased by >20,000 units with SPR + NNI branch swapping and a GTR + Γ model of sequence evolution. After correcting taxon names for the 21 loci where Song et al. (2012) inadvertently switched names for *Macropus* and *Macaca* (Fig. 7A), and excluding eight duplicated loci (Fig. 7B) and 26 loci that are characterized by misalignment of different splice variants (Fig. 8; Table 2), we selected 413 of 447 genes trees for further analyses We note that this curated set of 413 gene trees does not solve the more fundamental problem of concatalescence that afflicts Song et al.'s (2012) entire dataset (Section 3.1, Fig. 2).

The differences between Song et al.'s (2012) 447 gene trees and our 413 gene trees are striking and show that Song et al.'s (2012) gene trees (Fig. 11A) exhibit much higher levels of incongruence with the inferred species tree relative to our better gene trees (Fig. 11B). Song et al.'s (2012) gene trees show at least 5% more conflict with the species tree than our 413 gene trees for 21 different clades, and in five cases (Boreoeutheria, Laurasiatheria, Euarchontoglires, Glires, Rodentia) there is at least 30% more conflict (Fig. 11). Within Primates, 412 of 413 gene trees support Anthropoidea based on SPR + NNI searches with a GTR + Γ model of sequence evolution, whereas Song et al.'s (2012) analyses yielded 39 of 447 gene trees (8.7%) that conflict with Anthropoidea, a clade with an ∼27.6 MY stem branch (dos Reis et al., 2012). Deeper in Primates, 154 of 413 gene trees (37.2%) conflict with Haplorhini (Anthropoidea + *Tarsius*) whereas 202 of Song et al.'s (2012) 447 gene trees (45.2%) conflict with this clade. Although SPR + NNI yields some improvement relative to NNI, both analyses performed poorly in recovering Haplorhini. This clade is uniformly supported by 104 insertions of transposons with no transposon insertions favoring alternative relationships (Hartig et al., 2013) and is not a candidate for deep coalescence (Gatesy and Springer, 2014). Rather, difficulties associated with recovering Haplorhini likely reflect a long branch (*Tarsius*) and the antiquity of the Tarsiiformes-Anthropoidea split, which occurred in the Paleocene or even as far back as the latest Cretaceous (Meredith et al., 2011; Perelman et al., 2011; dos Reis et al., 2012, 2014; Springer et al., 2012).

Other ancient clades with unopposed transposon support include the superordinal groups Afrotheria (6 transposons), Laurasiatheria (9 transposons), Boreoeutheria (10 transposons), and Euarchontoglires (9 transposons) (Nishihara et al., 2009). Unanimous and unopposed support from transposon insertions suggests that accurately reconstructed gene trees should recover
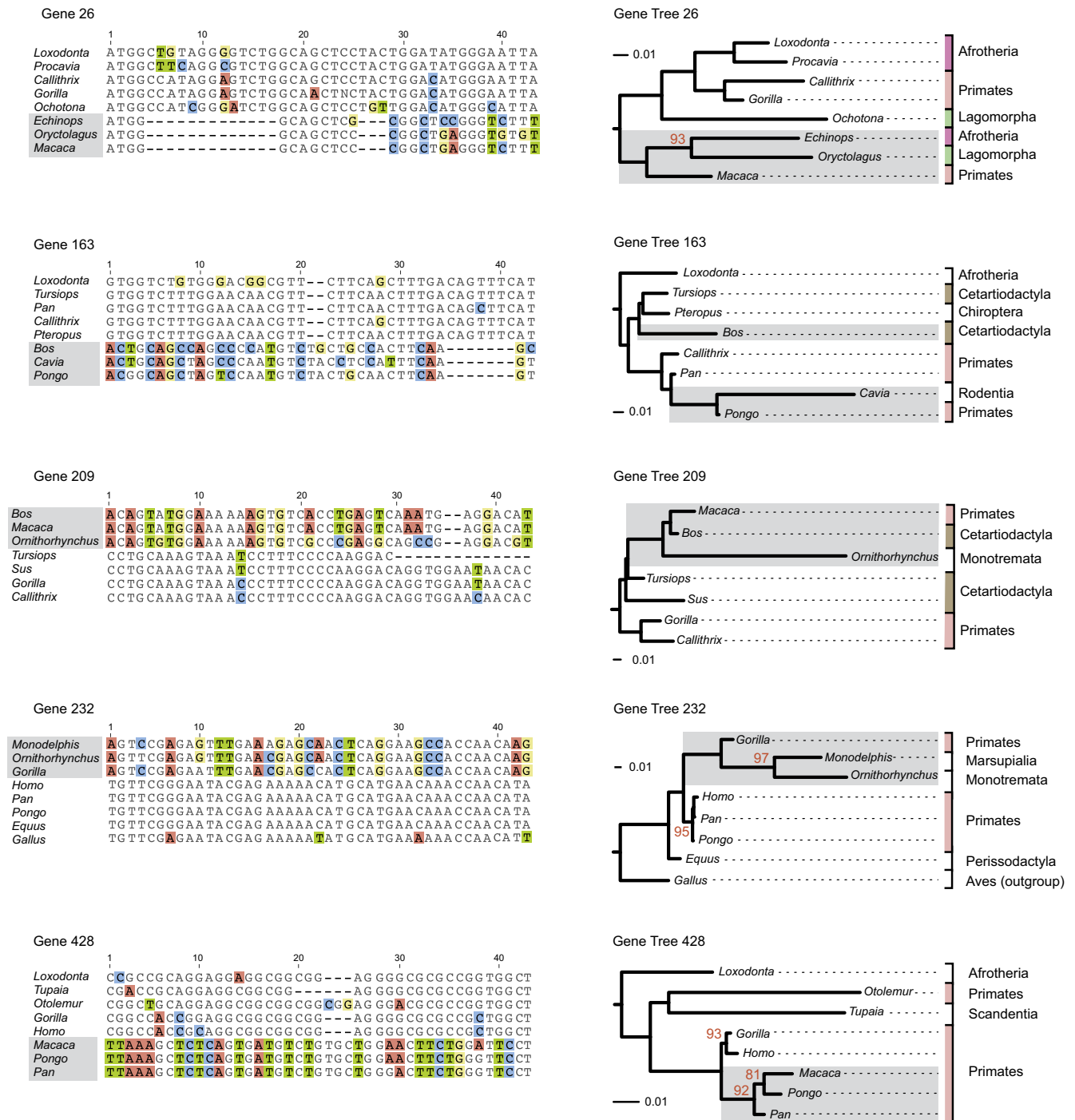
Fig. 8. Misaligned, non-homologous sequences (shaded versus non-shaded taxon names) for genes 26, 163, 209, 232, and 428 from Song et al. (2012). In each case, the problems caused by the non-homologous sequences are unrelated to incomplete lineage sorting and instead represent alignment errors between different splice site variants. In each gene tree, distantly related taxa with different splice site variants cluster together on gene trees (see Table 1 for additional examples); mammalian orders that are not monophyletic in the distorted gene trees are marked by colored bars to the right. For gene tree 428, misalignment of different splice variants separates *Pan* (chimp) from *Homo* (human) and *Gorilla* (gorilla) and disrupts monophyly of Hominidae (*Pan*, *Homo*, *Gorilla*, and *Pongo*), Homininae (*Pan*, *Homo*, *Gorilla*), and Hominini (*Pan*, *Homo*). Spurious clades with bootstrap percentages >50% are shown at internal branches. Note that trees shown are derived from full gene trees for all 36 mammalian species, but many taxa are pruned away to highlight conflicting relationships that are due to misalignment of non-homologous exons from different splice variants. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

these clades at a high rate given the absence of transposon evidence for ILS at these nodes. By contrast with transposons, Song et al.'s (2012) gene trees exhibit strong conflicts with each of these clades (Afrotheria = 100 conflicting trees [22.4%]; Laurasiatheria = 292 conflicting trees [65.3%], Boreoeutheria = 330 conflicting trees [73.8%]; Euarchontoglires = 303 conflicting trees [67.8%])

(Fig. 11A). The majority of conflicts are eliminated with SPR + NNI searches, a GTR + Γ model of sequence evolution, corrected taxon names for 21 loci, and the exclusion of redundant loci and genes with mis-aligned, non-homologous sequences (Afrotheria = 28 conflicting trees [6.8%]; Laurasiatheria = 89 conflicting trees [21.5%], Boreoeutheria = 89 conflicting trees [21.5%], Euarchontoglires = 89
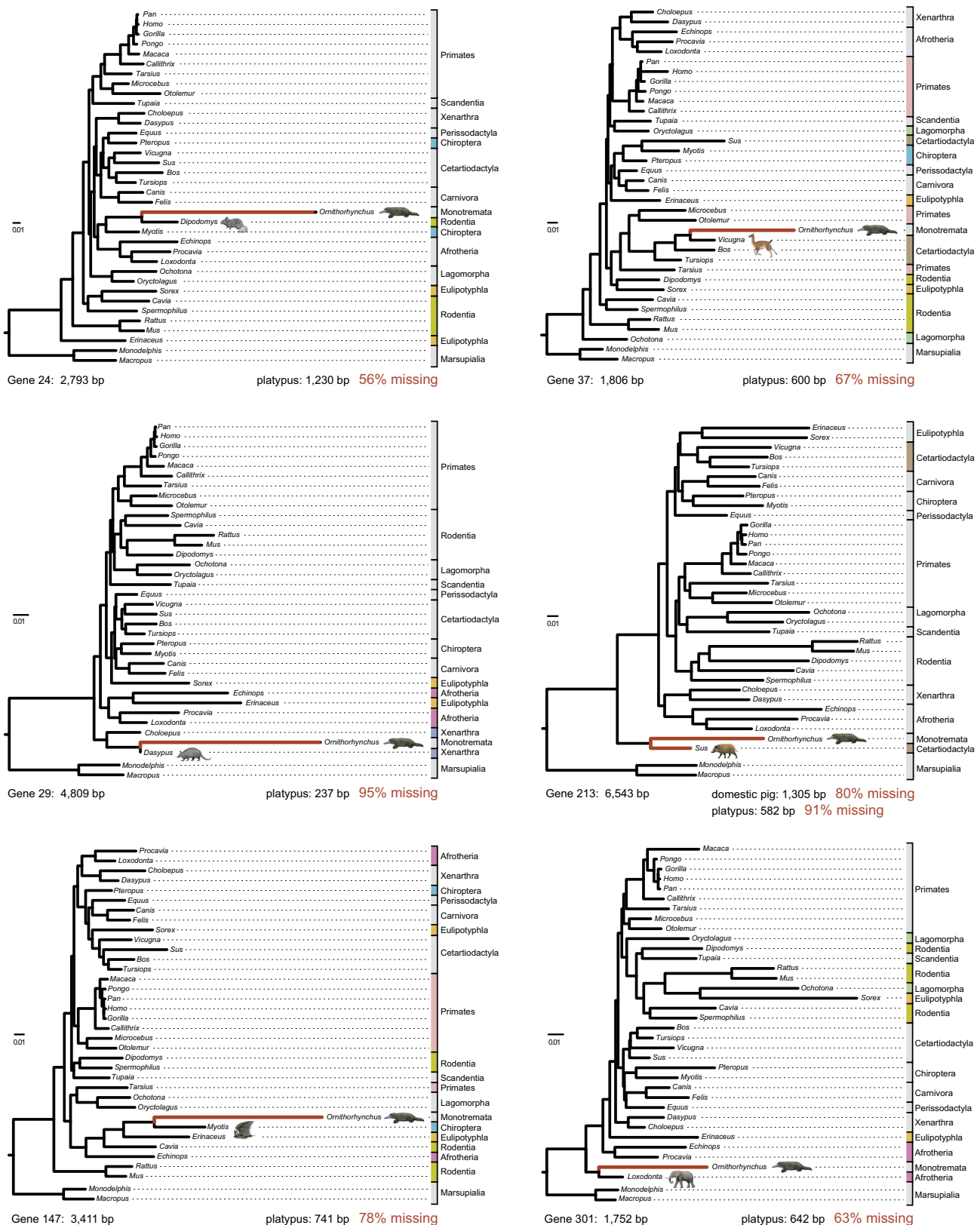
Fig. 9. Missing data problems for representative genes and gene trees from Song et al. (2012). Taxa with high percentages of missing data (red branches) are routinely misplaced on individual gene trees (see Table 3 for additional examples). Mammalian orders that are not monophyletic in the distorted gene trees are marked to the right by colored bars. In all trees shown, the egg-laying mammal *Ornithorhynchus* (platypus) clusters within placental mammals. Paintings of mammals are by C. Buell. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Taxa with high percentages of missing data and attendant phylogenetic problems among Song et al.'s (2012) genes and gene trees.

Gene #	Taxon w/Short sequence	Full length sequence (*Homo* unless otherwise noted)	Length of short sequence	% Missing in Taxon w/Short sequence	Position of Taxon w/Short sequence
24	*Ornithorhynchus*	2793	1230	56.0	Sister to *Dipodomys*
29	*Ornithorhynchus*	4809	237	95.1	Sister to *Dasypus*
37	*Ornithorhynchus*	1806[a]	600	66.8	Sister to *Vicugna*
53	*Sus*	3591	246	93.1	Sister to *Canis* (*Canis* 73.5% missing)
60	*Pongo*	2061	753	63.5	Sister to other Placentalia excepting Lagomorpha
91	*Equus*	843	411	51.2	Sister to *Vicugna*
104	*Pan*	1023	363	64.5	Sister to *Sorex*
121	*Sus*	9171	951	89.6	Sister to *Ornithorhynchus* (*Ornithorhynchus* 66.3% missing)
132	*Callithrix*	7326	1455	80.1	Sister to *Erinaceus*
132	*Sorex*	7326	3369	54.0	Sister to *Ornithorhynchus* (*Ornithorhynchus* 51.6% missing)
133	*Bos*	8190	972	88.1	Sister to other Placentalia
145	*Sus*	5997	843	85.9	Sister to *Microcebus*
147	*Ornithorhynchus*	3411	741	78.3	Sister to *Myotis*
148	*Choloepus*	2577	1084	57.9	Sister to *Felis*
154	*Ornithorhynchus*	2547	465	81.7	Sister to Theria (rooted between *Erinaceus* and other therians)
199	*Ornithorhynchus*	2199	534	75.7	Sister to Oryctolagus
213	*Ornithorhynchus*	6543	582	91.1	Sister to *Sus* (*Sus* 80.1% missing)
216	*Sus*	2292	1134	50.5	Sister to *Canis* + *Gorilla* + *Sorex* (*Canis* 50.5% missing)
216	*Ornithorhynchus*	2292	1140	50.3	Sister to *Dipodomys*
217	*Ornithorhynchus*	2304	963	58.2	Sister to *Sorex*
218	*Sorex*	2586	1241	52.0	Sister to Xenarthra
246	*Sus*	4155	378	90.9	Sister to *Sorex* (*Sorex* 49.6% missing)
250	*Macaca*	2646	633	76.1	Sister to *Sorex*
255	*Sus*	3972	2205	44.5	Sister to *Sorex*
259	*Sus*	1476	681	53.9	Sister to *Equus*
260	*Ornithorhynchus*	3813	1830	52.0	Sister to *Dipodomys*
263	*Sorex*	780	398	49.0	Sister to other mammals
266	*Sus*	2280	1536	32.6	Sister to other Placentalia
269	*Macaca*	837	237	71.7	Sister to *Pongo*
275	*Rattus*	1065	180	83.1	Sister to *Loxodonta*
276	*Gorilla*	1878	799	57.5	Outside of Primates
280	*Tarsius*	2619	1400	46.5	Sister to *Tupaia*
287	*Felis*	1917	789	58.8	Sister to *Dasypus*
298	*Ochotona*	2643	1355	48.7	Sister to other Placentalia
301	*Ornithorhynchus*	1752	642	63.4	Sister to *Loxodonta*
310	*Sus*	2925	1221	58.3	Sister to *Sorex*
314	*Sorex*	1206	566	53.1	Sister to *Sus*
315	*Myotis*	792	252	68.2	Sister to *Choloepus*
318	*Myotis*	849	318	62.5	Sister to *Procavia*
319	*Ornithorhynchus*	3573	603	83.1	Sister to *Cavia*
323	*Sus*	2094	429	79.5	Sister to other Boreoeutheria
326	*Ornithorhynchus*	3270	426	87.0	Sister to *Microcebus*
341	*Sorex*	2250	683	69.6	Eulipotyphla sister to *Ornithorhynchus* (*Erinaceus* 36.7% missing)
352	*Tupaia*	2229	1001	55.1	Sister to *Tarsius* (*Tarsius* 33.0% missing)
354	*Sorex*	2244	1183	47.3	Sister to *Ornithorhynchus*
355	*Sorex*	2079[a]	895	57.0	Sister to *Dasypus*
360	*Felis*	612	266	56.5	Sister to *Dipodomys*
393	*Dipodomys*	3348	1701	49.2	Sister to other Placentalia
395	*Tarsius*	1281	404	68.5	Sister to *Cavia*
402	*Bos*	1074	246	77.1	Sister to *Sorex* (*Sorex* 45.8% missing)
404	*Sus*	3948	531	86.6	Sister to *Otolemur*
412	*Sus*	1836	417	77.3	Sister to *Erinaceus*
425	*Sus*	3093	825	73.3	Sister to other Placentalia excepting Eulipotyphla
432	*Ornithorhynchus*	1897	1017	46.4	Sister to *Tupaia*
438	*Macaca*	2724	636	76.7	Sister to *Tarsius*
440	*Bos*	2301	1179	48.8	Sister to other Placentalia

[a] Reference sequence (complete) is *Pan* instead of *Homo*.

conflicting gene trees [21.5%]) (Fig. 11B). More thorough analysis of the data using a better branch-swapping algorithm (NNI + SPR) and the model of nucleotide substitution that Song et al. (2012) recommended, but did not use, eliminated >500 unnecessary gene tree conflicts at three nodes (Boreoeutheria, Laurasiatheria, Euarchontoglires).

In cases where Afrotheria, Laurasiatheria, Euarchontoglires, and Boreoeutheria are still not recovered, however, it is unlikely that ILS is the culprit given unopposed transposon support for these clades (Fig. 11). These ancient superordinal clades have historically proved difficult to resolve with short sequences in combination with poor taxon sampling and attendant problems with long

branches. Song et al.'s (2012) data set for 36 mammals and one outgroup is replete with long, undivided branches that are prone to long-branch misplacement including *Gallus*, *Ornithorhynchus*, *Echinops*, *Cavia*, *Tupaia*, *Tarsius*, *Sorex*, and *Erinaceus*.

MP-EST, STAR, and ASTRAL were used to infer species trees based on the 413 gene trees that were reconstructed with Phyml 3.0, SPR + NNI branch swapping, and a GTR + Γ model of sequence evolution (Fig. 12). The mean Robinson–Foulds distance between individual gene trees and the inferred species tree decreased from 25.7 (STAR and ASTRAL species trees) or 25.6 (MP-EST species tree) for Song et al.'s (2012) 447 NNI gene trees to 18.3 (STAR, ASTRAL, MP-EST species trees) for 413 SPR + NNI gene trees. The resulting
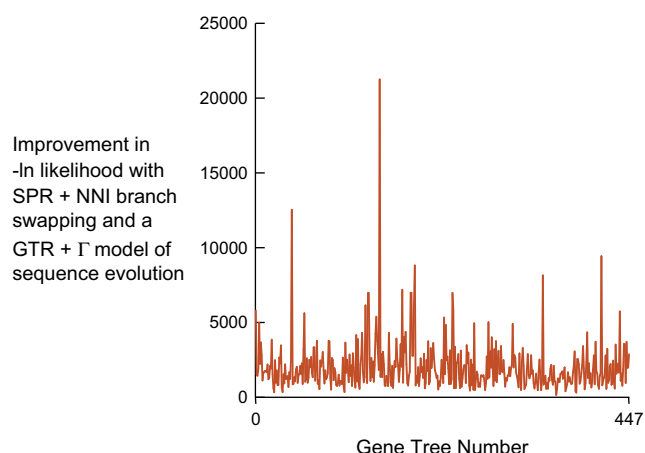
Fig. 10. The difference in −ln likelihood scores for 447 gene trees obtained with PhyML 3.0 with NNI + SPR branch swapping and a GTR + Γ model of sequence evolution versus PhyML 2.4.4 with NNI branch swapping and a GTR model of sequence evolution as employed by Song et al. (2012). The mean difference in likelihood scores for 447 gene trees is ~2022 ln likelihood units.

Song et al. (2012) claimed that MP-EST and STAR are robust to differential taxon and gene sampling and provide robust support for treeshrew + Primates, but our reanalysis of their own data yields a species tree that agrees with Meredith et al. (2011) in supporting Scandentia as the sister group to Glires (Fig. 12). Mirarab et al. (2014a,b,c) also recovered a sister group relationship between Glires and Scandentia based on a reanalysis of 424 loci from Song et al. (2012) with RAxML gene trees and ASTRAL. Mirarab et al.'s (2014b) 424 RAxML trees yielded a similar decrease in the mean RF distance (18.1) for STAR, ASTRAL, and MP-EST species trees relative to Song et al.'s (2012) calculations (mean RF distance = 25.6–25.7). These large decreases in RF distances among gene trees relative to Song et al.'s (2012) calculations demonstrate conclusively that Song et al.'s (2012) gene trees are plagued by heterogeneity that is completely unrelated to ILS. These drops in RF distance imply an average of ~3.5 incorrectly reconstructed nodes per gene tree (i.e., RF/2) that can be directly attributed to investigator error and have nothing to do with deep coalescence.

Bootstrap analyses based on MP-EST, STAR, and ASTRAL with 413 loci yielded inconsistent support for the position of treeshrew and unimpressive support scores for Perissodactyla + Carnivora (*Equus*, *Felis*, *Canis*) within Laurasiatheria. ASTRAL and STAR recovered 81% and 65% bootstrap support for treeshrew + Glires, respectively, whereas MP-EST recovered 65% support for treeshrew + Primates (65%) and much lower (35%) support for treeshrew + Glires (Fig. 12). The striking difference between ASTRAL and MP-EST may be explained by MP-EST's reliance on rooted gene trees whereas the input for ASTRAL is unrooted quartets (Simmons and Gatesy, 2015). ASTRAL bootstrap support for Perissodactyla + Carnivora (67%) based on 413 genes is also much lower than Song

species trees (Fig. 12) are similar to the MP-EST and STAR bootstrap trees that were reported by Song et al. (2012), but an important difference is the placement of treeshrew as sister to Glires rather than Primates. This relationship for treeshrew was confirmed with 413 RAxML gene trees, which resulted in MP-EST, STAR, and ASTRAL trees that are topologically identical to those that were obtained with 413 PhyML trees.

Fig. 11. Percentages of gene trees that conflict with the inferred MP-EST species tree for Song et al.'s (2012) 447 gene trees that employed NNI branch swapping and a GTR model of sequence evolution (A) and 413 better gene trees based on NNI + SPR branch swapping and a GTR + Γ model of sequence evolution (B). The percentage of gene trees that disagree with individual clades supported by the species tree is shown above internal branches. Note that Song et al.'s (2012) gene trees are at least 5% more incongruent at most nodes (colored circles and starbursts) but exhibit 6% more support for their inferred resolution of the placental root. In both trees, the root is positioned such that Afrotheria and Xenarthra form a clade that is sister to Boreoeutheria. Several clades with unopposed transposon support (Boreoeutheria, Laurasiatheria, Euarchontoglires, Haplorhini, Afrotheria) show extensive conflict for the gene trees of Song et al. (2012), which suggests wholesale errors in gene tree reconstruction (A). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
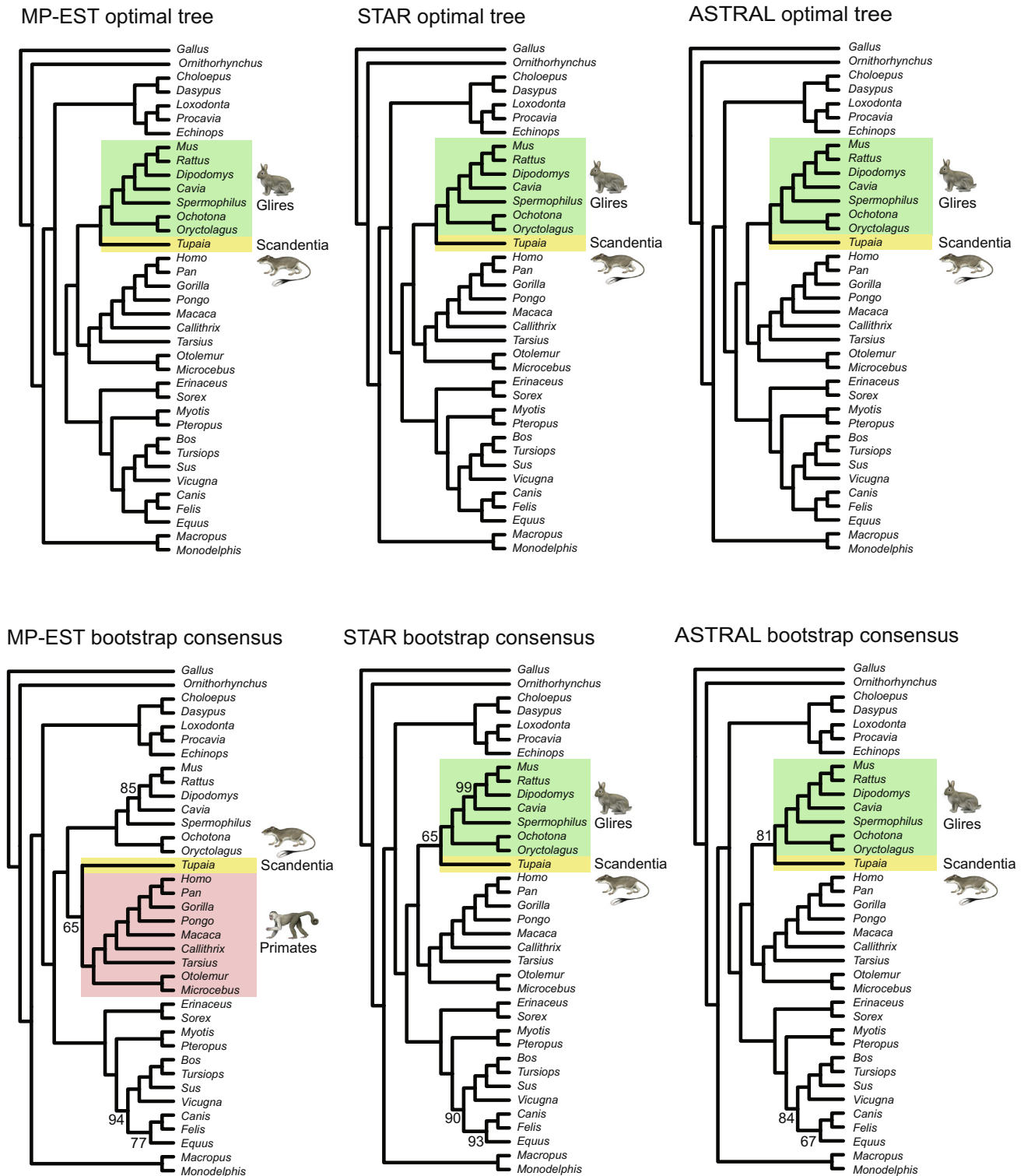
Fig. 12. Species trees (MP-EST, STAR, ASTRAL) based on 413 gene trees (top) and bootstrapped gene trees with 100 pseudoreplicates (bottom). Species trees shown in the top panel are based on gene trees that were inferred with PhyML 3.0, NNI + SPR branch swapping, and a GTR + Γ model of sequence evolution. Species trees based on 413 optimal gene trees that were inferred with RAxML and GTR + Γ were topologically identical to those that were obtained with PhyML gene trees (top). All three optimal trees position *Tupaia* (treeshrew) as the sister taxon to Glires and contradict the species tree advocated by Song et al. (2012). Bootstrap consensus trees (MP-EST, STAR, ASTRAL) are based on gene trees that were inferred with RAxML trees for 100 pseudoreplicates (bottom). Bootstrap support percentages are only shown for clades with less than 100% support. Paintings are by C. Buell.

et al.'s (2012) reported support values of 96% and 98% for this controversial clade based on MP-EST and STAR, respectively (Fig. 6).

Song et al. (2012) harshly critiqued concatenation for inconsistent placement of Scandentia in subsampling analyses where random subsets of genes from their phylogenomic matrix were analyzed (their Fig. 2). Song et al. (2012) argued that coalescence methods are superior because they provide consistent support for Scandentia + Primates when genes in their dataset are subsampled.

Ironically, Song et al.'s (2012) error-ridden gene trees fail to provide clear support for their preferred hypothesis (Fig. 6). Instead, the grouping of Scandentia with Primates may indicate a consistently wrong result for this dataset, since correcting the numerous errors in this matrix and analyzing the data using more appropriate substitution models supports Scandentia + Glires on all three optimal species trees as well as the ASTRAL and STAR bootstrap trees (Fig. 12). Given that Song et al. (2012) analyzed a dataset with wholesale errors in basic data compilation (i.e., gene duplications, swapped taxon names, misalignments of non-homologous sequences, long tracts of missing data, poor searches) and the wrong model of evolution, none of their conclusions regarding the inferiority of concatenation are justified. Given the numerous editing errors in Song et al.'s (2012) matrix, we would expect concatenation results from different subsamplings of loci to support different trees – perhaps depending on which subsets of genes contained monkey/wallaby swaps (Fig. 7A) or whether loci with gross alignment errors (Fig. 8) were included or not. This would be especially true if Song et al. (2012) again chose to analyze their subsampled datasets with an arbitrarily chosen model, instead of the model chosen by the AIC. We contend that Song et al.'s (2012) subsampling analyses need to be completely redone to determine whether any of their negative interpretations of concatenation relative to shortcut coalescence are valid.

3.2.8. Improbable deep coalescences

We calculated the maximum deep coalescence for each of Song et al.'s (2012) 447 gene trees based on dos Reis et al.'s (2012) timetree for mammals (Fig. 1), and recovered values that range from 3.1 to 174.2 MY across these trees (Table S1). Sixty-six gene trees imply deep coalescences that exceed 100 MY (Table 4). For example, cat and dog last shared a common ancestor 54.2 million years ago (MYA) whereas cat and platypus last shared a common ancestor 184.6 MYA based on dos Reis et al.'s (2012) time calibrated species tree (Fig. 1). Cat and platypus are sister taxa on Song et al.'s (2012) gene tree 80, which implies that cat and dog alleles at this locus coalesce 184.6 MYA or before. The difference between these values (i.e., 184.6 MY − 54.2 MY) results in a deep coalescence of 130.4 MY for the cat and dog alleles at locus 80 (Table 4). The deepest coalescence in each gene tree averages 42.0 MY, which is less than the mean of 101.5 MY for McCormack et al.'s (2012) data set of 183 UCE loci (Gatesy and Springer, 2014), but still of great concern for shortcut coalescence methods that assume all conflict between gene trees and the underlying species tree is attributable to ILS.

Moreover, the maximum deep coalescence for each gene tree is only part of the coalescence debt that each gene tree incurs, where coalescence debt is defined as the sum of deep coalescence across an entire gene tree that is required to fit that gene tree to the species tree. In some cases, the amount of coalescence debt that is implied by Song et al.'s (2012) gene trees is extraordinary. Dos Reis et al.'s (2012) timetree (Fig. 1) spans 2466.7 MY for the 36 mammalian species in Song et al.'s (2012) data set, yet the total amount of time that is required to account for mammalian branches on gene tree 269 is 4034.8 MY (Fig. 13). Thus, the total coalescence debt on this tree is 1568.1 MY and averages 82.5 MY per node for the 19 nodes in the gene tree that conflict with the species tree. We are unaware of any mammalian gene with deep coalescences of this magnitude. A more reasonable explanation is that garden-variety long-branch misplacement resulted in an association of two long branches, *Ornithorhynchus* (platypus) and *Myotis* (little brown bat), with the consequence that platypus is 16 internal nodes crownward of its accepted position as the sister taxon to therian mammals (placentals + marsupials) (Fig. 13). More generally, pervasive deep coalescences that are implied by Song et al.'s (2012) highly aberrant gene trees are unlikely to represent

ILS, and instead reflect traditional problems in phylogeny reconstruction such as poor searches that get trapped in local optima, long-branch misplacement, model misspecification, lack of phylogenetic signal, missing data, poor taxon sampling, alignment problems (Patel et al., 2013; Gatesy and Springer, 2014), and investigator errors (Figs. 7–11).

3.2.9. Stunted species trees and circular simulations

It has previously been shown that MP-EST trees have stunted internal branches due to inaccurate gene trees that imply unrealistic deep coalescences (Gatesy and Springer, 2014; Mirarab et al., 2014b, c; Springer and Gatesy, 2014). Fig. 14 shows an MP-EST tree based on 447 gene trees from Song et al. (2012) that employed NNI branch swapping and a GTR model of sequence evolution (Fig. 14A), and an alternative MP-EST tree based on 413 better gene trees estimated using NNI + SPR branch swapping and a GTR + Γ model of sequence evolution (Fig. 14B). Internal branch lengths on these trees are in CUs; terminal branch lengths cannot be meaningfully estimated using MP-EST and are arbitrarily nine CUs. Internal branches that are at least 2×, 3×, or 6× longer on one of the two MP-EST trees are highlighted in different shades of green. Twelve branches on the 413-gene MP-EST tree (Fig. 14B) are at least 2× longer than equivalent branches on the 447-gene MP-EST tree (Fig. 14A). In the case of the stem Catarrhini branch, there is an 11× difference, which in large part reflects the corrected taxonomic names for the marsupial *Macropus* and the catarrhine primate *Macaca*. The only internal branches that are at least 2× longer on the 447-gene MP-EST tree are at controversial polytomies (see Fig. 1). This finding implies that support for Song et al.'s (2012) phylogenetic hypothesis is biased at the most critical nodes and that model misfit and inadequate tree searches can provide artifactual support in coalescence analyses.

Table 5 shows branch lengths (in CUs) for three MP-EST trees that represent variations on Song et al.'s (2012) data set: the MP-EST tree based on 447 PhyML (NNI, GTR) gene trees from Song et al. (2012) (Fig. 14A), the MP-EST tree based on 424 RAxML (GTR + Γ) gene trees from Mirarab et al. (2014b), and our MP-EST tree based on 413 PhyML (SPR + NNI, GTR + Γ) gene trees (Fig. 14B). Table 5 also shows the number of CUs per branch based on dos Reis et al.'s (2012) timetree for mammals (Fig. 1) with four different estimates from the literature for the length of one CU. Patel et al. (2013) suggested that 1 CU = 400,000 years for many vertebrates. Carstens and Dewey (2010) estimated 1 CU = 1.0–1.5 MY for mouse-eared bats (*Myotis*). Rannala and Yang's (2003) estimate for a historical population size of 20,000 in the common ancestor of *Homo* + *Pan* yields 1 CU = 800,000 years if generation time equals 20 years (Rannala and Yang, 2003). Finally, Hobolth et al. (2007) estimated effective population sizes of 45,000 and 65,000 individuals in the ancestors of Hominae and *Homo* + *Pan*, respectively. If we use the mean of these two values (55,000 individuals) and Hobolth et al.'s (2007) estimate of 25 years for generation time then one CU = 2.75 MY. The total number of CUs for the stem Theria branch and all of its descendant branches (excluding terminal branches) is equal to 35.75 for the 447-gene MP-EST tree. This corresponds to only 2.89% of the total coalescence time (1237.5 CUs) based on dos Reis et al.'s (2012) timetree and Patel et al.'s (2013) estimate for the duration of one CU. At the other extreme, the 447-gene MP-EST tree accounts for 19.8% of the total number of CUs (180) if we choose a CU of 2.75 MY based on Hobolth et al.'s (2007) estimates for population size and generation time in long-lived hominids.

MP-EST trees based on 424 and 413 genes with better tree searches (RAxML analyses from Mirarab et al. (2014b), PhyML with best of SPR + NNI) show an increase in the total number of coalescence units (53.49 CUs for 424 genes, 57.47 CUs for 413 genes), but these totals comprise only 4.3 to 31.9% of the total coalescence

Table 4
Trees from Song et al. (2012) with maximum deep coalescence >100 million years.

Tree number (of 447)	Clade with maximum deep coalescence	Species tree divergence (MY)	Gene tree divergence (MY)	Magnitude of deep coalescence (MY)
1	Muridae + *Dipodomys*	56	184.6	128.6
5	Catarrhini	26.4	172.8	146.4
11	Catarrhini	26.4	172.8	146.4
24	Muridae + *Dipodomys*	56	184.6	128.6
29	Xenarthra	69.4	184.6	115.2
35	Catarrhini	26.4	172.8	146.4
37	*Bos + Tursiops*	52.8	184.6	131.8
47	Catarrhini	26.4	172.8	146.4
48	Chiroptera	59.1	184.6	125.5
49	Catarrhini	26.4	184.6	158.2
51	Xenarthra	69.4	184.6	115.2
53	*Sus + Bos + Tursiops*	58.1	184.6	126.5
60	Catarrhini	26.4	184.6	158.2
68	Strepsirrhini	55.1	184.6	129.5
74	Euarchonta	74.1	184.6	110.5
75	Catarrhini	26.4	184.6	158.2
77	*Homo + Pan + Gorilla*	10.4	172.8	162.4
80	Carnivora	54.2	184.6	130.4
86	Afrotheria	70.4	184.6	114.2
90	Catarrhini	26.4	184.6	158.2
96	Rodentia	64.4	172.8	108.4
119	Catarrhini	26.4	184.6	158.2
121	*Sus + Bos + Tursiops*	58.1	184.6	126.5
132	Anthropoidea	37.4	184.6	147.2
147	Muridae + *Dipodomys*	56	184.6	128.6
150	Catarrhini	26.4	184.6	158.2
152	Strepsirrhini	55.1	172.8	117.7
153	Strepsirrhini	55.1	184.6	129.5
154	Muridae + *Dipodomys*	56	172.8	116.8
163	Paenungulata	60.3	184.6	124.3
166	Strepsirrhini	55.1	184.6	129.5
168	Catarrhini	26.4	184.6	158.2
177	Catarrhini	26.4	172.8	146.4
195	Muridae + *Dipodomys*	56	172.8	116.8
197	Catarrhini	26.4	184.6	158.2
199	Catarrhini	26.4	184.6	158.2
200	Catarrhini	26.4	184.6	158.2
203	*Sus + Bos + Tursiops*	58.1	184.6	126.5
209	Catarrhini	26.4	184.6	158.2
213	*Sus + Bos + Tursiops*	58.1	184.6	126.5
216	Strepsirrhini	55.1	184.6	129.5
217	Eulipotyphla	61.5	184.6	123.1
232	*Homo + Pan + Gorilla*	10.4	184.6	174.2
243	Muridae + *Dipodomys*	56	184.6	128.6
257	Catarrhini	26.4	172.8	146.4
260	Muridae + *Dipodomys*	56	184.6	128.6
263	Eulipotyphla	61.5	184.6	123.1
268	Catarrhini	26.4	172.8	146.4
269	Muridae + *Dipodomys*	56	184.6	128.6
292	Muridae + *Dipodomys*	56	172.8	116.8
301	Paenungulata	60.3	184.6	124.3
303	Eulipotyphla	61.5	184.6	123.1
319	Muridae + *Dipodomys*	56	184.6	128.6
326	Strepsirrhini	55.1	184.6	129.5
329	Catarrhini	26.4	172.8	146.4
340	Lagomorpha	47.9	184.6	136.7
341	Laurasiatheria	76	184.6	108.6
354	Eulipotyphla	61.5	184.6	123.1
358	Catarrhini	26.4	172.8	146.4
369	Muridae + *Dipodomys*	56	172.8	116.8
371	Carnivora	54.2	184.6	130.4
373	Lagomorpha	47.9	172.8	124.9
405	Catarrhini	26.4	172.8	146.4
424	Catarrhini	26.4	172.8	146.4
432	Muridae + *Dipodomys*	56	184.6	128.6
434	Catarrhini	26.4	172.8	146.4
Mean of 66 maximum deep coalescences				135.5

time inferred from dos Reis et al.'s (2012) timetree and four different estimates for the length of one CU (Table 5). In reality there is no single conversion between time and CU length, and all four CU length estimates may be appropriate in different parts of the mammalian tree. However, all of our independent estimates

of branch lengths in CUs are much greater than those derived from Song et al.'s (2012) stunted MP-EST tree (Table 5).

Song et al. (2012) used Robinson–Foulds (RF) distances to argue that the multispecies coalescent model accounts for 77% of the topological variation among 447 gene trees in their data set
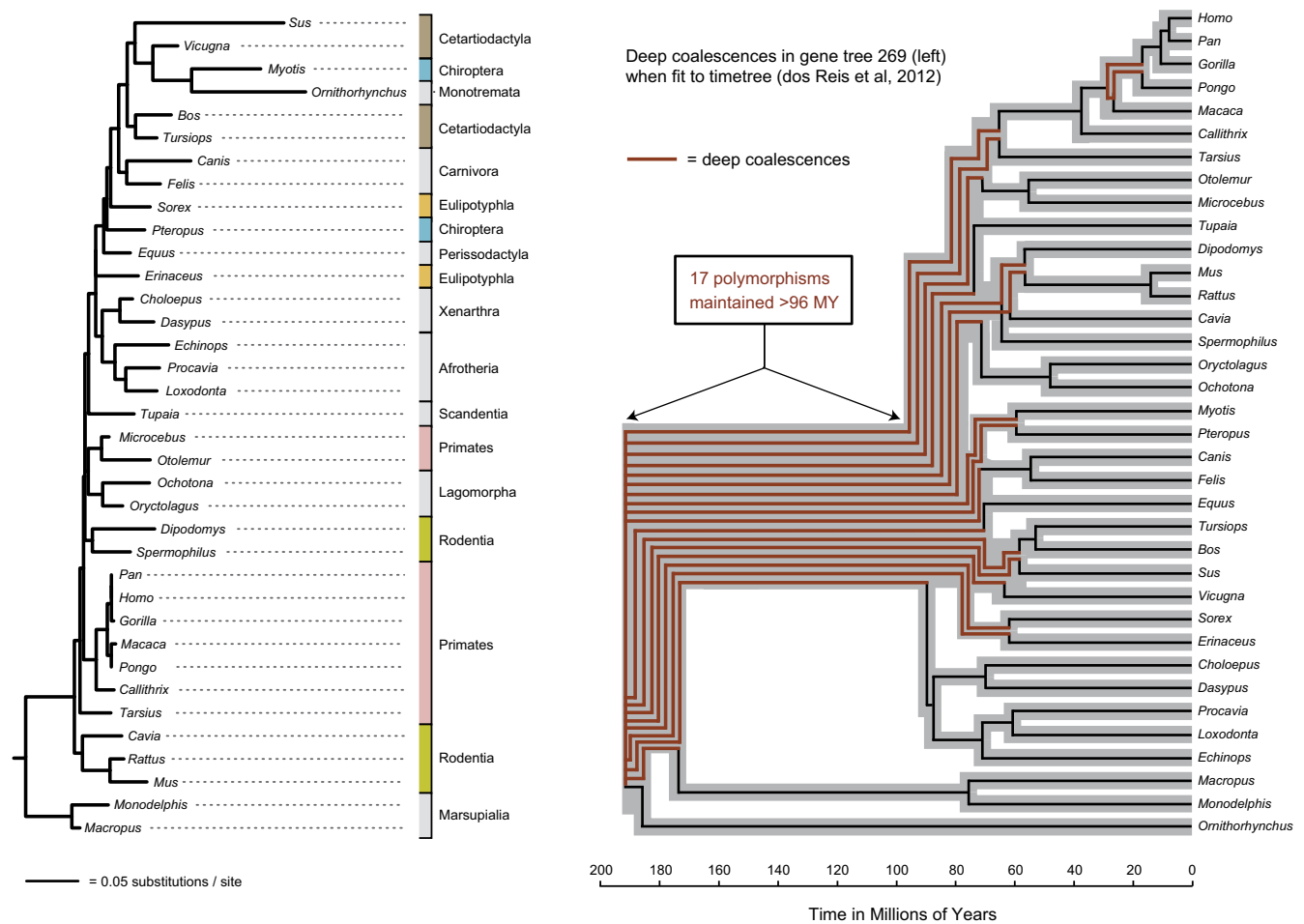
Fig. 13. Deep coalescences (right) implied by Song et al.'s (2012) gene tree 269 (left). Song et al.'s (2012) maximum likelihood tree for gene 269 was inferred with PhyML using NNI branch swapping and a GTR model of sequence evolution. Note that the monotreme *Ornithorhynchus* (platypus) is nested deeply within Placentalia, and multiple mammalian orders are not monophyletic (colored bars to the right of genus names). Deep coalescences (red branches) implied by gene tree 269 when this gene tree is fit to dos Reis et al.'s (2012) timetree suggest maintenance of 17 polymorphisms for >96 MY on the stem lineage of Placentalia. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 15). Specifically, the ratio between the expected RF distance, which was determined through simulations, and the observed RF distance, which was measured on real gene trees, was taken as a direct measure of the percentage of gene tree variation that is accounted for by the multispecies coalescent. This procedure is valid if all of the expected gene tree variation is the result of ILS, but Song et al.'s (2012) simulations were circular and their expected RF distance conflates multiple sources of gene tree incongruence that are unrelated to ILS. Song et al.'s (2012) simulations were based on their 447-gene MP-EST tree with its stunted branches that resulted from high levels of gene tree inaccuracy. This stunted MP-EST species tree was used to simulate 10,000 gene trees. RF distances were then calculated for the simulated trees and the median RF distance was used as an estimate of the expected gene tree variation under the multispecies coalescent. The expected variation was then compared to the observed gene tree variation, which was derived by calculating the median RF distance for the 447 gene trees estimated from their aligned DNA sequences. The ratio between the median expected RF distance and the median observed RF distance (20/26) was taken to indicate that the multispecies coalescent model accounts for 77% of the gene tree variation in the observed set of 447 gene trees. However, it would have been more appropriate to simulate gene trees based on an independent estimate of coalescent branch lengths to gauge

the magnitude of gene tree variation that is accounted for by the multispecies coalescent.

We employed dos Reis et al.'s (2012) timetree in conjunction with above-mentioned empirical estimates for the temporal duration of a CU in mammals to convert branch lengths in millions of years to branch lengths in CUs. These estimates for one CU are as follows: 0.4 MY (Patel et al., 2013); 0.8 MY (Rannala and Yang, 2003); 1.25 MY (Carstens and Dewey, 2010); and 2.75 MY (Hobolth et al., 2007). We simulated 1000 gene trees with species trees that employed each of these branch length conversions. Next, we calculated median RF distances relative to dos Reis et al.'s (2012) timetree. The median RF distance using Patel et al.'s (2013) CU of 0.4 MY is 0 and suggests that the multispecies coalescent accounts for 0% of the observed gene tree variation among Song et al.'s (2012) 447 gene trees (Fig. 15). If we use the mean instead of the median then the multispecies coalescent accounts for 0.67% (i.e., [0.172/25.6] × 100) of the observed gene tree variation. The median RF distance with Rannala and Yang's (2003) CU (0.8 MY) is also 0 and suggests that the multispecies coalescent accounts for 0% of the gene tree variation among Song et al.'s (2012) gene trees (3.88% if we use mean RF of 0.992) (Fig. 15). The median RF distance of 1 with Carstens and Dewey's (2010) conversion suggests that the multispecies coalescent accounts for 3.85% ([1/26] × 100) of the observed gene tree variation (8.88%
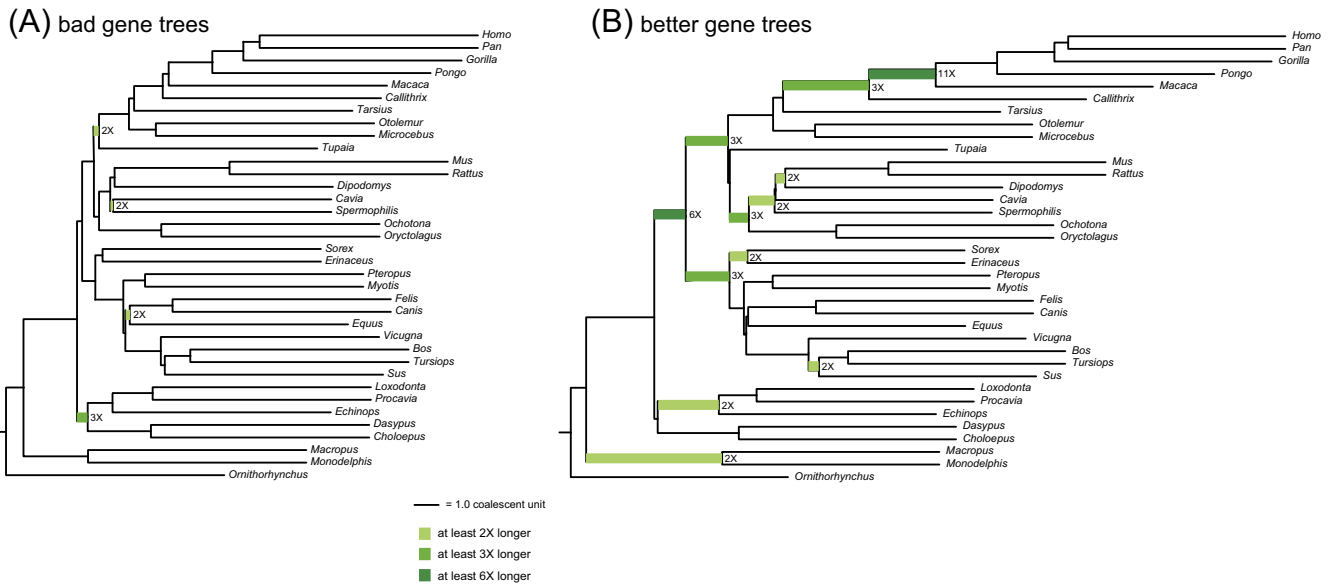
Fig. 14. MP-EST trees based on 447 gene trees from Song et al. (2012) that employed NNI branch swapping and a GTR model of sequence evolution (A) and 413 better gene trees based on NNI + SPR branch swapping and a GTR + Γ model of sequence evolution (B). Internal branch lengths on both MP-EST trees are to the same scale in coalescent units (CUs); terminal branch lengths are arbitrarily set at 9 CUs. Internal branches that are at least 2×, 3×, or 6× longer on one of the two MP-EST trees are highlighted in different shades of green. In two instances where the MP-EST trees differ (position of *Tupaia*, position of *Cavia*), branch lengths were compared for conflicting internal branches. See Fig. 11 for clade names. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with mean RF distances) (Fig. 15). Finally, the median RF distance of 4 with Hobolth et al.'s (2007) CU suggests that the multispecies coalescent accounts for as much as 15.38% ([4/26] × 100) of the variation in observed gene trees (30.1% with mean RF distances) (Fig. 15).

In all cases, gene tree variation that results from ILS is trumped by gene tree variation that results from other factors that affect gene tree accuracy including inadequate search algorithms, long branch misplacement, non-homology, missing data, investigator errors, etc. The percentage of gene tree variation that is accounted for by the multispecies coalescent is higher (0–22.2%) with more rigorous PhyML searches (SPR + NNI branch swapping) and a GTR + Γ model of sequence evolution for Song et al.'s (2012) 447 genes, as well as with Mirarab et al.'s (2014a) 424 RAxML gene trees (0 to 25.0%) and our reduced set of 413 PhyML gene trees (0–25.0%). However, other factors are still more important than ILS in accounting for gene tree variation. Importantly, the results of non-circular simulations (Fig. 15) contradict Song et al.'s (2012) assertion that ILS is the major source of gene tree heterogeneity in placental mammal phylogeny. This conclusion is strongly corroborated by (1) widespread gene tree conflicts at nodes where no conflict is expected based on many, unopposed, insertions of transposons (Fig. 11A), (2) gross conflicts in Song et al.'s (2012) gene trees that disappear after their data are analyzed using models of sequence evolution that they recommended, but chose not to use in their analyses (Fig. 11B), and (3) the numerous conflicts that disappear once errors that cannot be blamed on deep coalescence are fixed (e.g., mislabeled taxon names) (Fig. 7A).

Consensus summaries of the simulated gene trees that were generated with four different CUs suggest that each of the different conversions for time and CU length may be appropriate for different parts of the mammalian tree. Hobolth et al.'s (2007) conversion (2.75 MY per CU) yields a population of simulated trees with 68% support for *Homo + Pan*. This value broadly agrees with Hobolth et al.'s (2007) estimate of 62% (mean of four autosomal regions) for the fraction of the genome that supports *Homo + Pan*. However, only 78% of the simulated gene trees based on Hobolth et al.'s (2007) conversion support Haplorhini, which is supported by 104

transposons without opposition (Hartig et al., 2013) and is seemingly not a candidate for deep coalescence. Here, only Patel et al.'s (2013) conversion results in a population of gene trees with 100% support for Haplorhini. Other nodes with unopposed transposon support include Boreoeutheria, Euarchontoglires, and Laurasiatheria. Each of these nodes occurs on 100% of the simulated gene trees based on Patel et al.'s (2013) CU, but support for these nodes falls below 100% with Hobolth et al.'s (2007) CU. At the same time, Hobolth et al.'s (2007) CU yields reasonable support values for the laurasiatherian polytomy (e.g., 46% of simulated trees support Chiroptera + Perissodactyla + Carnivora). Unlike some of the above-mentioned clades that have been recovered with unanimous transposon support and are poor candidates for deep coalescence, a variety of competing hypotheses for relationships in the laurasiatherian polytomy are supported by one or more transposons and are candidates for deep coalescence (Nishihara et al., 2006; Hallström et al., 2011).

3.2.10. Summary of insufficiencies in Song et al.'s (2012) data and analyses

Song et al. (2012) claimed to have resolved difficult nodes in placental mammal phylogeny including the placental root, the placement of Scandentia, and the laurasiatherian polytomy using shortcut coalescence methods that ostensibly account for ILS. However, Song et al.'s (2012) dataset and analyses are beset by conceptual and mechanical problems (Figs. 2 and 6–9) that nullify all of their essential systematic conclusions (Fig. 6). STAR and ASTRAL bootstrap analyses with Song et al.'s (2012) gene trees fail to provide robust bootstrap support (>95%) for the position of treeshrews and one or more contentious nodes in the laurasiatherian polytomy that Song et al. (2012) claimed to have resolved with robust support. Indeed, two of three coalescence methods (STAR and ASTRAL) support a different tree from the one that Song et al. (2012) claimed was robustly supported. Further, Song et al.'s (2012) "coalescence" analyses are underpinned by switched taxonomic names (Fig. 7A), data duplications (Fig. 7B), non-homology problems (Fig. 8), and missing data for certain taxa (Fig. 9). These problems afflict at least 25% of the gene trees in Song

Table 5

Number of coalescent units (CUs) on the basis of MP-EST trees, branch lengths in millions of years (MY) on the basis of dos Reis et al.'s (2012) timetree, and the number of CUs on the basis of dos Reis et al.'s (2012) timetree with four different estimates from the literature (Patel et al., 2013; Carstens and Dewey, 2010; Rannala and Yang, 2003) for the duration of one CU.

Stem branch	Number of CUs on MP-EST tree for 447 gene trees from Song et al. (2012)	Number of CUs on MP-EST tree for 447 gene trees from Mirarab et al. (2014c)	Number of CUs on MP-EST tree for 413 gene trees	Millions of years on dos Reis et al. (2012) timetree	Number of CUs on dos Reis et al. (2012) timetree with Patel et al. (2013) conversion	Number of CUs on dos Reis et al. (2012) timetree with Carstens and Dewey (2010) conversion	Number of CUs on dos Reis et al. (2012) timetree with Rannala and Yang (2003) conversion	Number of CUs on dos Reis et al. (2012) timetree with Hobolth et al.'s (2007) conversion
Theria	0.7	0.8	0.6	11.8	29.5	9.4	14.8	4.3
Metatheria	2.7	2.6	5.6	97.7	244.3	78.2	122.1	35.5
Eutheria	2.2	1.9	2.8	83.7	209.3	67.0	104.6	30.4
Atlantogenata	0.4	0.1	0.1	2.0	5.0	1.6	2.5	0.7
Xenarthra	2.6	3.9	3.4	17.7	44.3	14.2	22.1	6.4
Afrotheria	1.0	2.6	2.5	16.7	41.8	13.4	20.9	6.1
Paenungulata	1.7	1.9	1.6	10.1	25.3	8.1	12.6	3.7
Boreoeutheria	0.2	1.3	1.3	5.3	13.3	4.2	6.6	1.9
Euarchontoglires	0.5	1.7	1.8	8.0	20.0	6.4	10.0	2.9
Primates	1.2	1.5	1.9	5.9	14.8	4.7	7.4	2.1
Strepsirrhini	1.1	1.9	1.7	13.1	32.8	10.5	16.4	4.8
Haplorhini	0.2	0.3	0.4	3.2	8.0	2.6	4.0	1.2
Anthropoidea	1.2	1.0	3.6	27.6	69.0	22.1	34.5	10.0
Catarrhini	0.2	0.3	2.8	11.0	27.5	8.8	13.8	4.0
Hominidae	1.8	2.0	2.5	9.3	23.3	7.4	11.6	3.4
Homininae	1.3	1.6	2.4	6.7	16.8	5.4	8.4	2.4
Homo + Pan	0.7	0.6	0.6	2.2	5.5	1.8	2.8	0.8
Tupaia + Glires[a]	0.2	0.1	0.1	1.7	4.3	1.4	2.1	0.6
Glires	0.2	0.8	0.8	5.0	12.5	4.0	6.3	1.8
Rodentia	0.4	1.0	1.1	6.4	16.0	5.1	8.0	2.3
Muridae + *Dipodomys* + *Cavia*[b]	0.1	0.1	.04	3.1	7.8	2.5	3.9	1.1
Muridae + *Dipodomys*	0.2	0.4	0.4	5.3	13.3	4.2	6.6	1.9
Muridae	4.7	2.9	4.3	42.0	105.0	33.6	52.5	15.3
Lagomorpha	2.6	3.0	3.6	22.9	57.3	18.3	28.6	8.3
Laurasiatheria	0.5	1.9	1.8	7.8	19.5	6.2	9.8	2.8
Eulipotyphla	0.3	0.8	0.7	14.5	36.3	11.6	18.1	5.3
Variamana	1.2	0.5	0.6	3.3	8.3	2.6	4.1	1.2
Chiroptera	0.9	1.4	1.2	12.6	31.5	10.1	15.8	4.6
Fereuungulata[c]	0.1	0.1	0.1	1.0	2.5	0.8	1.3	0.4
Zoomata	0.2	0.1	0.1	1.8	4.5	1.4	2.3	0.7
Carnivora	1.8	2.5	2.9	15.7	39.3	12.6	19.6	5.7
Cetartiodactyla	1.4	2.3	2.6	11.3	28.3	9.0	14.1	4.1
Sus + Bos + Tursiops	0.2	0.4	0.4	3.3	8.3	2.6	4.1	1.2
Bos + Tursiops	1.1	0.9	1.2	5.3	13.3	4.2	6.6	1.9
Total number of CUs or Ma	35.8 CUs	45.3 CUs	57.5 CUs	495 Ma	1237.5 CUs	396 CUs	618.8 CUs	180.0 CUs

[a] *Tupaia* + Primates for Song et al.'s (2012) MP-EST tree and dos Reis et al.'s (2012) timetree.
[b] *Cavia* + *Spermophilus* for Song et al.'s (2012) MP-EST tree.
[c] Zoomata + Chiroptera for dos Reis et al.'s (2012) timetree.

et al.'s (2012) dataset. Further, all 447 of Song et al.'s (2012) gene trees are compromised by gross concatalescence (Fig. 2) and gene tree searches with NNI branch swapping and an overly simple GTR model of sequence evolution (Fig. 11). Concatalescence analyses with improved gene trees for Song et al.'s (2012) dataset provide more support for an association of treeshrews with Glires than with Primates (Mirarab et al., 2014b) (Fig. 12), in contrast to another concatalescence analysis with even poorer taxon sampling that favored treeshrew + Primates (Kumar et al., 2013). Published coalescence/concatalescence studies of genome-scale datasets also favor topologies that directly conflict with those of Song et al. (2012) for both the placental root (McCormack et al., 2012) and the laurasiatherian polytomy (Hallström et al., 2011; Tsagkogeorga et al., 2013; Liu et al., 2015). Most recently, Liu et al. (2015) argued in favor of a Perissodactyla + Cetartiodactyla clade based on shortcut coalescence analyses even though this clade directly conflicts with Song et al.'s (2012) earlier resolution of this difficult problem with the same shortcut coalescence methods (Fig. 6). We make no strong claims regarding particular resolutions of these intransigent polytomies in mammalian phylogeny (Fig. 1), but contend that recent coalescence (McCormack et al., 2012) and concatalescence (Hallström et al., 2011; Song et al., 2012; Kumar et al., 2013; Tsagkogeorga et al., 2013; Liu et al., 2015) studies have not solved these difficult problems with robust support.

3.3. Simulations with illegitimate branch length switches

Gatesy and Springer (2014) reviewed simulations that directly compared the performance of coalescence methods to concatenation approaches for species tree reconstruction at deep timescales and concluded that concatenation had consistently outperformed shortcut coalescence on the basis of published studies (Leaché and Rannala, 2011; Aguiar and Schrago, 2013; Bayzid and Warnow, 2013; Patel et al., 2013). By contrast, Mirarab et al. (2014a,b) performed simulations based on an MP-EST species tree derived from analysis of Song et al.'s (2012) 447-gene dataset and concluded that the MP-EST coalescence method outperforms concatenation under biologically realistic conditions, except when gene trees have poor phylogenetic signal or ILS levels are too low. Zimmermann et al. (2014) concluded that BBCA (boosted-binned coalescent-based analysis), a method that combines *BEAST with
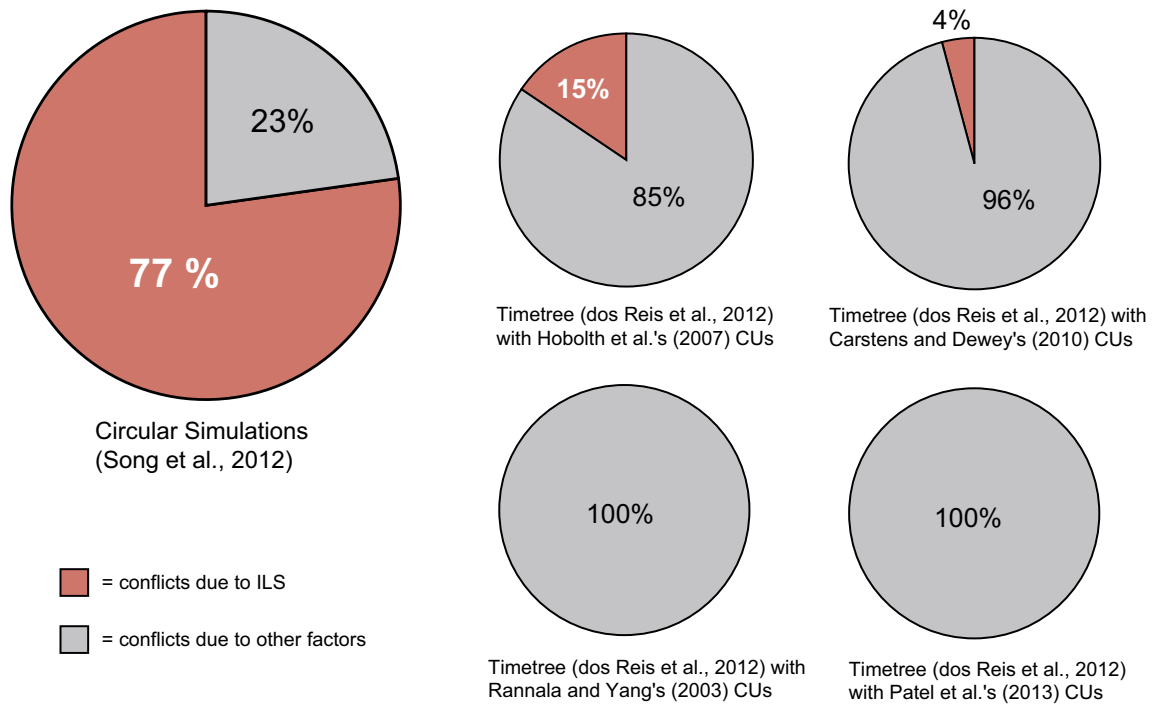
Fig. 15. Results of simulations that show the percentage of gene tree heterogeneity that is accounted for by incomplete lineage sorting (ILS; red) versus other sources of gene tree conflict such as model mis-specification, long branch misplacement, missing data, limited phylogenetic signal, etc. (gray). Song et al. (2012) concluded that the multispecies coalescent accounts for 77% of gene tree heterogeneity, but their simulations are circular and are based on a severely stunted MP-EST tree. By contrast, simulations based on dos Reis et al.'s (2012) timetree and four different estimates for the duration of one coalescent unit (Rannala and Yang, 2003; Hobolth et al., 2007; Carstens and Dewey, 2010; Patel et al., 2013) suggest that ILS accounts for a much smaller fraction of gene tree heterogeneity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

MP-EST, is more accurate than concatenation based on simulations that are underpinned by the same MP-EST tree. Mirarab et al. (2014c) reported the results of simulation studies that compared the performance of a new "coalescence" method, statistical binning with MP-EST, to other methods for constructing species trees including unbinned MP-EST and concatenation. These authors concluded that binned MP-EST is consistently and significantly more accurate than concatenation and unbinned MP-EST, and that unbinned MP-EST is also more accurate than concatenation when gene trees have sufficient accuracy. Based on these simulations, Roch and Warnow (2015) concluded that shortcut coalescence methods, especially statistical binning with MP-EST, are robust to gene tree estimation error and perform better than concatenation in some situations. These assertions purportedly contradict Gatesy and Springer's (2014) conclusion, based on simulation results that were published prior to 2014, that concatenation consistently outperforms shortcut coalescence methods in simulations at deep timescales. However, all of the new simulation studies (Mirarab et al., 2014a,b,c; Zimmermann et al., 2014), which underpin Roch and Warnow's (2015) conclusions, employed two different branch length switch procedures that invalidate their results.

Mirarab et al. (2014a,b,c) started with gene trees that they derived from Song et al.'s (2012) 447-gene dataset and generated an MP-EST species tree based on these gene trees. The individual gene trees exhibit high levels of topological variation that cannot be attributed to ILS and instead result from incorrect taxon labels, alignments of non-homologous sequences, missing data, and long branches (Figs. 7–11 and 16). A consequence of this gene tree inaccuracy is that branch lengths on Mirarab et al.'s (2014a,b,c) MP-EST species tree (in CUs) are stunted by at least an order of magnitude for some branches (Fig. 17, Table 5). Mirarab et al.'s (2014a,b,c) MP-EST species tree with its stunted branch lengths was then used to simulate gene trees with branch lengths in CUs. At the next critical step, Mirarab et al. (2014a,b,c) swapped stunted branch

lengths on their simulated gene trees with branch lengths from their real gene trees that are not similarly stunted and instead are often an order of magnitude too long. This is the first of two significant errors in Mirarab et al.'s (2014a,b,c) simulation procedure. Mirarab et al. (2014a,b,c) ostensibly employed this branch length swap to control for gene tree error among various simulations that they executed in which species tree depth was varied, but their branch length switch biased their simulations to favor coalescence methods by rescuing them from their Achilles' heel – inaccurately reconstructed gene trees (Huang and Knowles, 2009; Huang et al., 2010; Bayzid and Warnow, 2013; Patel et al., 2013; Gatesy and Springer, 2014; Mirarab et al., 2014b). Mirarab et al. (2014a,b,c) compounded this problem with branch length multipliers that exacerbated the disparity between branch lengths on their simulated trees and branch lengths from their empirically estimated gene trees. This is the second critical error in Mirarab et al.'s (2014a,b,c) simulation procedure. Details of these inappropriate branch length switches are outlined in Fig. 16 and described in more detail below.

The MP-EST tree that Mirarab et al. (2014a,b,c) used for their simulations was inferred from RAxML gene trees (GTR + Γ) derived from Song et al.'s (2012) 447-gene dataset. These trees are more congruent with the inferred MP-EST tree than are Song et al.'s (2012) gene trees, but are still compromised by inaccurate topological relationships that are unrelated to ILS. Twenty-one trees retain incorrect taxon labels wherein a wallaby (*Macropus*) and a primate (*Macaca*) were inadvertently switched (Fig. 7A). Mirarab et al. (2014a) reported this error in Song et al.'s (2012) dataset, but their discovery of this error post-dated their own simulations (Mirarab, pers. com). In addition, Mirarab et al.'s (2014a,b,c) gene trees are based on a dataset that includes eight duplicated loci (Fig. 7B), 26 loci with alignments of non-homologous sequences (Fig. 8), and numerous loci with high percentages of missing data for one or more taxa (Fig. 9). These and other problems such as long-branch

Illogical Simulation Protocol

Inaccurate Gene trees with "Normal" Branches
1) Many gene trees compromised by missing data (Fig. 9)
2) 21 gene trees with "monkeyroo" swaps (Fig. 7)
3) 26 gene trees with non-homology issues (Fig. 8)
4) 42 gene trees with deep coalescences >100 MY
5) Gene trees are **inaccurate** but branches **not stunted**

Use inaccurate gene trees to infer species tree with MP-EST

MP-EST Species Tree with Stunted Branches
1) Branches of species tree are **severely stunted** due to **inaccurately reconstructed** gene trees
2) Independent analyses suggest species tree branch lengths are severely **stunted** (Fig. 17; Table 5)

Simulate gene trees based on stunted MP-EST species tree

Simulated Gene Trees (Conflicting and Stunted)
1) Extensive **ILS + conflicts** in simulated gene trees due to **stunted branches** of MP-EST species tree
2) Internal branches in simulated gene trees are **stunted** and reflect **stunted** branches of MP-EST species tree

Replace stunted gene tree branches with real gene tree branches that are not stunted and simulate sequence data for coalescence versus concatenation comparisons

Simulation Results Invalid due to Branch Swaps
1) Coalescence rescued from gene tree reconstruction errors by switched, **artificially long** gene tree branches
2) Concatenation hindered by **inflated gene tree conflicts**
3) **Errors compounded** by 0.5X and 0.2X multipliers of the **already stunted** MP-EST model species tree (Fig. 18)
4) Simulation results are not valid due to inadvertent **branch length shell games** in the protocol
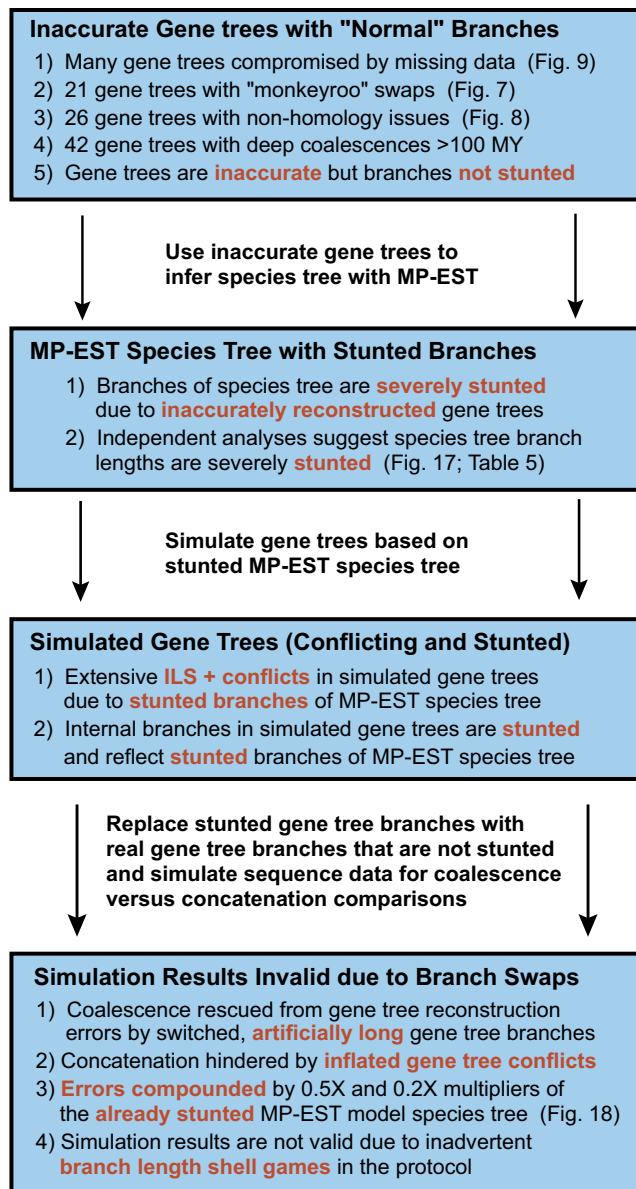
Fig. 16. Flowchart that illustrates illegitimate branch length switches and branch length multipliers that were employed by Mirarab et al. (2014a,b,c) in their simulations that compared coalescence, binning, and concatenation methods for species tree estimation. Mirarab et al. (2014a,b,c) employed better searches than Song et al. (2012), but branches on their MP-EST tree (Figs. 17 and 18) are still severely stunted because their 447 gene trees are based on a dataset that included 21 genes with switched taxon labels (*Macropus* and *Macaca*) (Fig. 7), 26 genes with non-homologous sequences (Fig. 8; Table 2), and numerous genes with missing data problems (Fig. 9; Table 3). Forty-two of their gene trees imply unrealistic deep coalescences that exceed 100 million years, which is greater than the age of placental mammals based on dos Reis et al. (2012).

misplacement resulted in extensive gene tree error among Mirarab et al.'s (2014a,b,c) 447 gene trees. Forty-two of their gene trees imply unrealistic deep coalescences that exceed 100 MY. Within Placentalia, 41.4% of these trees did not recover Haplorhini (*Tarsius* + Anthropoidea). The failure of these gene trees to recover Haplorhini cannot be attributed to deep coalescence given that this clade is supported by 104 transposons without any opposing support (Hartig et al., 2013). Due to abundant gene tree errors that cannot be attributed solely to ILS (Figs. 7–9 and 11), Mirarab et al.'s (2014a,b,c) MP-EST tree, like that of Song et al. (2012), is severely

stunted with internal branch lengths (in CUs) that comprise only 3.7–25.2% of the expected sum of internal branch lengths based on a well calibrated molecular timetree (dos Reis et al., 2012) and four different estimates for the duration of one CU (Rannala and Yang, 2003; Hobolth et al., 2007; Carstens and Dewey, 2010; Patel et al., 2013) (Table 5). Fig. 17 shows internal branches (in CUs) on Mirarab et al.'s (2014a,b,c) MP-EST tree that are dwarfed by internal branches (in CUs) on dos Reis et al.'s (2012) timetree with Carstens and Dewey's (2010) estimate for the duration of one CU (~9× difference). These differences are less exaggerated when Mirarab et al.'s (2014a,b,c) MP-EST tree is compared to dos Reis et al.'s (2012) timetree with Hobolth et al.'s (2007) CU estimate (~4× difference), but more exaggerated when Mirarab et al.'s (2014a,b,c) MP-EST tree is compared to dos Reis et al.'s (2012) timetree with CU estimates from Rannala and Yang (2003) (~14× difference) and Patel et al. (2013) (~27× difference) (Table 5).

After constructing their "model species tree" with MP-EST based on 447 gene trees, Mirarab et al. (2014a,b,c) simulated gene trees and sequence alignments that were utilized in comparisons that competed shortcut coalescence methods against concatenation. Mirarab et al. (2014a,b,c) replaced the coalescent branch lengths on their simulated gene trees with branch lengths that were drawn from their biological dataset (i.e., 447 gene trees that were reconstructed from real sequences with RAxML). Terminal branch lengths for each taxon (e.g., *Homo*) in the simulated trees were randomly drawn from a pool of branch lengths from the biological dataset, in units of substitutions per site, for the same taxon. Internal branches on both the simulated and real trees (i.e., 447 gene trees that were reconstructed from real sequences with RAxML) were ordered and partitioned into 100 equally sized bins based on branch length. Internal branches in the simulated gene trees were then exchanged for internal branches from the real 'biological' trees that were in the same percentile, e.g., the top 1% of simulated internal branches (based on length) were switched with branches from the top 1% of internal branch lengths among the 447 biological trees. These 'true' gene trees with branch lengths in substitutions per site were then used to simulate sequence data, and gene trees were estimated from these sequences using RAxML.

Mirarab et al.'s (2014a,b,c) replacement procedure wherein internal branch lengths on their 1× simulated gene trees were replaced with branch lengths from their biological data set (Fig. 16) invalidated all of their coalescence versus concatenation comparisons by ignoring increased gene tree reconstruction error that is inexorably linked to shorter coalescence branch lengths. This is the first critical error in Mirarab et al.'s (2014a,b,c) simulation procedure. Branch lengths in substitutions per site on the 447 biological gene trees are not systematically stunted whereas coalescence branch lengths on the 1× simulated gene trees are too short by more than an order of magnitude for some branches. For example, the stem Catarrhini (Old World monkeys and apes) branch is 0.3 CUs on Mirarab et al.'s (2014a,b,c) compressed MP-EST tree, but is >29× longer, 8.8 CUs, on dos Reis et al.'s (2012) timetree with Carstens and Dewey's (2010) estimate for the duration of one CU (Table 5). Thus, coalescence branch lengths on Mirarab et al.'s (2014a,b,c) 1× simulated gene trees are not commensurate with branch lengths for 447 real genes that are in substitutions per site. The net result of this illegitimate switching procedure is that coalescence methods are liberated from gene tree error by gene tree branches that are inappropriately long whereas concatenation methods are penalized by the high diversity of gene trees that results from simulations with stunted coalescence branch lengths (Fig. 16).

This problem was exacerbated when Mirarab et al. (2014a,b,c) employed multipliers of 0.5× and 0.2× to internal branch lengths (in CUs) on their "model species tree" (=1× MP-EST tree) to increase ILS and the amount of conflict between individual gene
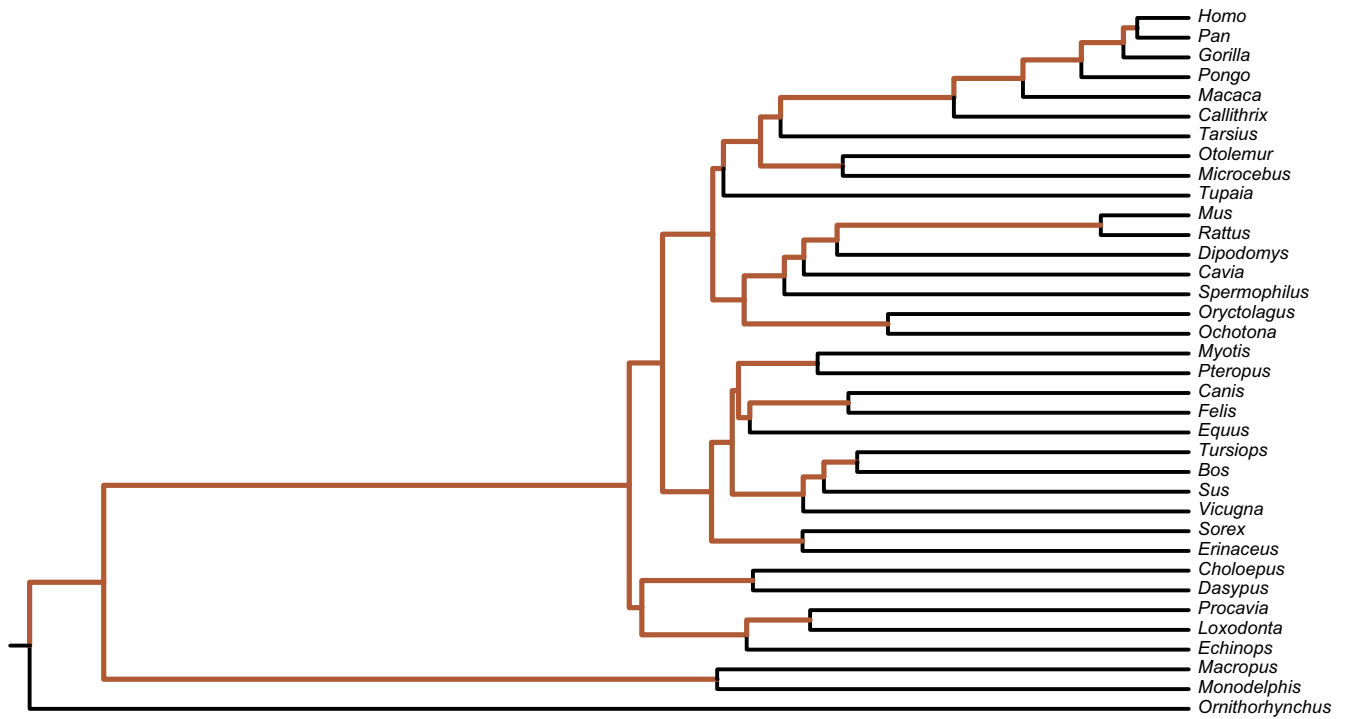
Carstens and Dewey (2010) CUs for dos Reis et al.'s (2012) timetree
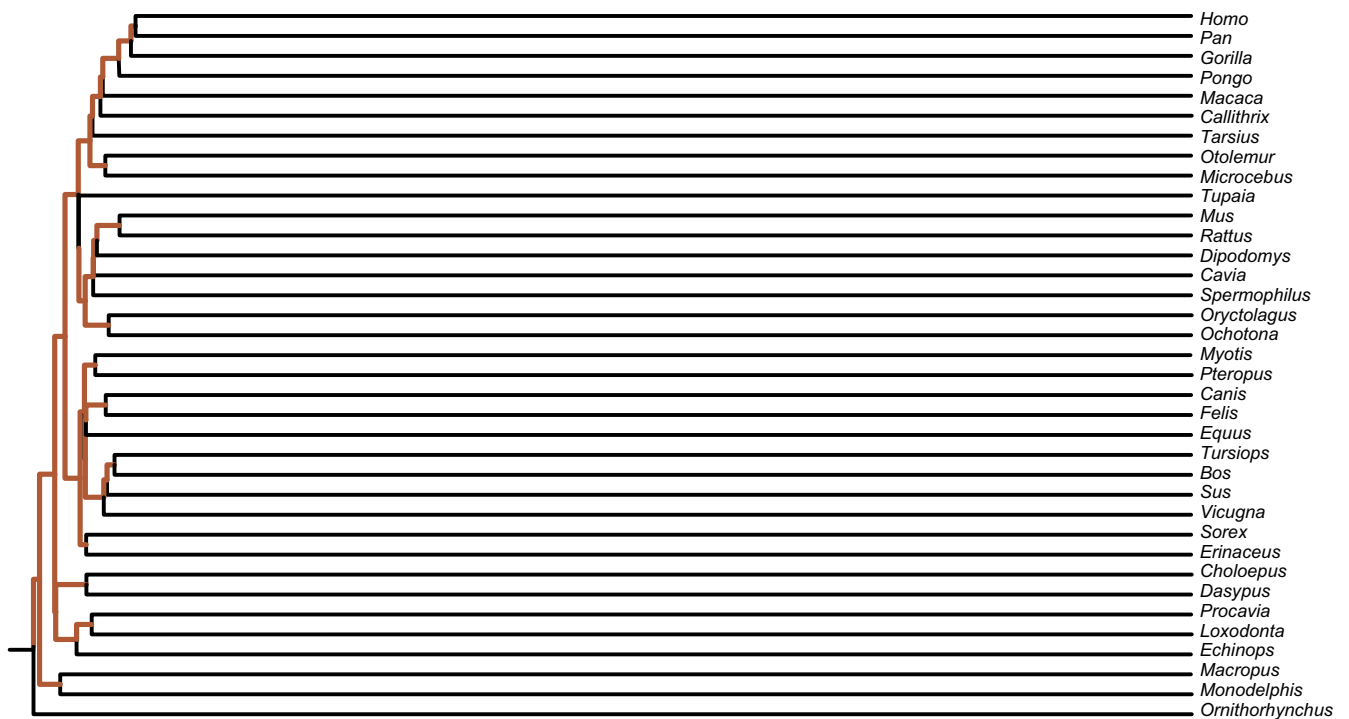
Mirarab et al.'s (2014a, b, c) 1X MP-EST tree used in simulations

Fig. 17. Species trees with internal branches in coalescent units (CUs) based on dos Reis et al.'s (2012) timetree for mammals with Carstens and Dewey's (2010) estimation for the duration of one CU (top) versus Mirarab et al.'s (2014a,b,c) MP-EST species tree utilized in simulations (bottom). Internal branches in both trees are to the same scale. Terminal branch lengths on MP-EST trees are undefined (Liu et al., 2010); terminal branches were extended on Mirarab et al.'s (2014a,b,c) MP-EST tree so that both trees in the figure have equivalent root to tip distances.
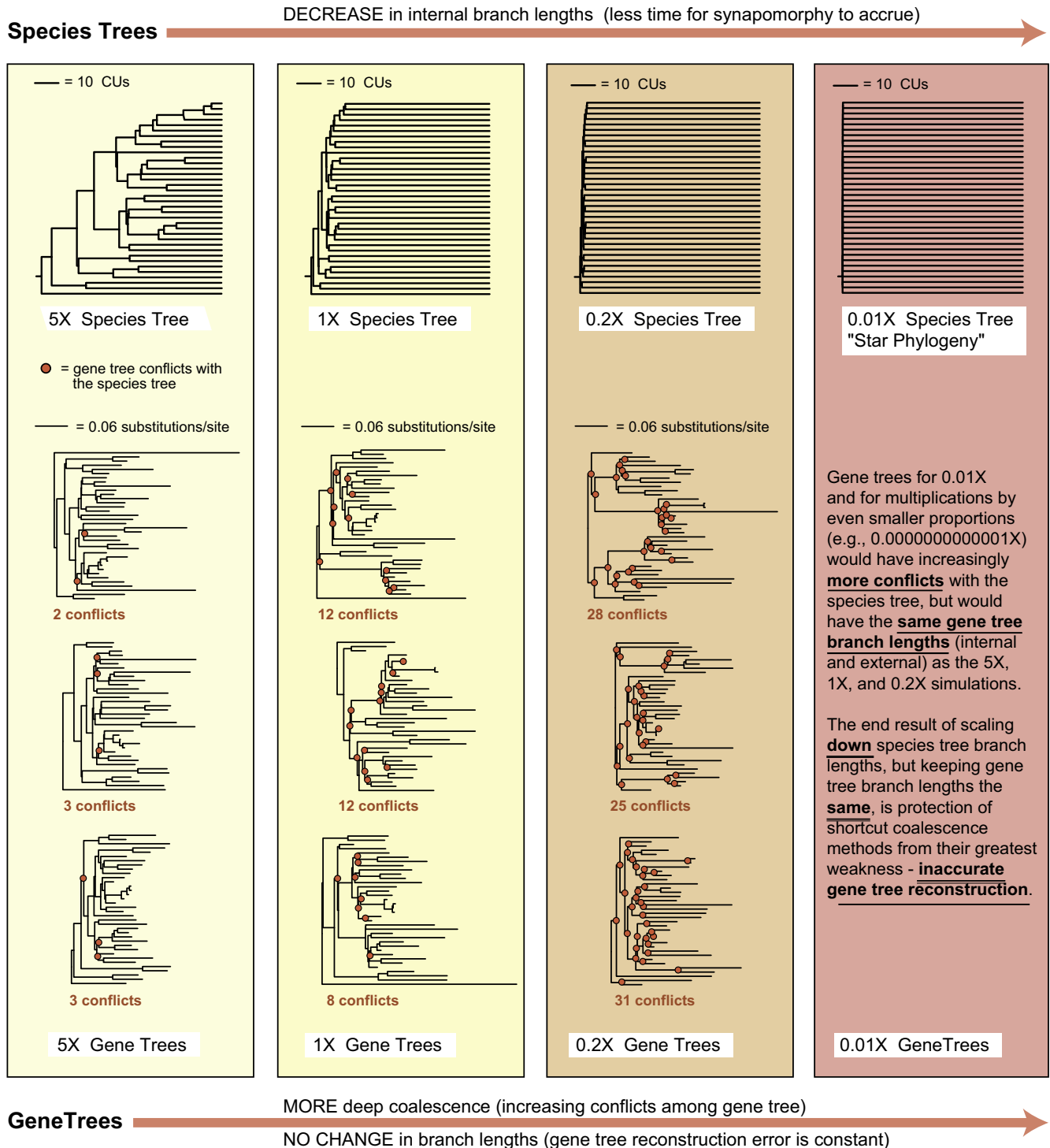
Fig. 18. Illegitimate swapping of branch lengths in Mirarab et al.'s (2014a,b,c) simulated gene trees that limited gene tree error for species trees that used multipliers of <1 to increase incomplete lineage sorting. Mirarab et al.'s (2014a,b,c) species tree (1× MP-EST) is shown along with additional species trees that result from 5×, 0.2×, and 0.01× multiplications of internal branch lengths. Mirarab et al. (2014a,b,c) did not employ a 0.01× multiplier, but this multiplier illustrates the logical problem with their procedure. A 0.01× multiplier yields a star phylogeny with internal branch lengths that approach zero coalescent units (CUs). Mirarab et al. (2014a,b,c) switched branch lengths in CUs on their simulated gene trees with branch lengths in substitutions per site from 447 RAxML gene trees. Gene trees that were simulated based on the 0.2× MP-EST species tree show much higher levels of conflict with the species tree than gene trees that were simulated based on the 1× MP-EST species tree (conflicts in three representative gene trees are shown for each simulation condition). Shorter internal branches should accompany gene trees that were simulated based on the 0.2× MP-EST tree, but Mirarab et al. (2014a,b,c) switched internal branches on all of their simulated gene trees (in CUs) with the same set of internal branches (in substitutions/site) from their 447 RAxML gene trees. This procedure rescued 0.2× simulations from high levels of gene tree reconstruction error that would have reduced the performance of shortcut coalescence methods. Mirarab et al.'s (2014a,b,c) procedure for switching branch lengths likewise would have protected a star phylogeny (e.g., 0.01× MP-EST) from gene tree error.

trees and the species tree (Fig. 18). However, instead of simulating sequence alignments based on these simulated gene trees with their shorter internal branch lengths, Mirarab et al. (2014a,b,c)

executed their second fundamental error and replaced coalescence branch lengths on their 0.5× and 0.2× simulated gene trees with real branch lengths from their 447 biological trees (Fig. 18).

A simple example illustrates this problem (Fig. 18). Suppose that Mirarab et al. (2014a,b,c) had employed a multiplier of 0.01× to internal branch lengths on their 1× MP-EST tree (instead of 0.5× or 0.2×). Such a multiplier would have resulted in a species tree that is essentially a star phylogeny with internal branch lengths that all approach zero (Fig. 18). If gene trees and sequence datasets were subsequently simulated from this species tree with internal branch lengths that approach zero, it would be impossible to recover robust, bifurcating gene trees and instead there would be high gene tree inaccuracy, which would then compromise the performance of shortcut coalescence methods as in previous studies that have used more legitimate simulation strategies (Leaché and Rannala, 2011; Bayzid and Warnow, 2013; Patel et al., 2013). By contrast, the procedure employed by Mirarab et al. (2014a,b,c) would have replaced all of these ~zero length internal branches with internal branch lengths from their real genes trees (Figs. 16 and 18). In some cases, these internal branch lengths exceed 0.25 substitutions per site. The result of employing these multipliers is that there would be no more gene tree reconstruction error associated with gene trees that were simulated from a star phylogeny (0.01× branches) than a tree with 1× branch lengths. Likewise, the 0.2× simulations of Mirarab et al. (2014a,b,c) have equivalent gene tree error relative to the 5× simulations from this same study (Fig. 18).

This "control" for gene tree error is described in detail in Mirarab et al.'s (2014b) initial simulation study, but this internal branch length switch cannot be justified from a biological perspective. Indeed, much of the debate surrounding concatenation versus coalescence has centered on the negative effects of increased gene tree reconstruction error when shortcut coalescence methods are employed to unravel rapid radiations in the Tree of Life (Bayzid and Warnow, 2013; Patel et al., 2013; Gatesy and Springer, 2014). When radiations occur rapidly, short species tree branch lengths provide no time for synapomorphies to accrue. This problem is accentuated at deep timescales where accurate reconstruction of gene trees is impossible for c-genes of realistic length (Fig. 3).

In summary, recent simulations (Mirarab et al., 2014a,b,c; Zimmermann et al., 2014) where coalescence methods outperform supermatrix methods are invalidated by improper internal branch length replacement procedures that unfairly favor coalescence methods. Valid simulations (e.g., Bayzid and Warnow, 2013; Patel et al., 2013) support Gatesy and Springer's (2014) previous conclusion that concatenation has consistently outperformed shortcut coalescence at deep timescales in head-to-head comparisons to date. Indeed, the "standard" 100 taxon/25 gene simulations that Mirarab et al. (2014a) executed in which gene tree topologies and branch lengths were properly generated via coalescence models once again favored concatenation relative to shortcut coalescence approaches (MP-EST and ASTRAL). This set of simulations from Mirarab et al. (2014a) is the exception that proves the rule and further validates our explanation for the exceptional performance of coalescence methods in the majority of Mirarab et al.'s (2014a,b,c) simulations. Thus, we argue once again that valid simulations of deep divergences favor concatenation over coalescence. Given the continued success of concatenation in comparison to coalescence methods at deep timescales, we encourage researchers who have shied away from directly competing concatenation versus shortcut coalescence in simulations studies to do so in future work (e.g., Lanier and Knowles, 2012, 2015; Lanier et al., 2014).

3.4. The gene tree delusion

Song et al. (2012) and other advocates of the shortcut coalescence approach (Shaw et al., 2012; Kumar et al., 2013; Zhong et al., 2013) have routinely applied shortcut coalescence methods, which assume no within-locus recombination, to analyze complete protein-coding sequences extracted from genomes. Loci employed in these studies often span more than 100,000 bp and in some case more than one megabase of the genome from start codon to stop codon. By contrast, estimates of recombination breakpoints for three taxa (human, chimp, gorilla) suggest that c-genes have a mean size of only 111 bp for this three-taxon problem (Hobolth et al., 2007). C-genes become smaller with more taxa owing to the recombination ratchet (Fig. 3). Calculations based on Hobolth et al.'s (2007) empirical data suggest that c-genes for Song et al.'s 37-taxon dataset have a mean size of 12 bp or less, which is more than four orders of magnitude shorter than the average length of Song et al.'s (2012) loci (139.6 kb). Moreover, this estimate (~12 bp) ignores branch length heterogeneity, which is an important source of gene tree variation (Edwards, 2009) that must be accounted for to achieve balanced gene tree stoichiometry (Fig. 4). Real c-genes may approach one or a few bp for Song et al.'s (2012) dataset when branch-length heterogeneity is properly accounted for, and approximate the basic unit of analysis in the concatenation approach, a single bp (Gatesy and Springer, 2014).

ILS is potentially an important source of gene tree variation at tight internodes, especially in the anomaly zone where the species tree may differ from the most probable gene tree (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007), but problems with ILS cannot be solved by applying shortcut coalescence methods to unrealistically long, "pseudo" c-genes that deny recombination and erase the unique histories of individual c-genes through concatalesence. This is the gene tree delusion that has been promoted by advocates of shortcut coalescence methods. C-gene lengths are determined exclusively by recombination. Advocates of coalescence methods may prefer longer c-genes with their more accurate "gene" trees, but c-genes that are detached from biological reality are not useful. Coalescence methods based on distantly linked, evolutionarily independent SNPs (RoyChoudhury et al., 2008; Bryant et al., 2012; Chifman and Kubatko, 2014) avoid the recombination ratchet. For this reason alone, such methods merit further pursuit and critical evaluation.

Empirical estimates of c-gene length (Fig. 3A) also have critical implications for simulation studies where coalescence methods and concatenation are competed against each other. These studies typically simulate genes that are 500 bp, 1000 bp, or even longer. Mirarab et al. (2014a,b,c) employed simulated gene lengths of 250, 500, 1000, and 1500 bp for 37-taxon trees that were modeled after Song et al.'s (2012) mammalian data set. However, calculations based on empirical results (Table 1; Hobolth et al., 2007, 2011) suggest that it would have been more appropriate for Mirarab et al. (2014a,b,c) to simulate 10 bp c-genes for their studies that competed coalescence against concatenation. Mirarab et al.'s (2014a,b,c) simulated c-genes that are too long, coupled with inappropriate branch-length switches and branch length multipliers (see previous section), yielded simulation conditions that unfairly favored shortcut coalescence methods relative to concatenation. Future simulation studies that model ancient divergences in the Tree of Life should be informed by biologically realistic estimates of recombination rates and c-gene sizes, and should avoid branch length switches and multipliers that unfairly rescue coalescence methods from their Achilles' heel, inaccurate reconstruction of gene trees (Bayzid and Warnow, 2013; Patel et al., 2013; Gatesy and Springer, 2014).

3.5. The future of systematics at deep timescales: shortcut coalescence, concatenation, and beyond

It is perhaps counterintuitive that we have harshly criticized the shortcut concatalescence approach to systematics (Song et al.,

2012) when the supermatrix paradigm entails concatenation of all c-genes, rather than just some, into a single matrix (de Queiroz and Gatesy, 2007). Theory and simulations under completely neutral conditions suggest that concatenation can fail to reconstruct the correct species tree in the anomaly zone due to this merging of c-genes with divergent histories (Kubatko and Degnan, 2007), but unlike shortcut coalescence methods, the supermatrix approach is impacted much less by gene tree reconstruction errors because estimation of many separate gene trees is unnecessary when all characters are concatenated. Instead, an overall species tree is derived under the assumptions that individual gene trees have the same topology as the species tree and that conflicting character evidence at a particular node may be due to "standard" homoplasy (i.e., reversal and convergence) as well as "mixed signal" homoplasy (e.g., hemiplasy, lateral transfer, introgression; Doyle, 1997).

These assumptions of concatenation are not entirely realistic in view of ILS and anomaly zone problems, hitchhiking/linkage effects, etc. However, the critical issue is whether these assumptions are better or worse than the various assumptions of shortcut coalescence/concatalescence methods, especially if actual c-genes are as tiny as suggested above and approach the size of the basic unit of analysis in concatenation (Figs. 3–5; Gatesy and Springer, 2014). Advocates of shortcut coalescence methods have repeatedly asserted that these methods are superior to concatenation based on theory and simulations (Liu and Edwards, 2009; Song et al., 2012), but these same authors have turned a blind eye to the deficiencies of their own methods. Recently, Lemmon and Lemmon (2013) outlined a pipeline procedure for phylogenomic data analyses (their Fig. 4) and suggested three different bins for existing phylogeny estimation methods: recommended, suboptimal or requires validation, and not recommended. Supermatrix methods were placed in the "not recommended" waste bin whereas shortcut coalescence methods (STAR, STEAC, STEM, MP-EST) were categorized as "recommended" because these methods ostensibly accommodate gene tree variation. However, concatenation and coalescence methods are both compromised by unrealistic assumptions (Gatesy and Springer, 2014), and Lemmon and Lemmon's (2013) unjustified dismissal of concatenation in favor of coalescence is unwarranted given that coalescence methods are confronted by their own problems (e.g., distorted gene tree stoichiometry, the recombination ratchet, and the inaccuracy of estimated gene trees). Valid simulations of deep divergences that have compared concatenation to shortcut coalescence are currently lacking due to the use of c-genes that are likely too long relative to empirical estimates (Fig. 3; Hobolth et al., 2007, 2011). Yet, in simulations that were otherwise properly executed, concatenation has consistently bested shortcut coalescence methods at depth despite these unrealistically long c-gene sizes that should have favored coalescence methods (Leaché and Rannala, 2010; Bayzid and Warnow, 2013; Patel et al., 2013; 100 taxon/25 gene simulations of Mirarab et al., 2014a).

Inaccurate reconstruction of c-gene trees is a widely acknowledged weakness of shortcut coalescence methods (Huang and Knowles, 2009; Huang et al., 2010; Bayzid and Warnow, 2013; Patel et al., 2013; Gatesy and Springer, 2014; Mirarab et al., 2014b). This weakness is even more pronounced than generally acknowledged if conflicting c-genes are as short as suggested by recent estimates from mammalian genomes. Even with "pseudo-c-genes" that average 3100 bp, extensive gene tree reconstruction errors are apparent where none are expected based on independent evidence (e.g., Haplorhini, Boreoeutheria, Laurasiatheria, Euarchontoglires, and Afrotheria nodes; Fig. 11), and statistical consistency of shortcut coalescence methods is not guaranteed unless gene trees are reconstructed accurately (Liu et al., 2009a,b, 2010; Roch and Warnow, 2015). Thus,

concatalescence of multiple c-genes as in Song et al. (2012) provides a compounded set of problems: extensive gene tree reconstruction error resulting from a variety of factors (long branch misplacement, insufficient phylogenetic signal, model misfit, etc.), as well as direct violation of a basic assumption of shortcut coalescence methods – no recombination within loci (i.e., error due to concatenation of multiple c-genes). The question then becomes, "which defects of the various methods (coalescence versus concatenation versus concatalescence) are most problematic in particular phylogenetic circumstances?".

An additional concern for the application of shortcut coalescence/concatalescence methods centers on unrealistic assumptions of the multispecies coalescent. Although rarely acknowledged, the assumption of complete neutrality in the coalescence approach to systematics (Liu et al., 2009a) describes an anti-Darwinian dystopia in which there is no natural selection of any kind. With their assumption of neutrality, coalescence methods require constant N_e across the genome. However, the differential action of selection translates to variation in N_e across the genome (Charlesworth, 2009). Balancing selection results in an elevation of N_e at sites that are in close proximity to the target of selection (Charlesworth, 2009). By contrast, purifying or background selection results in reduced N_e at sites that are closely linked to a deleterious mutation (Charlesworth, 1993, 2009; Hobolth et al., 2011). The effect of positive directional selection is similar to that of background selection, but is more extreme and can lead to an extended period without any ILS due to severe reduction of N_e at, and near, the target of selection (Hobolth et al., 2011). Scally et al. (2012) concluded that ILS varies with respect to genome position for the human–chimp–gorilla problem and that suppression of ILS extends well beyond coding genes, so that selection "has affected almost all of the genome throughout hominid evolution" (p. 171). If selection regimes that reduce ILS (background selection, directional positive selection) are more prevalent than selection regimes that increase ILS (diversifying selection) (McVicker et al., 2009; Mukherjee et al., 2009; Scally et al., 2012), gene tree stoichiometry generally will be systematically biased in favor of gene trees that exhibit higher congruence with the species tree than would be the case for neutrally evolving loci. The impact of this potential bias on species tree estimation with shortcut coalescence methods that assume neutrality remains unclear, especially in the anomaly zone where the most common gene tree may differ from the species tree for a given node.

It is paradoxical that the effects of selection on N_e are well documented, whereas the "no selection" assumption of coalescence methods for species tree estimation has mostly been ignored. Predictable gene tree stoichiometry in current coalescence methods is predicated on the assumption of complete neutrality across every genome of each species in a phylogenetic analysis. Panmixia within each species and the absence of population bottlenecks are additional problematic assumptions of current shortcut coalescence methods. When these assumptions are not met (e.g., some loci or lineages under strong purifying selection, others under positive selection, strong population subdivision, severe bottlenecks), gene tree stoichiometry is not predictable and becomes distorted relative to neutral expectations. Recent comparative analyses suggest that large percentages of primate genomes have been impacted by selection, either directly or indirectly via hitchhiking (McVicker et al., 2009; Hobolth et al., 2011; Scally et al., 2012). This Darwinian world does not match the alternate, unrealistic neutral world of the coalescent, and this disconnect surely impacts gene tree stoichiometry and therefore phylogenetic inference.

As the field of systematics moves forward, it will be important to consider how genomes should be partitioned, sampled, and analyzed to derive robust estimates of species trees for ever larger samples of taxa. Given complete genome sequences from many

species, what is the basic unit of phylogenetic inference that should be applied and how should these basic units be delimited and analyzed? Possible units of analysis include genes, c-genes based on topological shifts along a chromosome, c-genes that account for all recombination break points (i.e., topological incongruence and branch length heterogeneity), and individual nucleotide sites. Aside from individual sites, however, delimitation of these various units is not a trivial exercise (Ané, 2010). A serious discussion of this issue in the context of empirical phylogenomic studies, as well as more realistic simulations, will be required to resolve the tightest phylogenetic splits that are deep in the Tree of Life and to determine the ultimate merits of concatenation, coalescence, and concatalescence strategies.

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.07.018.

References

Aguiar, B.O., Schrago, C.G., 2013. Conventional simulation of biological sequences leads to a biased assessment of multi-loci phylogenetic analysis. Evol. Bioinf. 9, 317–325.

Ané, C., 2010. Reconstructing concordance trees and testing the coalescent model from genome-wide data sets. In: Knowles, L.L., Kubatko, L.S. (Eds.), Estimating Species Trees: Practical and Theoretical Aspects. Wiley-Blackwell, Hoboken, NJ, pp. 35–52.

Arnason, U., Adegoke, J.A., Bodin, K., Born, E.W., Esa, Y.B., Gullberg, A., Nilsson, M., Short, R.V., Xu, X., Janke, A., 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. USA 99, 8151–8156.

Arnason, U., Adegoke, J.A., Gullberg, A., Harley, E.H., Janke, A., Kullberg, M., 2008. Mitogenomic relationships of placental mammals and molecular estimates of their divergences. Gene 421, 37–51.

Asher, R.J., Novacek, M.J., Geisler, J.H., 2003. Relationships of endemic African mammals and their fossil relatives based on morphological and molecular evidence. J. Mamm. Evol. 10, 131–194.

Bayzid, M.S., Warnow, T., 2013. Naive binning improves phylogenomic analyses. Bioinformatics 29, 2277–2284.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.

Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L., Purvis, A., 2007. The delayed rise of present-day mammals. Nature 446, 507–512.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., Roy-Choudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932.

Carstens, B.C., Dewey, T.A., 2010. Species delimitation using a combined coalescent and information theoretic approach: an example from North American *Myotis* bats. Syst. Biol. 59, 400–414.

Charlesworth, B., 2009. Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10, 195–205.

Charlesworth, B., Morgan, M.T., Charlesworth, D., 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134, 1289–1303.

Chiari, Y., Cahais, V., Galtier, N., Delsuc, F., 2012. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). BMC Biol. 10, 65.

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics, btu530.

D'Erchia, A.M., Gissi, C., Pesole, G., Saccone, C., Arnason, U., 1996. The guinea-pig is not a rodent. Nature 381, 597–600.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, 762–768.

de Queiroz, A., Gatesy, J., 2007. The supermatrix approach to systematics. Trends Ecol. Evol. 22, 34–41.

dos Reis, M., Inoue, J., Hasegawa, M., Asher, R.J., Donoghue, P.C.J., Yang, Z., 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc. R. Soc. B 279, 3491–3500.

dos Reis, M., Donoghue, P.C.J., Yang, Z., 2014. Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. Biol. Lett. 10, 20131003.

Doyle, J.J., 1997. Trees within trees: genes and species, molecules and morphology. Syst. Biol. 46, 537–553.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.

Gatesy, J., Springer, M.S., 2013. Concatenation versus coalescence versus "concatalescence". Proc. Natl. Acad. Sci. USA 110, E1179.

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

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

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

Hallström, B.M., Schneider, A., Zoller, S., Janke, A., 2011. A genomic approach to examine the complex evolution of laurasiatherian mammals. PLoS One 6, e28199.

Hartig, G., Churakov, G., Warren, W.C., Brosius, J., Makalowski, W., Schmitz, J., 2013. Retrophylogenomics place tarsiers on the evolutionary branch of anthropoids. Sci. Rep. 3, 1756.

Hobolth, A., Christensen, O.F., Mailund, T., Schierup, M.H., 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. PLoS Genet. 3, e7.

Hobolth, A., Dutheil, J.Y., Hawks, J., Schierup, M.H., Mailund, T., 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res. 21, 349–356.

Huang, H., Knowles, L.L., 2009. What is the danger of the anomaly zone for empirical phylogenetics? Syst. Biol. 58, 527–536.

Huang, H., He, Q., Kubatko, L.S., Knowles, L.L., 2010. Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. Syst. Biol. 59, 573–583.

Janečka, J.E., Miller, W., Pringle, T.H., Wiens, F., Zitzmann, A., Helgen, K.M., Springer, M.S., Murphy, W.J., 2007. Molecular and genomic data identify the closest living relative of primates. Science 318, 792–794.

Kriegs, J.O., Churakov, G., Keifmann, M., Jordan, U., Brosius, J., Schmitz, J., 2006. Retroposed elements as archives for the evolutionary history of placental mammals. PLoS Biol. 4, e91.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24.

Kumar, V., Hallström, B.M., Janke, A., 2013. Coalescent-based genome analyses resolve the early branches of the euarchontoglires. PLoS One 8, e60019.

Lanier, H.C., Knowles, L.L., 2012. Is recombination a problem for species-tree analyses? Syst. Biol. 61, 691–701.

Lanier, H.C., Knowles, L.L., 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.

Lanier, H.C., Huang, H., Knowles, L.L., 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Mol. Phylogenet. Evol. 70, 112–119.

Leaché, A.D., Rannala, B., 2011. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60, 126–137.

Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99–121.

Liang, D., Shen, X.X., Zhang, P., 2013. One thousand two hundred ninety nuclear genes from a genome-wide survey support lungfishes as the sister group of tetrapods. Mol. Biol. Evol. 30, 1803–1807.

Lin, Y.-H., McLenachan, P.A., Gore, A.R., Phillips, M.J., Ota, R., Hendy, M.D., Penny, D., 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol. Biol. Evol. 19, 2060–2070.

Liu, L., Edwards, S.V., 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58, 452–460.

Liu, L., Yu, L., Kubatko, L.S., Pearl, D.K., Edwards, S.V., 2009a. Coalescent methods for estimating phylogenetic trees. Mol. Phylogenet. Evol. 53, 320–328.

Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009b. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58, 468–477.

Liu, L., Yu, L., Edwards, S.V., 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10, 302.

Liu, L., Xi, Z., Davis, C.C., 2015. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. Mol. Biol. Evol. 32, 791–805.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, H.M., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409, 610–614.

McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements are novel phylogenomic markers that

resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 22, 746–754.

McVicker, G., Gordon, D., Davis, C., Green, P., 2009. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 5, e1000471.

Meredith, R.W., Janecka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L.L., Stadler, T., Rabosky, D.L., Honeycutt, R.L., Flynn, J.J., Ingram, C.M., Steiner, C., Williams, T.L., Robinson, T.J., Burk-Herrick, A., Westerman, M., Ayoub, N.A., Springer, M.S., Murphy, W.J., 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334, 521–524.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop, pp. 1–8.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014a. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, i541–i548.

Mirarab, S., Bayzid, M.S., Warnow, T., 2014b. Evaluating summary methods for multi-locus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol., syu063

Mirarab, S., Bayzid, M.S., Boussau, B., Warnow, T., 2014c. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science 346, 1250463-1.

Morgan, C.C., Foster, P.G., Webb, A.E., Pisani, D., McInerney, J.O., O'Connell, M.J., 2013. Heterogeneous models place the root of the placental mammal phylogeny. Mol. Biol. Evol. 30, 2145–2156.

Mukherjee, S., Sarkar-Roy, N., Wagener, D.K., Majumder, P.P., 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. Proc. Natl. Acad. Sci. USA 106, 7073–7078.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001a. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.

Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294, 2348–2351.

Murphy, W.J., Pringle, T.H., Crider, T.A., Springer, M.S., Miller, W., 2007. Using genomic data to unravel the root of the placental mammal phylogeny. Genome Res. 17, 413–421.

Nery, M.F., González, D.J., Hoffmann, F.G., Opazo, J.C., 2012. Resolution of the laurasiatherian phylogeny: evidence from genomic data. Mol. Phylogenet. Evol. 64, 685–689.

Nishihara, H., Hasegawa, M., Okada, N., 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. Proc. Natl. Acad. Sci. USA 103, 9929–9934.

Nishihara, H., Maruyama, S., Okada, N., 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. Proc. Natl. Acad. Sci. USA 106, 5235–5240.

O'Leary, M.A., Bloch, J.I., Flynn, J.J., Gaudin, T.J., Giallombardo, A., Giannini, N.P., Goldberg, S.L., Kraatz, B.P., Luo, Z.-X., Meng, J., Ni, X., Novacek, M.J., Perini, F.A., Randall, Z.S., Rougier, G.W., Sargis, E.J., Silcox, M.T., Simmons, N.B., Spaulding, M., Velazco, P.M., Weksler, M., Wible, J.R., Cirranello, A.L., 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. Science 339, 662–667.

Patel, S., Kimball, R.T., Braun, E.L., 2013. Error in phylogenetic estimation for bushes in the tree of life. J. Phylogenet. Evol. Biol. 1, 110.

Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A.M., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M.P.C., Silva, A., O'Brien, S.J., Pecon-Slattery, J., 2011. A molecular phylogeny of living primates. PLoS Genet. 7, e1001342.

Rannala, B., Yang, Z., 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164, 1645–1656.

Ranwez, V., Delsuc, F., Ranwez, S., Belkhir, K., Tilak, M., Douzery, E.J.P., 2007. OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics. BMC Evol. Biol. 7, 241.

Reyes, A., Gissi, C., Catzeflis, F., Nevo, E., Pesole, G., Saccone, C., 2004. Congruent mammalian trees from mitochondrial and nuclear genes using Bayesian methods. Mol. Biol. Evol. 21, 397–403.

Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53, 131–147.

Roch, S., Warnow, T., 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Syst. Biol., syv016

Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N., Douzery, E.J.P., 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. Mol. Biol. Evol. 30, 2124–2144.

RoyChoudhury, A., Felsenstein, J., Thompson, E.A., 2008. A two-stage pruning algorithm for likelihood computation for a population tree. Genetics 180, 1095–1105.

Scally, M., Madsen, O., Douady, C.J., de Jong, W.W., Stanhope, M.J., Springer, M.S., 2001. Molecular evidence for the major clades of placental mammals. J. Mamm. Evol. 8, 239–277.

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., et al., 2012. Insights into hominoid evolution from the gorilla genome sequence. Nature 483, 169–175.

Shaw, T.I., Srivastava, A., Chou, W.-C., Liu, L., Hawkinson, A., Glenn, T.C., Adams, R., Schountz, T., 2012. Transcriptome sequencing and annotation for the Jamaican fruit bat (*Artibeus jamaicensis*). PLoS One 7, e48472.

Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: species TRee analysis web server. Nucleic Acids Res. 41 (Web Server issue), W238–W241.

Shoshani, J., McKenna, M.C., 1998. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. Mol. Phylogenet. Evol. 9, 572–584.

Simmons, M.P., Gatesy, J., 2015. Coalescence vs. concatenation: sophisticated analyses vs. first principles applied to rooting the angiosperms. Mol. Phylogenet. Evol. 91, 98–122.

Song, S., Liu, L., Edwards, S.V., Wu, S., 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc. Natl. Acad. Sci. USA 109, 14942–14947.

Springer, M.S., Gatesy, J., 2014. Land plant origins and coalescence confusion. Trends Plant Sci. 19, 267–269.

Springer, M.S., Murphy, W.J., 2007. Mammalian evolution and biomedicine: new views from phylogeny. Biol. Rev. 82, 375–392.

Springer, M.S., Murphy, W.J., Eizirik, E., O'Brien, S.J., 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc. Natl. Acad. Sci. USA 100, 1056–1061.

Springer, M.S., Stanhope, M.J., Madsen, O., de Jong, W.W., 2004. Molecules consolidate the placental mammal tree. Trends Ecol. Evol. 19, 430–438.

Springer, M.S., Murphy, W.J., Eizirik, E., O'Brien, S.J., 2005. Evidence for major placental clades. In: Rose, K.D., Archibald, J.D. (Eds.), The Rise of Placental Mammals: Origins and Relationships of the Major Extant Clades. Johns Hopkins University Press, Baltimore, pp. 37–49.

Springer, M.S., Murphy, W.J., Eizirik, E., Madsen, O., Scally, M., Douady, C.J., Teeling, E.C., Stanhope, M.J., de Jong, W.W., O'Brien, S.J., 2007. A molecular classification for the living orders of placental mammals and the phylogenetic placement of primates. In: Ravosa, M.J., Dagosto, M. (Eds.), Primate Origins: Adaptations and Evolution. Springer, New York, pp. 1–28.

Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J., Rabosky, D.L., Stadler, T., Steiner, C., Ryder, O.A., Janečka, J.E., Fisher, C.A., Murphy, W.J., 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PLoS One 7, e49521.

Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571.

Sul, S.-J., Williams, T.L., 2008. An experimental analysis of Robinson–Foulds distance matrix algorithms. In: Halperin, D., Mehlhorn, K. (Eds.), Algorithms-ESA. Springer, Berlin, Heidelberg, pp. 793–804.

Sullivan, J., Swofford, D.L., 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mamm. Evol. 4, 77–86.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (∗and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Tsagkogeorga, G., Parker, J., Stupka, E., Cotton, J.A., Rossiter, S.J., 2013. Phylogenomic analyses elucidate the evolutionary relationships of bats. Curr. Biol. 23, 2262–2267.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P., Graham, S.W., Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw, A.J., DeGironimo, L., Stevenson, D.W., Surek, B., Villarreal, J.C., Roure, B., Philippe, H., dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G.K.-S., Leebens-Mack, J., 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA 111, E4859–E4868.

Wildman, D.E., Uddin, M., Opazo, J.C., Liu, G., Lefort, V., Guindon, S., Gascuel, O., Grossman, L.I., Romero, R., Goodman, M., 2007. Genomics, biogeography, and the diversification of placental mammals. Proc. Natl. Acad. Sci. USA 104, 14395–14400.

Xi, Z., Rest, J.S., Davis, C.C., 2013. Phylogenomics and coalescent analyses resolve extant seed plant relationships. PLoS One 8, e80870.

Xi, Z., Liu, L., Rest, J.S., Davis, C.C., 2014. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. Syst. Biol. 63, 919–932.

Xu, L., Chen, S.-Y., Nie, W.-H., Jiang, X.-L., Yao, Y.-G., 2012. Evaluating the phylogenetic position of Chinese tree shrew (*Tupaia belangeri chinensis*) based on complete mitochondrial genome: Implication for using tree shrew as an alternative experimental animal to Primates in biomedical research. J. Genet. Genom. 39, 131–137.

Zhong, B., Liu, L., Yan, Z., Penny, D., 2013. Origin of land plants using the multispecies coalescent model. Trends Plant Sci. 18, 492–495.

Zimmermann, T., Mirarab, S., Warnow, T., 2014. BBCA: improving the scalability of *BEAST using random binning. BMC Genom. 15, S11.

Glossary

Anomaly zone: a set of conditions wherein the most probable gene tree is different from the species tree

Branch length heterogeneity: a type of gene tree heterogeneity that results from recombination and leads to differences in branch lengths (in CUs) but not topology

Coalescence debt: the sum of all deep coalescences, measured in years, that are required to fit a gene tree to a species tree

Coalescence gene (c-gene): a segment of the genome for which there has been no recombination over the phylogenetic history of a clade

Concatalescence: concatenation of two or more c-genes followed by application of coalescence methods to reconstruct a species tree

Deep coalescence: "...the failure of ancestral copies to coalesce (looking backwards in time) into a common ancestral copy until deeper than previous speciation events" (Maddison, 1997; p. 523)

Maximum deep coalescence: the maximum discrepancy (in years) between the age of a crown clade on the species tree and the age of the least inclusive clade on the gene tree that includes all of the same taxa

Gene tree stoichiometry: the quantitative relationship between gene trees and the species tree

Recombination ratchet: the inverse relationship between c-gene size and the number of taxa that is driven by recombination in different regions of a phylogenetic tree

Robinson–Foulds distance: a measure of the distance between two phylogenetic trees that is equal to the sum of partitions on the first tree but not the second tree plus the number of partitions on the second tree but not the first tree

Shortcut coalescence method: a coalescence method that does not co-estimate gene trees with the species tree and instead estimates a species tree using summary statistics derived from the estimated gene trees