Molecular Phylogenetics and Evolution xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum

John Gatesy *, Mark S. Springer

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

ARTICLE INFO

Article history:
Received 10 May 2014
Revised 26 July 2014
Accepted 10 August 2014
Available online xxxx

Keywords:
Mammalia
Supermatrix
Species tree
Gene tree
Deep coalescence

ABSTRACT

Large datasets that include many taxa and many genes are required to solve difficult phylogenetic problems that are deep in the Tree of Life. Currently, two divergent systematic methods are applicable and commonly applied to such datasets: the traditional supermatrix approach (= concatenation) and "shortcut" coalescence methods (= coalescence methods wherein gene trees and the species tree are not co-estimated). When applied to ancient clades, these contrasting frameworks often produce congruent results, but in recent phylogenetic analyses of Placentalia (placental mammals), this is not the case, with proponents of coalescence arguing that their novel approach robustly supports several interordinal clades that have resisted consistent resolution using standard concatenation methods. A recent series of papers has alternatively disputed and defended the utility of shortcut coalescence methods at deep phylogenetic scales (Meredith et al., 2011; Song et al., 2012; Gatesy and Springer, 2013; Wu et al., 2013; Zhong et al., 2013, 2014; Springer and Gatesy, 2014). Here, we examine this exchange in the context of published phylogenomic data from Mammalia (183 loci; McCormack et al., 2012); in particular we explore two critical and related issues – the delimitation of data partitions ("genes") in coalescence analysis and hidden support that emerges with the combination of such partitions in phylogenetic studies. Hidden support – increased support for a clade in combined analysis of all data partitions relative to the support evident in separate analyses of the various data partitions, is a hallmark of the supermatrix approach and a primary rationale for concatenating all characters into a single matrix. In the most extreme cases of hidden support, relationships that are contradicted by all gene trees are supported when all of the genes are analyzed together. A valid fear is that shortcut coalescence methods might bypass or distort character support that is hidden in individual loci because small gene fragments are analyzed in isolation. Given the extensive database of molecular, phenotypic, and fossil data for Mammalia, the assumptions and applicability of shortcut coalescence methods can be assessed with rigor to complement a small but growing body of simulation work that has directly compared these methods to concatenation. We document several remarkable cases of hidden support in both supermatrix and coalescence paradigms and argue that in most instances, the emergent support in the shortcut coalescence analyses is an artifact. By referencing rigorous molecular clock studies of Mammalia, we suggest that inaccurate gene trees that imply unrealistically deep coalescences debilitate shortcut coalescence analyses of the placental dataset. We document contradictory coalescence results for Placentalia, and outline a critical conundrum that challenges the general utility of shortcut coalescence methods at deep phylogenetic scales. In particular, the basic unit of analysis in coalescence analysis, the coalescence-gene (a minimal non-recombining stretch of aligned genomes) is expected to shrink in size as more taxa are analyzed, but as the amount of data for reconstruction of a gene tree ratchets downward, the number of nodes in the gene tree that need to be resolved ratchets upward. One possible solution to this inevitable equation is to concatenate multiple coalescence-genes to yield "gene trees" that better match the species tree. However, this hybrid concatenation/coalescence approach, "concatalescence," contradicts the most basic biological rationale for performing a coalescence analysis in the first place. We discuss this reality in the context of recent

Abbreviations: AGT, anomalous gene tree; bp, basepairs; c-gene, coalescence gene; CUs, coalescent units; MRP, matrix representation with parsimony; ML, maximum likelihood; MP-EST, maximum pseudo-likelihood for estimating species trees; MY, million years; MDC, minimization of deep coalescences; STEAC, species tree estimation using average coalescence times; STAR, species tree estimation using average ranks of coalescences; STEM, species tree estimation using maximum likelihood; UCEs, ultra-conserved elements.

* Corresponding author. Fax: +1 951 827 2421.
E-mail address: john.gatesy@ucr.edu (J. Gatesy).

http://dx.doi.org/10.1016/j.ympev.2014.08.013

simulation work that also suggests inaccurate reconstruction of gene trees is more problematic for shortcut coalescence methods than deep coalescence of independently segregating loci is for concatenation methods.

1. Introduction

A revolution in systematic methods appears to be imminent; implementation of coalescence theory, a central concept in population genetics (Hudson, 1990), is impacting systematic biology at all levels of analysis (Degnan and Rosenberg, 2009; Edwards, 2009; Liu et al., 2009a, 2010; Kubatko, 2009; Knowles and Kubatko, 2010; Zhong et al., 2013). However, with the emergence of any new scientific paradigm, critical analysis and a fair-minded, broad comparison to incumbent methodologies should precede abandonment of the status quo. In the context of resolving ancient phylogenetic problems in the history of Life, multiple authors recently have argued that shortcut coalescence methods that do not co-estimate gene trees with the species tree – STAR (Liu et al., 2009b), STEAC (Liu et al., 2009b), MP-EST (Liu et al., 2010), STEM (Kubatko et al., 2009), minimization of deep coalescences (MDC; Maddison, 1997) – are superior to the more traditional supermatrix (concatenation) approach to phylogenetic analysis (de Queiroz and Gatesy, 2007) because gene tree heterogeneity is accounted for by the former (Edwards, 2009; Liu et al., 2009a; Knowles and Kubatko, 2010). Here, we examine this viewpoint in the context of two critical and related aspects of systematics at deep phylogenetic scales: (1) the delimitation of data partitions and (2) hidden support that emerges with the combination of such partitions in phylogenetic analysis (Barrett et al., 1991; Chippindale and Wiens, 1994; Gatesy et al., 1999).

This comparison between concatenation and shortcut coalescence is critical, because better-justified coalescence methods that co-estimate gene trees and species trees such as BEST (Liu and Pearl, 2007; Liu, 2008) and *BEAST (Heled and Drummond, 2010) are irrelevant for reconstructing large sectors of the Tree of Life due to daunting computational constraints. Solid sampling of both genes and taxa surely will be required to resolve intransigent phylogenetic problems at deep timescales (e.g., Zhong et al., 2013), but for even moderately sized datasets, co-estimation methods cannot be applied because many free parameters must be optimized simultaneously (Bayzid and Warnow, 2013). To speed computation, shortcut coalescence methods instead entail isolated estimation of gene trees, which are assumed to be accurate based on limited evidence (Kubatko et al., 2009; Liu et al., 2009b, 2010). Nearly all, recent, multigene coalescence studies at deep phylogenetic scales that have sampled >20 taxa and >150 loci applied shortcut methods (e.g., McCormack et al., 2012; Song et al., 2012; Xi et al., 2013; Zhong et al., 2013) as have studies with >100 taxa and >20 loci (e.g., Meredith et al., 2011; Song et al., 2011).

Hidden support can be defined as increased support for a clade in combined analysis of all data partitions relative to the sum of support for that clade indicated in separate analyses of the individual data partitions (Gatesy et al., 1999). In the most extreme cases, multiple data partitions, each of which contradicts a clade in separate analysis, favor that clade in combination (Fig. 1A; Gatesy and Baker, 2005). Supermatrix analyses have repeatedly shown extensive hidden support with the concatenation of small data partitions in both parsimony and maximum likelihood (ML) contexts (Gatesy et al., 1999; Gatesy and Arctander, 2000; Gatesy et al., 2003, 2004; Lee and Hugall, 2003; Lambkin, 2004; Gatesy and Baker, 2005; Wahlberg et al., 2005). This synergistic support is a hallmark of the supermatrix method (de Queiroz and Gatesy, 2007).

The ineffectiveness of phylogenetic reconstruction based on small sets of characters was a primary impetus for the development of the supermatrix approach to systematics over 20 years ago (Miyamoto, 1985; Kluge, 1989; de Queiroz et al., 1995; Nixon and Carpenter, 1996). In this framework, all characters are concatenated into a single matrix, and the data are analyzed simultaneously to decrease sampling error and to offset quirky homoplasy that is restricted to particular data partitions (Fig. 1A). The primary alternative at that time, the taxonomic congruence approach (Fig. 1B; Nelson, 1979; Miyamoto and Fitch, 1995), stressed the independent corroboration provided by trees supported by separate analyses of individual data partitions, but taxonomic congruence was criticized and fell out of favor when it was realized how much common support was ignored by separating, rather than combining, data (e.g., Nixon and Carpenter, 1996; Gatesy et al., 1999; Gatesy and Baker, 2005). Hidden character support cannot emerge in the taxonomic congruence approach, which simply summarizes the clades supported by isolated analyses of individual data partitions (Fig. 1B).

An analogous critique subsequently was leveled against the supertree approach (Fig. 1C) in which various subsets of characters are analyzed separately to produce subtrees, which are then used to reconstruct phylogeny (Springer and de Jong, 2001; Gatesy et al., 2002; Gatesy and Springer, 2004; Janies et al., 2013). As in taxonomic congruence, the detailed character information within datasets is lost when sets of characters are summarized as trees, and hidden character support again is ignored or distorted (Gatesy et al., 2004). In certain supertree methods, novel relationships can emerge that conflict with all input trees (e.g., matrix representation with parsimony – MRP), but this effect is generally interpreted as a defect, not an advantage, of these methods (Rodrigo, 1993; Goloboff and Pol, 2002; Pisani and Wilkinson, 2002; Gatesy et al., 2004).

A valid fear is that with the advent of shortcut coalescence methods (Fig. 1D), the field of systematics has come back full circle to trust in the analysis of small data partitions, "genes," that are compromised by sampling error (Meredith et al., 2011; Townsend et al., 2011; Chiari et al., 2012; Rosenfeld et al., 2012; Bayzid and Warnow, 2013; Gatesy and Springer, 2013; McCormack et al., 2013; Patel et al., 2013; Rokas et al., 2013). It is unknown whether the new coalescence methods are able to successfully quantify emergent support for deep phylogenetic relationships from separate analyses of individual loci and whether resolution of conflicts among inaccurately reconstructed gene trees can yield coherent systematic results (Meredith et al., 2011; Gatesy and Springer, 2013; Springer and Gatesy, 2014). These concerns are justified, because coalescence methods that presently can be applied to large systematic datasets do not condition on a common species tree (Maddison, 1997; Kubatko et al., 2009; Liu et al., 2009a, 2009b, 2010), generally presume that *all* gene tree incongruence is due to deep coalescence, and are predicated on the further assumption that gene trees are reconstructed accurately in isolation. When this last assumption does not hold, statistical consistency – convergence on the correct species tree with the

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

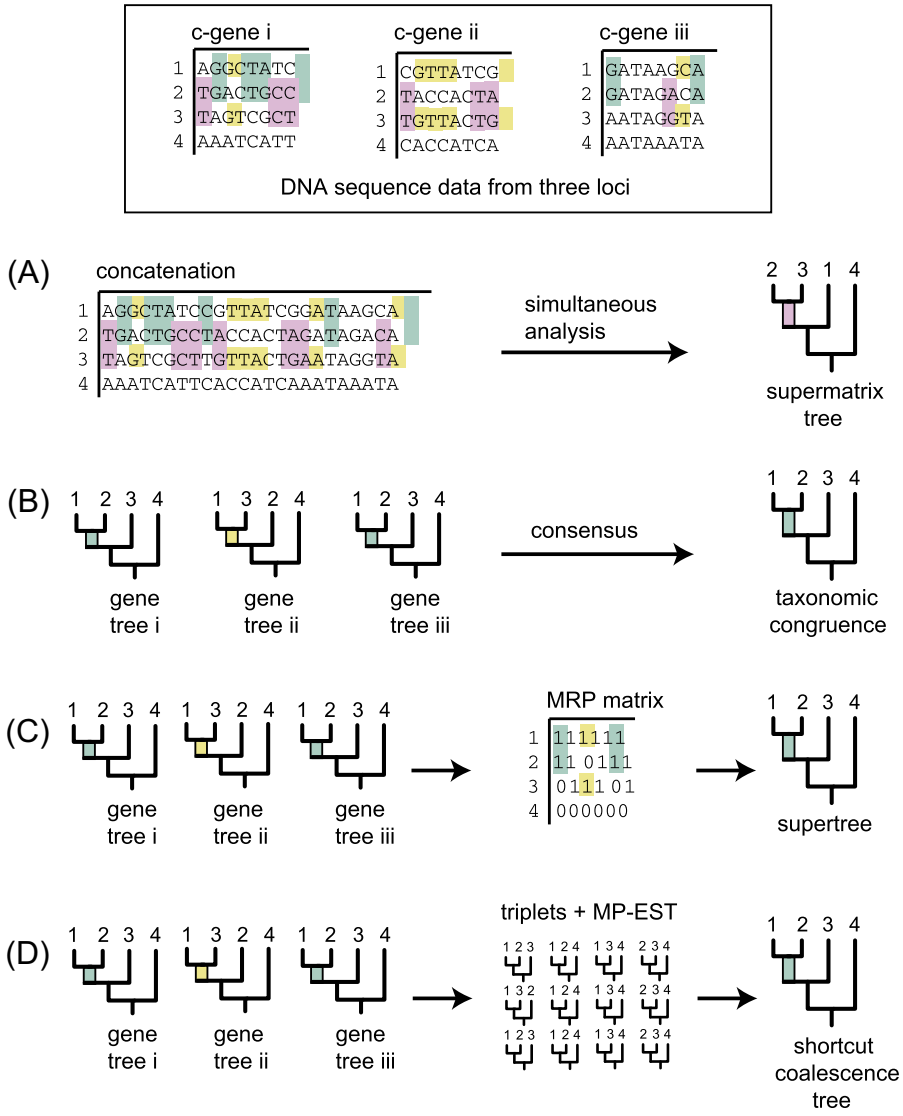

Fig. 1. Three c-genes (i–iii) from four taxa (1–4) and phylogenetic results for four systematic approaches: supermatrix analysis (A), taxonomic congruence (B), matrix representation with parsimony (MRP) supertree (C), and MP-EST shortcut coalescence (D). Taxon 4 is specifed as the outgroup. Optimal species trees are shown for each method. Note that in this contrived scenario, hidden character support that emerges with concatenation (A) is bypassed by methods B–D that first reconstruct subtrees from small data partitions, and then derive a species tree based on subtrees rather than through direct analysis of characters. For simplicity, phylogenetic analyses of character data (supermatrix plus individual c-genes) and the MRP supertree matrix are based on equally weighted parsimony. STAR shortcut coalescence analysis of the data supports the same tree as MP-EST (D). Parsimony informative characters that support 1 + 2 (green), 1 + 3 (yellow), and 2 + 3 (pink) are highlighted, and the same colors mark these clades in gene trees and species trees. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

accumulation of more data – is not guaranteed (Liu et al., 2009b, 2010; DeGiorgio and Degnan, 2014).

A directly related concern is that support for particular clades might be underestimated by coalescence methods relative to concatenation due to division of the overall dataset into small, misleading data partitions (Meredith et al., 2011; Townsend et al., 2011; Gatesy and Springer, 2013; McCormack et al., 2013; Springer and Gatesy, 2014). Proponents of coalescence repeatedly have stressed that supermatrix analysis can overestimate nodal support due to mis-specification of evolutionary models (Edwards, 2009; Song et al., 2012; Wu et al., 2013), a valid concern (Kubatko and Degnan, 2007). It is also possible, however, that the lower support values commonly reported in coalescence studies (Townsend et al., 2011; McCormack et al., 2012, 2013; Song et al., 2012) are due to reliance on small data partitions that are prone to reconstruction errors (Meredith et al., 2011; Gatesy and Springer, 2013; Lanier et al., 2014). Furthermore, the delimitation of data partitions in the coalescence approach is controversial and challenging (Ané, 2010; Gatesy and Springer, 2013; Springer and Gatesy, 2014). Longer stretches of DNA might support robust topologies (Song et al., 2013) but also likely span regions of the genome that have experienced numerous recombination events (Hobolth et al., 2007, 2011), contradicting the primary logic of employing coalescence methods in the first place (Bayzid and Warnow, 2013; Gatesy and Springer, 2013; Springer and Gatesy, 2014). When subsequent internodes are very short, coalescence methods can prefer a minority gene tree in the so-called "anomaly zone" (Kubatko and Degnan, 2007; Liu and Edwards, 2009; Rosenberg, 2013), but there is no precedent in the literature for coalescence methods favoring a clade that is contradicted by all input gene trees, as can occur in supermatrix analysis (Fig. 1A).

For deep phylogenetic splits, the relative utility of a coalescence approach to systematics is debatable (Edwards et al., 2007; Meredith et al., 2011; Song et al., 2012; Gatesy and Springer,

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

2013; Oliver et al., 2013; Wu et al., 2013; Zhong et al., 2013, 2014; Springer and Gatesy, 2014). When allelic mutants in terminal taxa arose *subsequent* to all divergences among species in a phylogenetic tree, this intraspecific variation is expected to provide limited information for reconstructing ancient phylogenetic relationships, offsetting a primary advantage of the coalescence paradigm (Maddison and Knowles, 2006; Patel et al., 2013). Coalescence methods could still be advantageous for resolving deep phylogeny, however, because ancient speciation events can be tightly spaced in time relative to each other, and random sorting of ancestral polymorphism might still account for much of the character conflict at such nodes (Fig. 2), so-called "mixed-signal homoplasy" (Doyle, 1997) or "hemiplasy" (Avise and Robinson, 2008). Coalescence methods are designed to handle these conflicts that are due to retention of ancestral polymorphism and subsequent sorting of this intraspecific variation (Fig. 2; Edwards, 2009).

Several published coalescence analyses of higher-level phylogenetic relationships are largely congruent with supermatrix results (e.g., Edwards et al., 2007; Oaks, 2011; Chiari et al., 2012). However, a more meaningful comparison between shortcut coalescence versus supermatrix analysis of ancient clades must instead focus on taxa that show disparate evolutionary reconstructions when these competing methods are applied (e.g., Townsend et al., 2011; Kimball et al., 2013; Zhong et al., 2013). In this context, perhaps the most high profile discrepancy between coalescence and concatenation to date has centered on interrelationships among orders of placental mammals – a spirited, but unresolved, debate that has transpired in *Science* (Meredith et al., 2011), *Proceedings of the National Academy of Sciences USA* (Song et al., 2012; Gatesy and Springer, 2013; Wu et al., 2013), and *Genome Research* (McCormack et al., 2012). Fortunately, Mammalia is characterized by a huge database of systematic information relative to most other groups and includes anatomical characters, fossils, transposon insertion data, as well as multiple nuclear and mitochondrial genomes (e.g., Nishihara et al., 2006, 2009; Hallström et al., 2011; Lindblad-Toh et al., 2011; Hassanin et al., 2012; Gatesy et al., 2013; Kumar et al., 2013; O'Leary et al., 2013). The rich catalog of empirical data from mammals permits a rigorous discrimination between the competing viewpoints of the concatenation (Meredith et al., 2011; Gatesy and Springer, 2013) and coalescence (Song et al., 2012; McCormack et al., 2012; Wu et al., 2013) camps, and complements conclusions drawn from coalescence simulations (Kubatko et al., 2009; Liu et al., 2009b, 2010; Leaché and Rannala, 2011; Bayzid and Warnow, 2013; Patel et al., 2013), in which conditions are highly controlled but likely do not express the complexities of real genetic data.

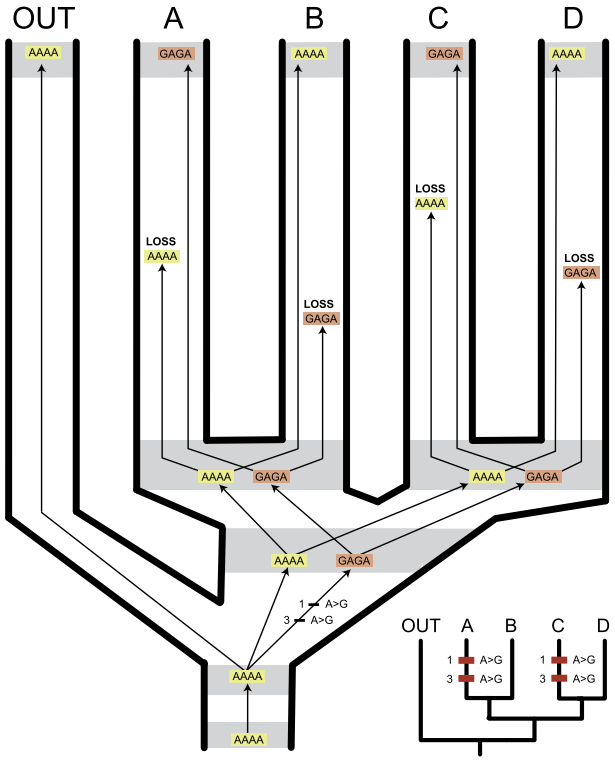

Fig. 2. A hypothetical scenario that shows mixed signal homoplasy ("hemiplasy") due to retention of ancestral polymorphism across multiple internal nodes and convergent sorting of alleles in different lineages. The true evolutionary history is shown in the large tree with four tightly spaced speciation events. A single allele is ancestral. Mutations at nucleotide positions 1 and 3 near the base of the tree yield a second allele. Alternative alleles are then fixed in each of the four terminal ingroup lineages (A–D). The small tree (lower right) shows a parsimony reconstruction of character state changes for the tree shown; note the inference of homoplasy at two sites that is due to the convergent, correlated fixation of mutations in alternative alleles, and not due to "standard" homoplasy (convergent mutation + fixation of the same state or a mutation + fixation that results in reversal to the primitive state). Note that substitution matrices commonly utilized by systematists model mutation + fixation and not just mutation (Patel et al., 2013). For example, different substitution models and branch lengths are commonly applied to different codon positions of protein-coding sequences (Ren et al., 2005). This is not because rates of mutation are thought to be very different at the three codon positions, but instead because rates of fixation are unequal due to different selective constraints.

To our knowledge, no study has documented emergent support in recently developed procedures that infer phylogenetic relationships based on the relative depths of coalescences in gene trees. Here, we describe several remarkable cases of hidden support in coalescence and supermatrix analyses of the same phylogenomic dataset for Mammalia (McCormack et al., 2012). For both methods, we record examples where >100 loci each contradict a particular clade when analyzed separately, but support that clade when the >100 conflicting loci are analyzed in combination. By utilizing the diverse systematic database compiled for Mammalia, we suggest that the surprising, synapomorphy-free, hidden support detected by coalescence approaches in a few instances represents a methodological artifact. More commonly, the consistent hidden support that emerged in concatenation is bypassed by shortcut coalescence approaches. Examination of the DNA sequence data from McCormack et al. (2012) supports the hypothesis that uninformative data partitions, gross incongruence among inferred gene trees, and inexplicably deep coalescences mask emergent support – a pattern corroborated by recent empirical work (Meredith et al., 2011; Kimball et al., 2013; McCormack et al., 2013) and simulations that are not flattering to shortcut coalescence (Leaché and Rannala, 2011; Bayzid and Warnow, 2013; Patel et al., 2013). We discuss the general implications of these results and outline the current limitations of a coalescence approach for reconstructing ancient divergences in the Tree of Life using large datasets, focusing in particular on a critical conundrum regarding delimitation of "genes," the basic units of analysis in this systematic paradigm.

2. Materials and methods

This study focuses on a reanalysis of the 183-locus dataset for placental mammals compiled by McCormack et al. (2012). McCormack et al. (2012) utilized short sequences (~410 basepairs [bp]) that flank ultra-conserved elements (UCEs) in the genome (24 mammalian taxa, 183 loci, 74,750 bp) and concluded that coalescence analysis was necessary for the resolution of several long-standing controversies in the deep phylogeny of placental mammals (Fig. 3). By contrast, a recent supermatrix analysis of Mammalia at the family level (164 mammalian taxa, 26 loci, 35,603 bp) suggested that the polytomies resolved by McCormack et al. (2012) could not be distinguished based on current phylogenetic methods and datasets (Meredith et al., 2011). Analyses of

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

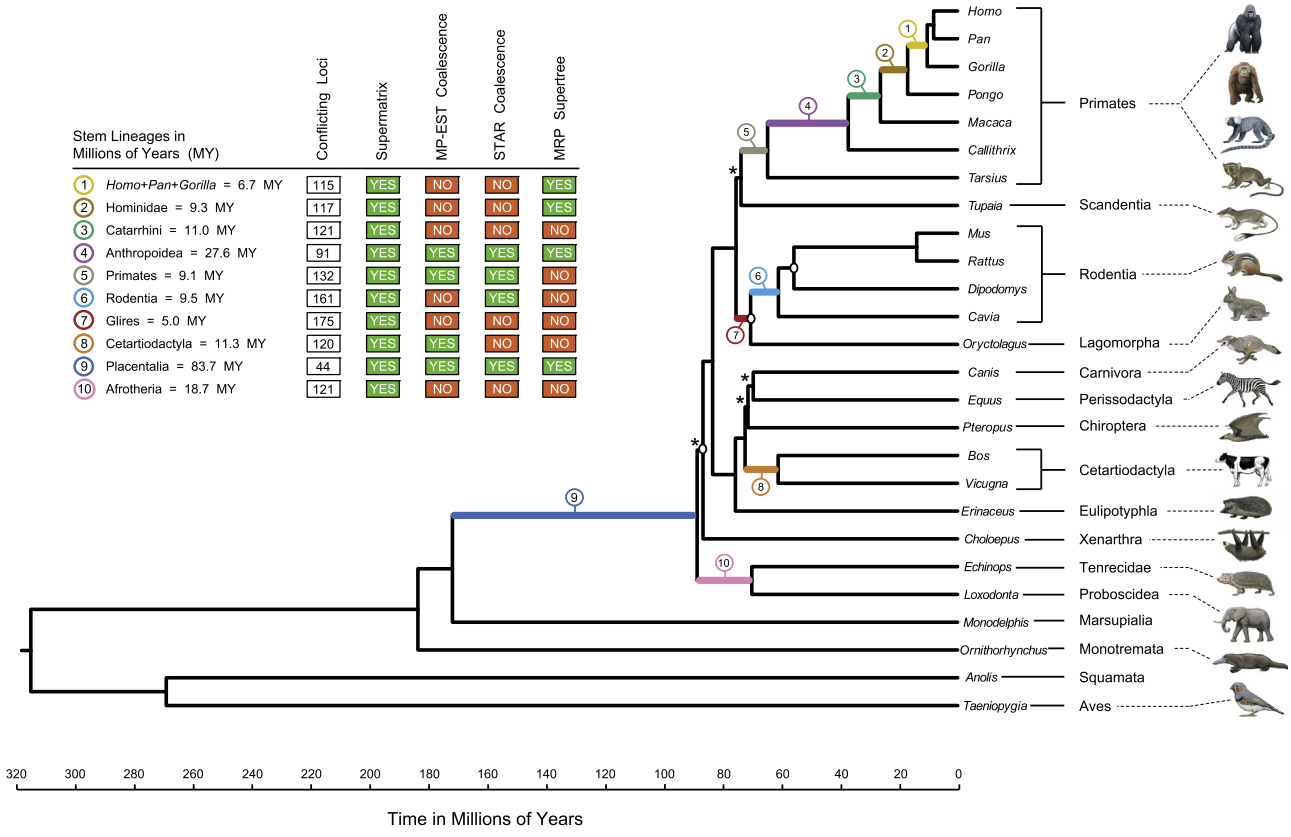

Fig. 3. The mammalian timetree used in this paper for measuring deep coalescences, and hidden support results for ten uncontroversial clades in this tree. Topology is the tree supported by STAR coalescence analysis of 183 UCE loci in McCormack et al. (2012), and divergence dates are from dos Reis et al. (2012). Four traditionally difficult to resolve nodes are marked by asterisks (*); collapse of these four nodes results in three polytomies. The three nodes in the STAR shortcut coalescence tree that are contradicted by MP-EST shortcut coalescence analysis are marked by white ovals. Hidden support was assessed for ten test clades; these groups are numbered and the duration of the stem lineage to each clade is shown. In separate analyses of individual loci, numerous gene trees contradict a given clade; the number of conflicting loci, out of 183, is listed for each clade. "YES" indicates that combined analysis of these conflicting loci supports that clade. "NO" indicates that combined analysis of these conflicting loci does not support that clade. Results for supermatrix, MP-EST coalescence, STAR coalescence, and MRP supertree analyses are shown. Paintings are by C. Buell.

transposon insertions support this second interpretation and record extensive evidence for mis-sorting of ancestral polymorphism at these polytomies with nearly equivalent support for competing hypotheses (Nishihara et al., 2006, 2009).

To assess contrasting patterns of emergent support, we focus on ten uncontroversial clades that were supported robustly by both coalescence and supermatrix analyses of this dataset (McCormack et al., 2012) and have long stem lineages according to multiple molecular clock studies (Fig. 3; Bininda-Emonds et al., 2007; Arnason et al., 2008; Meredith et al., 2011; dos Reis et al., 2012, 2014). These clades include: Glires (~5 MY stem), Cetartiodactyla (~11 MY stem), Afrotheria (~19 MY stem), Rodentia (~10 MY stem), Anthropoidea (~27 MY stem), Placentalia (~84 MY stem), Primates (~9 MY stem), Catarrhini (~11 MY stem), Hominidae (~9 MY stem), and *Pan* + *Homo* + *Gorilla* (~7 MY stem) (Fig. 3). In its most basic form, hidden support for a clade is apparent when individual loci that *do not* support a clade in separate analyses *do* support that clade when these same contradictory loci are analyzed in combination (Fig. 1A). We therefore re-analyzed all loci that conflict with a particular clade using both shortcut coalescence methods (STAR and MP-EST) and ML concatenation to document extreme emergent support for the ten clades of interest.

ML gene trees for the 183 UCE loci were as in McCormack et al. (2012). ML gene trees that strictly conflict with a given clade were utilized to build species trees using the STAR and MP-EST shortcut coalescence methods, both of which have been employed in recent phylogenetic analyses of mammalian genomes (McCormack et al., 2012; Song et al., 2012; Kumar et al., 2013). STAR was implemented at the STRAW website (Shaw et al., 2013; http://bioinformatics.publichealth.uga.edu/SpeciesTreeAnalysis/), and MP-EST (version 1.2) was executed on a desktop computer with more thorough search settings than employed in the default version. Specifically, 'MAXROUND' was changed from 1,000,000 to 20,000,000 and 'NUM_NOCHANGE' was changed from 10,000 to 1,000,000. ML supermatrix analyses of loci that uniformly contradicted a particular clade in isolated ML analyses (McCormack et al., 2012) were executed with RAxML (Stamatakis, 2006) and Garli 2.01 (Zwickl, 2006) on CIPRES (Miller et al., 2010). Analyses with both programs were performed with a single GTR + Γ model for each complete, non-partitioned dataset. Bootstrapping of concatenated data in RAxML employed 500 replications with rapid bootstrapping. We employed the GTRGAMMA option for both bootstrapping and inference of the final ML tree. Analyses with Garli employed 100 bootstrap replications (maximum allowable number on CIPRES), each of which was performed with stepwise addition (attachmentspertaxon = 59) and subtree pruning regrafting (spr) branch swapping (limsprrange = 8). For comparison to shortcut coalescence and supermatrix results, MRP supertree analyses (Fig. 1C) were executed in PAUP* (Swofford, 2002) using the ML gene trees from McCormack et al. (2012). Searches were heuristic with 100 random taxon addition replicates and tree bisection reconnection (tbr) branch swapping to assess emergent support in supertree analysis.

Recent relaxed clock studies have obtained similar divergence time estimates for both interordinal and intraordinal divergences that are included in McCormack et al.'s (2012) topology for 24 mammalian species (Meredith et al., 2011; dos Reis et al., 2012; Springer et al., 2012). Meredith et al. (2011) and dos Reis et al.

(2012) employed calibrations at 82 and 27 nodes, respectively, and have dense calibration coverage relative to other timetree analyses for mammalian interordinal relationships. We employed dates from dos Reis et al. (2012) to calculate deep coalescence times that are implied by McCormack et al.'s (2012) gene trees given complete taxonomic overlap and topological congruence with McCormack et al.'s (2012) species tree. We assumed a most recent common ancestry of 200,000 years for *Homo sapiens* (Underhill and Kivisild, 2007).

3. Results and discussion

3.1. Remarkable and consistent hidden support in supermatrix analyses of placental mammals

For the supermatrix approach, when loci that strictly conflict with a given clade were concatenated and analyzed in combination, solid hidden support emerged for all ten clades that were examined (Fig. 3). For example, Glires is supported by both STAR coalescence (bootstrap = 66%) and ML supermatrix analysis (bootstrap = 95%) with the full 183-locus data set. However, Glires is not supported by 175/183 (96%) loci (McCormack et al., 2012). When the 175 loci that do not individually support Glires are concatenated and analyzed using the ML supermatrix approach, however, the clade is supported with a RAxML bootstrap score of 81% (Fig. 4; 79% for Garli). This is perhaps the most extreme example of hidden support ever documented; 175 "wrongs" in separate analysis add up to an apparent "right" in supermatrix analysis. Remarkably, Glires (175 conflicting loci) as well as Cetartiodactyla (120 conflicting loci; bootstrap = 99% RAxML; 100% Garli), Afrotheria (121 conflicting loci; bootstrap = 100%; 100%), Rodentia (161 conflicting loci; bootstrap = 95%; 96%), Primates (132 conflicting loci; bootstrap = 100%; 98%), Anthropoidea (91 conflicting loci; bootstrap = 100%; 100%), Catarrhini (121 conflicting loci; bootstrap = 78%; 80%), Hominidae (117 conflicting loci; bootstrap = 71%; 75%), *Pan* + *Homo* + *Gorilla* (115 conflicting loci; bootstrap = 93%; 92%), and Placentalia (44 conflicting loci; bootstrap = 100%; 100%) were recovered with solid support by supermatrix analyses of all conflicting loci for a given clade.

Given that all ten examples of hidden support (Fig. 3) favor widely accepted clades (e.g., Meredith et al., 2011; O'Leary et al., 2013; Song et al., 2012), it is difficult to argue that hidden support is a mere artifact (Fig. 4). Indeed, both coalescence and concatenated analyses of McCormack et al.'s (2012) complete dataset validate the monophyly of all ten clades. The simplest interpretation of the overwhelming and consistent hidden support in these supermatrix analyses is that sampling error and discrepant patterns of homoplasy in the separate analyses of short loci (~410 bp) are overcome by concatenation into a much larger supermatrix. Individual synapomorphies that support Glires monophyly presumably were masked by homoplasy in separate analyses of the 175 small data partitions, but the strong common character support across loci emerged in supermatrix analysis of a much larger set of characters (Figs. 3 and 4).

3.2. Surprising, but inconsistent, hidden support in shortcut coalescence analyses of placental mammals

Coalescence methods showed a much more scattered and incoherent pattern of hidden support that is difficult to interpret (Fig. 3). For example, when the 175 "wrong" gene trees that do not support Glires monophyly are analyzed using STAR or MP-EST, Glires is not supported (Fig. 4). From our perspective this is the expected result – hidden character support that emerges via character congruence in supermatrix analyses should not emerge with coalescence methods that instead process subtrees (Gatesy et al., 1999, 2004). Coalescence methods that interpret all conflicts among gene trees as due to deep coalescence might not be expected to yield a coherent result when many of the conflicts among gene trees are likely due to other factors (sampling error, long branch misplacement, uneven homoplasy, model mis-specification, poor taxonomic sampling, arbitrary resolution). When detailed character information is hijacked from final analysis – cloaked by the cruder summaries provided by gene trees – hidden character support might remain hidden (Gatesy and Baker, 2005).

If deep coalescence is accepted as the sole explanation for gene trees that fail to support Glires, as is assumed by shortcut coalescence methods, this requires that (1) ancestral polymorphisms were carried across the ~5 MY stem branch of Glires at 175/183 loci, and (2) alleles at each of these loci were subsequently mis-sorted to yield 175 gene trees that conflict with Glires monophyly. This is an implausible scenario. If this were the case, why would supermatrix analysis of these same 175 conflicting loci robustly support Glires monophyly (Fig. 4)? It is much simpler to argue that homoplasy and the small size of the loci yielded many inaccurate gene trees, and that methods that assume all gene tree conflict is due to deep coalescence were unable to mine the hidden signal *within* each of these 175 conflicting loci. For both STAR and MP-EST, only 40% of the clades tested showed emergent support, whereas 100% of the clades expressed hidden support in supermatrix analyses. In five cases (Fig. 3), clades that were not recovered in coalescence analyses of conflicting loci were supported with 71–100% bootstrap support by supermatrix analyses of the same data. This implies that shortcut coalescence methods mask much of the phylogenetic signal that is embedded in McCormack et al.'s (2012) phylogenomic dataset.

By contrast with the results for Glires and several other clades, where hidden support emerged with concatenation but not coalescence, unexpected hidden coalescence support emerged at some nodes with both STAR and MP-EST (Figs. 3 and 5). For MP-EST, four clades showed apparent hidden support (Placentalia, Anthropoidea, Cetartiodactyla, Primates) and six clades did not (Rodentia, Afrotheria, Glires, Catarrhini, Hominidae, *Pan* + *Homo* + *Gorilla*). For STAR, there also were four clades that expressed hidden support (Placentalia, Anthropoidea, Rodentia, Primates), and six that did not (Cetartiodactyla, Afrotheria, Glires, Catarrhini, Hominidae, *Pan* + *Homo* + *Gorilla*). In the cases of Rodentia (STAR: 161 conflicting loci), Cetartiodactyla (MP-EST: 120 conflicting loci), and Primates (STAR and MP-EST; 132 conflicting loci), combinations of over 100 "wrong" loci yielded the "right" clade when all contrary gene trees were analyzed using shortcut coalescence methods. To our knowledge, these examples represent, by far, the most extreme documented cases of hidden support in coalescence analysis and demonstrate that clades contradicted by all input gene trees can emerge with this approach to systematics. Coalescence methods can prefer a minority gene tree in the anomaly zone (Degnan and Salter, 2005; Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007), but there is no precedent for coalescence methods favoring a clade that is contradicted by all input gene trees. For shortcut coalescence (Fig. 1D), it is theoretically possible for extreme hidden support to reveal itself with the combination of many conflicting loci, and in the empirical examples explored here, it is evident that this effect is not just hypothetical (Figs. 3 and 5). Like the MRP supertree method (Figs. 1C and 3), extreme emergent support with shortcut coalescence approaches may be a methodological defect that is unrelated to valid hidden support. This topic merits further investigation.

3.3. Synapomorphy-free clades in coalescence analyses

Hidden *coalescence* support should not be confused with hidden *character* support recorded in supermatrix studies. In the coalescence examples described above, emergent clades are not

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

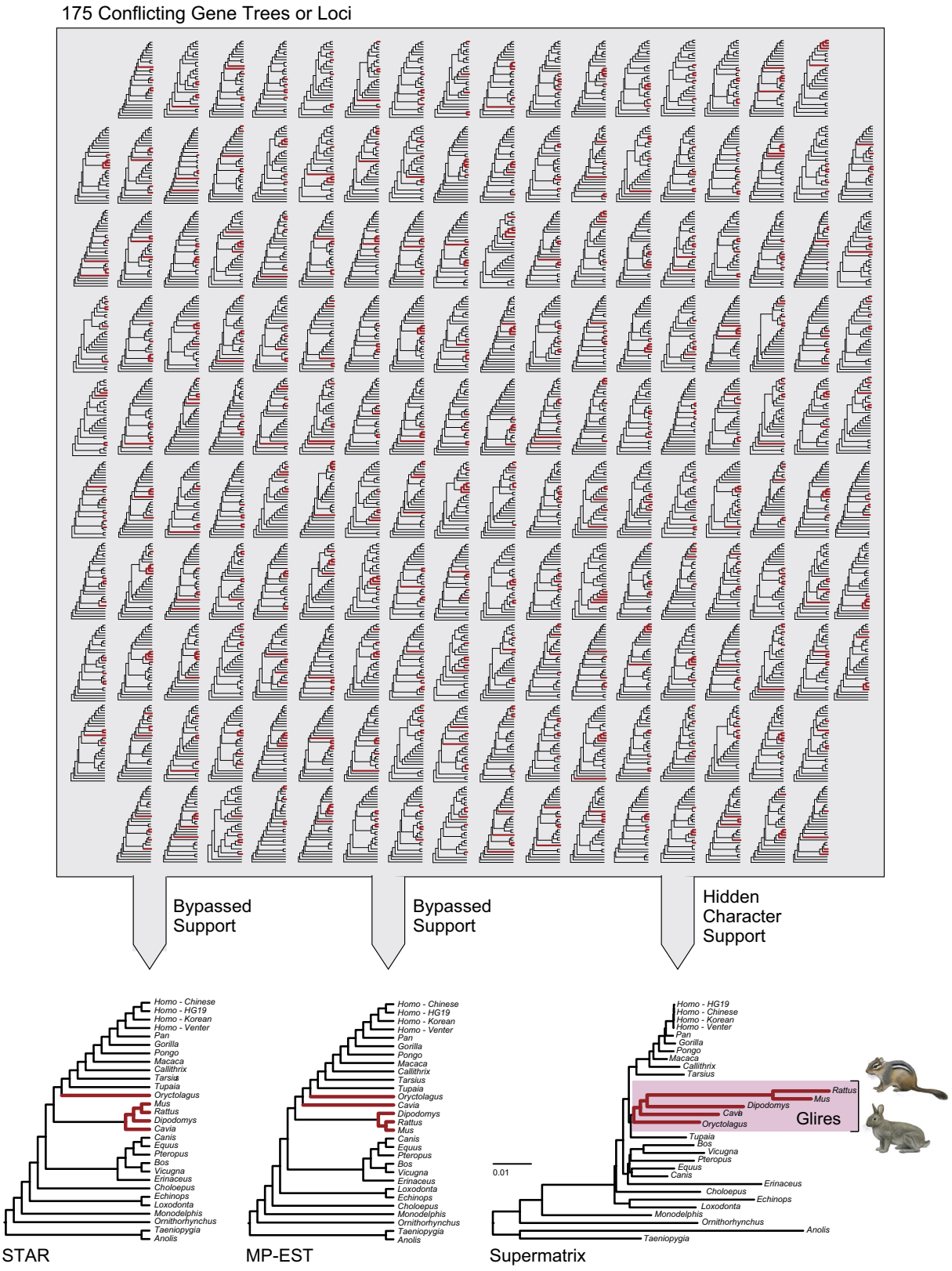

Fig. 4. Extreme hidden support that emerges in concatenation but not in shortcut coalescence analyses of 175 conflicting loci. Glires – the clade that includes rodents and rabbits – is supported by concatenation and coalescence (STAR) analysis of all 183 UCE loci, but only eight loci favor this clade in isolated analyses of individual loci. ML supermatrix analysis of the 175 conflicting loci resolves Glires with 81% bootstrap support (RAxML). Glires is not supported by STAR and MP-EST coalescence analyses of the 175 contradictory gene trees; the MP-EST coalescence analysis also did not support Rodentia, another traditionally recognized clade. Red branches in trees connect to the five members of Glires. Note the scattered distribution of the four rodents and single rabbit in the conflicting gene trees. The rapid rate of DNA substitution for rodents is evident in the supermatrix tree; branch lengths in this tree are proportional to change in expected substitutions per site. *Tamias* (chipmunk) and *Sylvilagus* (cottontail) are illustrated as representatives of Rodentia and Lagomorpha, respectively. Paintings are by C. Buell.

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Fig. 5. Ninety-one gene trees that do not support monophyly of Anthropoidea, extreme hidden character support in supermatrix analysis of these conflicting loci, and artifactual, synapomorphy-free coalescence support in STAR/MP-EST coalescence analyses of the 91 contradictory gene trees. Anthropoidea – the clade that includes humans, apes, and monkeys – is supported by concatenation and coalescence (STAR, MP-EST) analyses of all 183 UCE loci, but only 92 loci favor this clade in isolated analyses of each locus. ML supermatrix analysis of the 91 conflicting loci resolves Anthropoidea with high bootstrap support (100%). Anthropoidea also is supported by STAR and MP-EST coalescence analyses of the 91 contradictory gene trees shown. However, the stem branch of Anthropoidea is so long (~27 MY; Fig. 3) that no, or very little, retention of ancestral polymorphism is expected over the entire duration of this lineage (Fig. 11; Appendix A). Purple branches in trees connect to the six anthropoid species in each tree (four individuals of *Homo* are sampled). Branch lengths in the supermatrix tree are proportional to change in expected substitutions per site. Paintings are by C. Buell. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

underpinned by even a *single* synapomorphy. This is because (1) each individual input gene tree is interpreted as an accurate representation of evolutionary history in shortcut coalescence methods, (2) none of the gene trees support the clade of interest, (3) there are therefore no synapomorphies for that clade in *any* of the gene trees, and yet the clade is supported (Fig. 5). In the STAR and MP-EST analyses that yielded hidden support, the overall pattern of inferred deep coalescences across many loci supports a "synapomorphy-free" monophyletic group. This fact might be disconcerting to those who have been trained in traditional systematic methods, particularly cladistics, but is an outcome of the coalescence approach to systematics – for better or worse. Synapomorphy-free clades also have been generated in ML and Bayesian analyses of supermatrices with missing data, due to the assumption of common branch lengths across loci (Lemmon et al., 2009; Simmons and Goloboff, 2013). Whether such synapomorphy-free clades are desirable in supermatrix analysis is debatable (Wolsan and Sato, 2010; Simmons, 2012). In the case of shortcut coalescence methods, we contend that emergent, synapomorphy-free clades might in many cases be undesirable artifacts – examples of getting the right answer for the wrong reasons (see below).

3.4. Artifactual hidden support in coalescence analyses?

Both STAR and MP-EST expressed emergent support for Anthropoidea (Fig. 5; ~27 MY stem lineage) and for Placentalia (~84 MY stem lineage), implying that coalescence methods had rooted out common phylogenetic signals by correctly interpreting complex patterns of deep coalescence in the many gene trees that contradict these clades. However, for neutral loci, virtually no deep coalescence is expected on the long stem lineages to these clades (Appendix A), and even less under strong purifying selection – a likely scenario for UCEs that are hyper-conserved regions of the genome (McCormack et al., 2012). Molecular clock studies suggest that these clades do not represent anomaly zone situations (Fig. 3), and retention of ancestral polymorphism for ⩾84 MY at 44 independent loci (24% of the total) is not a tenable scientific explanation for the numerous gene trees that contradict placental monophyly. We contend that *none* of these conflicts are due to deep coalescence but instead result from a combination of factors that have been characterized in previous systematic work. These problems include lack of informative variation in short loci (Bull et al., 1993; Huelsenbeck and Hillis, 1993; Swofford et al., 2001), arbitrary resolution (Simmons and Norton, 2013), long branches (Felsenstein, 1978; Swofford et al., 2001; Anderson and Swofford, 2004; Bergsten, 2005), unevenly distributed homoplasy (Sanderson, 1989; Gatesy, 2000), poor taxon sampling (Zwickl and Hillis, 2002; Hedtke et al., 2006), and general substitution model misfit (Sullivan and Swofford, 1997; Collins et al., 2005; Ren et al., 2005; Morgan et al., 2013), all of which generate gene tree conflicts at ancient divergences.

It is perplexing that coalescence methods, which assume all gene tree conflicts are due to deep coalescence, can process 44 wrong gene trees and arrive at the right answer, placental monophyly, when none of the gene trees are wrong due to deep coalescence. STAR and MP-EST derive species trees by sorting out deep coalescences among gene trees, not by reconciling homoplasy at individual sites as in the supermatrix approach. One could argue that shortcut coalescence methods are robust to a few aberrant gene trees (e.g., Liu et al., 2010; Song et al., 2012), and uncritically accept the hidden support results, but for the case of Placentalia, *all* 44 gene trees are outliers with deep coalescences that in every case are ⩾84 MY. The probability for retention of ancestral polymorphism across 84 MY is vanishingly small (Appendix A). For Anthropoidea, 91 loci contradict this clade in separate analyses (Fig. 5); if the 91 contrary gene trees are accurate as assumed in the coalescence methods, this presumes that ~50% of 183 loci retained polymorphism for ⩾27 MY, a pattern that is inconsistent with both theory and previously published empirical work (Appendix A). Our preferred interpretation is that the coalescence methods were right for the wrong reasons at the Placentalia and Anthropoidea nodes, and the emergent support is artifactual "pseudo-hidden support," a serendipitous result that has no obvious biological interpretation.

The above arguments are based on the assumption that the short loci utilized by McCormack et al. (2012) do not yield accurate gene trees. It is therefore necessary to examine in more detail the basic data partitions in McCormack et al.'s (2012) phylogenomic dataset and compare this information to loci analyzed in other large-scale datasets for Mammalia (Meredith et al., 2011; Song et al., 2012). Because deep evolutionary relationships within Mammalia have been well studied, the credibility of gene trees generated from different studies can be assessed directly. Furthermore, the timing of inferred deep coalescences can be measured with rigor based on well-calibrated molecular clock studies that constrain ancient divergences among placental mammals (Fig. 3).

3.5. McCormack et al. (2012) versus Song et al. (2012): coalescence versus "concatalescence"

Coalescence-genes ("c-genes"; Doyle, 1992, 1997) are the basic units of analysis in the coalescence approach to systematics (Slowinski and Page, 1999). C-genes are DNA segments characterized by no internal recombination over the history of a clade and are therefore considered independent evolutionary modules with a common branching history (Fig. 6; Maddison, 1997). Depending on the number of recombination events, c-genes might range from thousands of sites for non-recombining genomic regions (e.g., some organellar DNA) to a single nucleotide position (Doyle, 1997; Maddison, 1997; Edwards, 2009). Gene trees that represent the hierarchical evolutionary histories of multiple c-genes are used to reconstruct phylogeny in the coalescence paradigm (reviewed in Edwards, 2009). Proponents of coalescence methods contend that the ability to accommodate the conflicting histories of independent c-genes is critical and contrasts with supermatrix methods that do not explicitly account for retention of ancestral polymorphism across multiple nodes of a tree and instead impose a single branching history on multiple c-genes (Edwards, 2009; Knowles, 2010). Reconstructing the evolutionary histories of these minimal units of non-recombining DNA (Fig. 6) is therefore *the most basic rationale for the coalescence approach to systematics* (Doyle, 1997; Maddison, 1997; Slowinski and Page, 1999; Edwards, 2009).

In phylogenomic analyses of placental mammals, Song et al. (2012) and McCormack et al. (2012) argued assertively and repeatedly for the clear-cut superiority of shortcut coalescence approaches relative to supermatrix methods (e.g., Meredith et al., 2011). These authors focused their attention on the few clades resolved by coalescence analyses of 183–917 loci that conflicted with supermatrix studies; these areas of conflict are at three difficult to resolve polytomies, and multiple molecular clock studies suggest that all three date to >70 MYA and represent rapid splitting events (Fig. 3). Both Song et al. (2012) and McCormack et al. (2012) argued that supermatrix analyses have been grossly misled at these tightly spaced nodes by not properly accounting for incomplete lineage sorting at different loci, resulting in disagreements with their technically superior coalescence results. Both studies concluded that coalescence had robustly resolved difficult polytomies that were intractable with a supermatrix approach. Within each study, phylogenetic results derived from shortcut coalescence analyses were consistent at controversial nodes. In each case, various subsets of loci from their total sample supported the same topology, which led these authors to conclude that coalescence

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

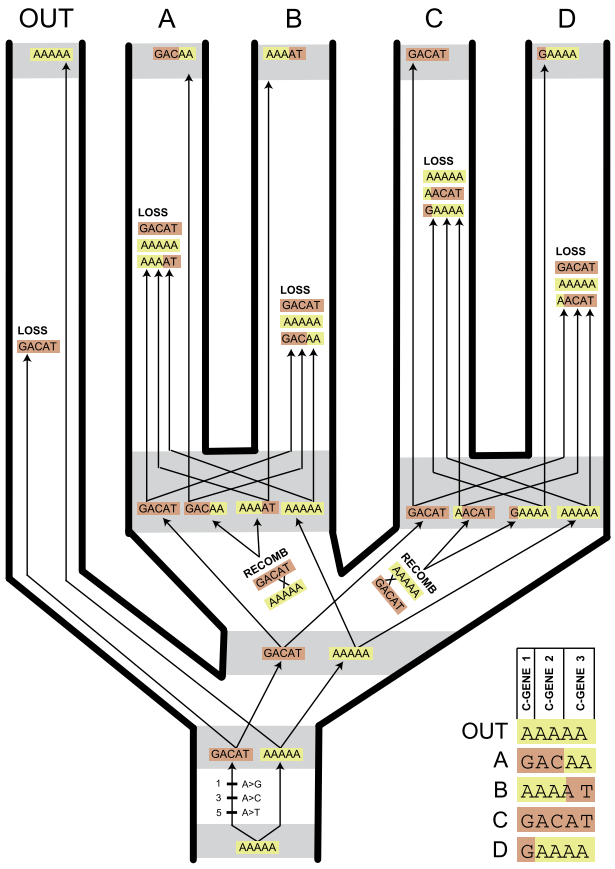

Fig. 6. Recombination reduces the length of c-genes. In this hypothetical scenario, the true evolutionary history is shown in the large tree with four tightly spaced speciation events. A single allele is ancestral. Mutations at nucleotide positions 1, 3, and 5 at the base of the tree yield a second allele. Alternative alleles then undergo two recombination events and various alleles are fixed in each of the five terminal lineages (A–D and OUT). The data matrix (lower right) shows the delimitation of the three c-genes derived from the evolutionary history shown. Note that if fewer species were sampled, there would be fewer c-genes in some cases; for example if just 'OUT' and C were sampled, there would be only one c-gene, and if 'OUT, B, and C were sampled, there would be only two c-genes. With the sampling of even more taxa than shown here, the number of c-genes for this locus could ratchet even higher, but could not decrease. With the addition of taxa, note that the number of internodes in each gene tree goes higher, but the length of the c-gene simultaneously gets shorter – a debilitating reality for current shortcut coalescence methods that estimate gene trees in isolation from estimation of the species tree.

methods had determined the correct answer (McCormack et al., 2012; Song et al., 2012).

Unfortunately, the strongly supported results of these two phylogenomic coalescence studies directly contradict each other at the root of the placental tree (Fig. 3; McCormack et al., 2012; Song et al., 2012). It is therefore impossible for both to be right, no matter how adamantly the authors argued their respective cases. Either one, or both, of the coalescence studies must be flawed – both cannot represent the proper species tree for Placentalia (Gatesy and Springer, 2013). Furthermore, within Laurasiatheria, yet another recent phylogenomic coalescence analysis of placental mammals (Hallström et al., 2011) conflicts robustly with the coalescence tree favored by Song et al. (2012) and also contradicts McCormack et al. (2012) (Fig. 3). The discrepant nodes in the three coalescence studies are the same ones that could not be resolved previously by supermatrix and coalescence analyses, even with much denser taxon sampling (Meredith et al., 2011). In sum, discordant coalescence results directly refute the claim that supermatrix analyses, but not coalescence studies, are prone to robustly supported conflicts, such as these, in analyses of genome-scale data (Song et al., 2012; Knowles et al., 2014; Zhong et al., 2014). Instead, the contrasting results between different coalescence studies (Hallström et al., 2011; McCormack et al., 2012; Song et al., 2012) suggest that closer examination might provide general lessons regarding limitations of particular coalescence studies or the entire coalescence approach to systematics. These results also demonstrate that the procurement of consistent results for different subsets of genes should not be confused with accuracy, as was argued in Song et al. (2012), McCormack et al. (2012), and in other coalescence analyses (e.g., Xi et al., 2013; Zhong et al., 2013). Consistent results may be right, or they may be incorrect and instead reflect a shared underlying bias in a particular method or dataset (Gatesy and Baker, 2005; Gatesy et al., 2007; Springer and Gatesy, 2014).

The most striking difference between the coalescence work of McCormack et al. (2012) and Song et al. (2012) is the strategy used to delimit c-genes in the two contradictory analyses (Gatesy and Springer, 2013). Published mammalian genome sequences were partitioned in both studies. McCormack et al. (2012) took the more justified approach and constructed gene trees based on relatively short (~410 bp), distantly linked regions that flank UCEs in the genome (Fig. 7). Based on previous genomic comparisons of mammals (e.g., Hobolth et al., 2007, 2011), it is at least possible that sequences of this length have not been subdivided by numerous recombination events over the ~200 million year history of Mammalia (Fig. 3) and might roughly correspond to c-genes. Song et al. (2012), by contrast, chose to analyze much longer sections of the genome to increase the information content of each locus included in their study. Song et al. (2012) reported an average c-gene length of ~3100 bp (Fig. 7), but assembled their dataset from the OrthoMaM database (Ranwez et al., 2007), which is composed of protein-coding sequences trimmed of introns, and concatenated exons that are spread across broad swaths of the genome. Many of the genes employed by Song et al. (2012) span >100,000 bp, with the average length being 139,600 bp (median = 96,000 bp). Protein-coding sequences of three loci (*PRKG1, GALNTL6, IL1RAPL1*) each cover more than 1,200,000 bp in the genome of *Homo sapiens* (Gatesy and Springer, 2013).

Song et al. (2012) therefore merged aspects of supermatrix methods – concatenation of exons up to a million bp apart – with features of coalescence methods – resolution of conflicts among gene trees based solely on the relative depths of inferred coalescences. This hybrid "concatalescence" approach is inconsistent with the most basic rationale for performing a coalescence analysis in the first place (Gatesy and Springer, 2013; Springer and Gatesy, 2014). If multiple c-genes are concatenated, an almost certain prospect when individual exons of a protein-coding gene span >100,000 bp of the genome, coalescence analysis of these constructs is confronted with the same critique that proponents of coalescence have leveled against the supermatrix approach – a confusing mixture of different historical signals (Edwards, 2009; Lanier and Knowles, 2012). The robustly conflicting results of Song et al. (2012) and McCormack et al. (2012) thus might be due to the fact that the former is not a valid *coalescence* analysis and is instead a *concatalescence* analysis (Gatesy and Springer, 2013). Song et al. (2012) cited coalescence simulations and theory as assurances that their study had accurately resolved three controversial polytomies at and near the base of Placentalia (Fig. 3). However, the concatalescence procedure employed by Song et al. (2012) is a composite approach, so it would be premature to attribute any successes or failures of this procedure to its coalescence features, or to its supermatrix features (see Bayzid and Warnow, 2013).

The intermingling of coalescence and concatenation in Song et al.'s (2012) concatalescence analyses does not imply that the strongly discordant coalescence results of McCormack et al.

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

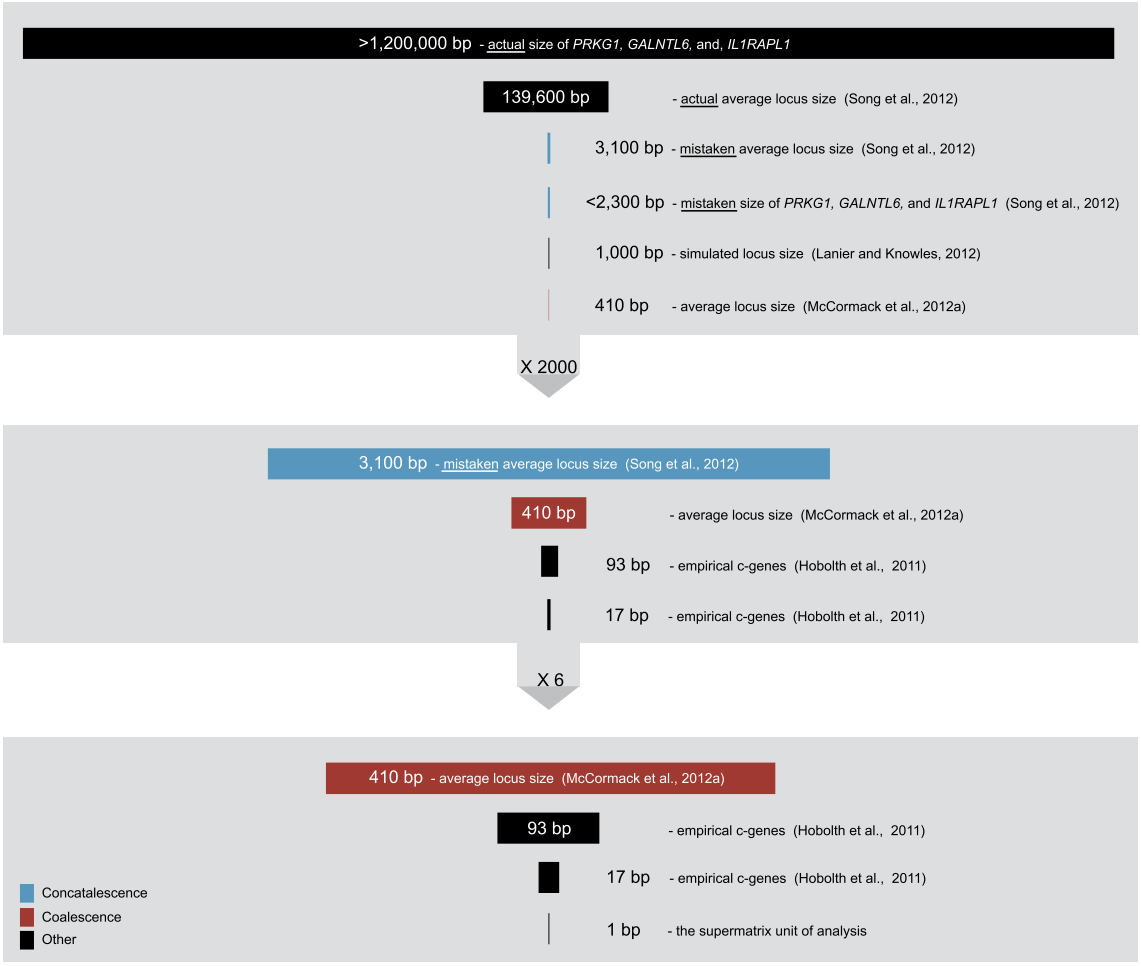

Fig. 7. Lengths of loci analyzed in various coalescence studies (McCormack et al., 2012; Song et al., 2012; Lanier and Knowles, 2012) in comparison to empirically estimated c-genes in primates (Hobolth et al., 2007, 2011). The following are shown: (1) actual sizes of *PRKG1*, *GALNTL6*, and *IL1RAPL1* genes in *Homo*, (2) mistaken lengths of *PRKG1*, *GALNTL6*, and *IL1RAPL1* genes from Song et al. (2012), (3) actual average locus length for 447 loci analyzed by Song et al. (2012), (4) mistaken average locus length for 447 loci according to Song et al. (2012) with introns not accounted for, (5) size of simulated loci in recombination analysis of Lanier and Knowles (2012), (6) average locus size in McCormack et al. (2012), (7) sizes of conflicting c-genes in Hobolth et al. (2011) for four species of primates, and (8) the supermatrix unit of analysis. Note the multipliers (arrows) between gray boxes, and that empirical c-genes in primates are closer in size to the basic unit used in supermatrix studies (1 bp) than to loci delimited in coalescence analyses of mammals (McCormack et al., 2012; Song et al., 2012). In Hobolth et al. (2011), conflicting c-gene size was <100 bp on average with 75% of c-genes estimated to be between 17 bp and 93 bp. These two endpoints are shown in the figure, but given that many more taxa were included in the UCE dataset, conflicting c-gene size might be expected to be much smaller due to additional recombination events (Fig. 6) at tight internodes (Fig. 3).

(2012) represent a significant advance relative to Song et al. (2012) or to prior supermatrix analyses of placental mammals (e.g., Meredith et al., 2011). Indeed, an ongoing critique of coalescence analyses at deep phylogenetic levels is the poor information content of c-genes. Numerous authors have suggested that analysis of very short c-genes might undermine coalescence analysis (Huang et al., 2010; Meredith et al., 2011; Townsend et al., 2011; Chiari et al., 2012; Knowles et al., 2012; Rosenfeld et al., 2012; Bayzid and Warnow, 2013; Gatesy and Springer, 2013; Patel et al., 2013; Salichos and Rokas, 2013; DeGiorgio and Degnan, 2014; Lanier et al., 2014; Springer and Gatesy, 2014). In particular, because shortcut coalescence methods are critically dependent on the accurate reconstruction of gene trees (Kubatko et al., 2009; Liu et al., 2009b, 2010), mis-estimated gene trees could debilitate phylogenetic inference, especially at deep nodes in the Tree of Life. Given the short length of c-genes analyzed in McCormack et al. (2012), poor gene trees based on uninformative loci might be another valid explanation for topological conflicts with concatalescence analyses (Hallström et al., 2011; Song et al., 2012).

3.6. Short c-gene alignments are short on informative variation

McCormack et al. (2012) argued that, because of extreme negative selection on UCEs, their data showed less evidence of saturation through multiple hits relative to protein-coding exons that have been utilized in previous studies of mammalian phylogeny (e.g., Meredith et al., 2011). This might be correct, but the loci sampled by McCormack et al. (2012) also show precariously low levels of informative variation (Table 1). For the 22 placental taxa in the dataset, there are 20 internal nodes that define relationships among these ingroup species (Fig. 3). Yet, for the ingroup sequences, 38 of 183 loci include ten or fewer parsimony informative sites, 101 of the 183 alignments include ⩽20 informative sites, and 134 of 183 have ⩽30 informative sites (Table 1). By seeking to minimize homoplasy and by analyzing relatively short c-genes of justifiable length (~410 bp), McCormack et al. (2012) ended up with loci that provide insufficient information for reconstructing individual gene trees. Eighty-eight loci have fewer informative sites than internal branches to resolve (Table 1).

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Table 1
Locus number, locus name, the number of parsimony informative (PI) sites with and without the outgroup taxa, and alignment length for 183 loci that were employed by McCormack et al. (2012) to construct a mammalian species tree with shortcut coalescence methods.

Locus #	Locus name	# PI with outgroups	#PI without outgroups	Alignment length (bp)
1	chr10_4411	23	14	190
2	chr10_4414	19	15	352
3	chr11_1600	16	9	251
4	chr11_3210	53	33	1036
5	chr11_3432	15	13	266
6	chr11_4723	152	115	315
7	chr11_4735	44	23	232
8	chr11_4777	54	32	364
9	chr12_5828	59	42	228
10	chr12_5837	112	63	814
11	chr12_5895	32	21	386
12	chr12_5908	51	39	291
13	chr12_5999	105	70	717
14	chr13_2821	104	83	735
15	chr13_2870	68	53	462
16	chr13_3033	138	91	608
17	chr13_3036	42	31	344
18	chr13_710	27	13	386
19	chr13_714	17	10	226
20	chr17_3088	30	12	338
21	chr18_3123	39	28	449
22	chr19_365	6	3	191
23	chr1_15480	45	24	294
24	chr1_28829	50	43	637
25	chr1_31654	40	26	567
26	chr1_31767	38	20	378
27	chr1_32215	2	2	321
28	chr1_32246	37	27	547
29	chr1_32247	8	1	119
30	chr1_32248	20	10	197
31	chr1_32424	31	21	387
32	chr1_5279	35	19	604
33	chr1_5365	19	10	281
34	chr1_5460	140	81	690
35	chr1_5478	86	62	470
36	chr1_5522	105	40	655
37	chr1_8696	10	4	222
38	chr20_5193	23	16	215
39	chr22_640	42	30	378
40	chr2_11740	39	31	362
41	chr2_13064	100	82	324
42	chr2_17003	84	62	483
43	chr2_18144	38	16	526
44	chr2_18608	37	27	371
45	chr2_22071	45	29	488
46	chr2_22128	31	21	361
47	chr2_22616	11	9	184
48	chr2_22628	7	7	280
49	chr2_23124	49	36	365
50	chr2_23281	264	202	834
51	chr2_23548	37	23	456
52	chr2_23668	12	7	557
53	chr2_23670	19	15	314
54	chr2_24815	96	62	996
55	chr2_24825	17	6	315
56	chr2_25783	22	16	276
57	chr2_25816	30	20	302
58	chr2_25941	32	20	317
59	chr2_25984	8	2	103
60	chr2_26677	75	47	746
61	chr2_27968	27	20	400
62	chr2_6767	66	42	452
63	chr2_7915	40	28	205
64	chr2_7954	23	15	319
65	chr2_7957	48	27	670
66	chr2_7982	24	15	367
67	chr2_8593	17	16	192
68	chr2_8655	18	12	491
69	chr2_8698	15	6	516
70	chr3_14142	37	24	364
71	chr3_17712	10	6	166
72	chr3_17769	33	27	338
73	chr3_18148	27	15	197

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Table 1 (*continued*)

Locus #	Locus name	# PI with outgroups	#PI without outgroups	Alignment length (bp)
74	chr3_18256	6	6	118
75	chr3_226	22	13	311
76	chr3_23229	25	7	344
77	chr3_24775	23	18	264
78	chr3_24903	50	20	413
79	chr3_24921	52	38	267
80	chr3_3073	31	19	401
81	chr3_3107	18	16	171
82	chr3_3177	25	13	230
83	chr3_3180	22	14	264
84	chr3_454	64	40	506
85	chr3_5441	62	57	535
86	chr3_5834	39	27	267
87	chr3_5924	65	58	316
88	chr3_6115	48	40	333
89	chr3_6135	108	80	781
90	chr3_6248	33	17	356
91	chr3_764	51	21	685
92	chr3_9696	26	15	297
93	chr3_9700	13	6	240
94	chr3_9709	25	17	389
95	chr3_9719	26	13	332
96	chr3_9749	33	22	392
97	chr4_13431	20	18	251
98	chr4_15987	19	9	555
99	chr4_16780	34	18	302
100	chr4_17854	27	15	217
101	chr4_4009	20	12	343
102	chr4_4926	33	23	364
103	chr4_7559	23	14	313
104	chr4_7570	46	25	307
105	chr4_7577	16	6	342
106	chr4_9731	50	29	357
107	chr5_10047	66	38	605
108	chr5_10180	29	15	209
109	chr5_10260	45	20	637
110	chr5_10841	30	20	371
111	chr5_11146	6	5	124
112	chr5_11231	11	5	269
113	chr5_11292	84	50	815
114	chr5_11304	15	9	161
115	chr5_12015	21	11	300
116	chr5_1649	32	19	298
117	chr5_2303	52	22	538
118	chr5_2978	21	6	370
119	chr5_3310	179	114	966
120	chr6_3147	66	53	323
121	chr6_3228	34	22	382
122	chr6_3274	36	19	248
123	chr6_3284	10	7	160
124	chr6_3291	21	15	330
125	chr6_4403	18	11	307
126	chr6_4409	114	75	601
127	chr6_4410	8	2	216
128	chr6_8057	24	20	401
129	chr6_8064	19	9	629
130	chr6_8786	27	20	372
131	chr6_9631	19	10	240
132	chr7_10269	16	10	508
133	chr7_10353	8	7	212
134	chr7_10443	17	15	288
135	chr7_10445	56	47	564
136	chr7_10687	78	51	623
137	chr7_10849	25	14	440
138	chr7_11467	10	8	231
139	chr7_11510	116	77	613
140	chr7_11511	19	9	339
141	chr7_11565	18	9	426
142	chr7_4670	42	17	606
143	chr7_4698	41	28	647
144	chr7_4732	30	19	323
145	chr7_4747	31	18	421
146	chr7_4777	72	46	422
147	chr7_4941	56	43	678
148	chr7_4994	34	26	297

(*continued on next page*)

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Table 1 (*continued*)

Locus #	Locus name	# PI with outgroups	#PI without outgroups	Alignment length (bp)
149	chr7_5167	75	49	674
150	chr7_5228	35	24	654
151	chr8_1574	22	12	211
152	chr8_3279	20	8	287
153	chr8_4221	32	25	445
154	chr8_4241	36	25	463
155	chr8_4319	70	53	324
156	chr8_4947	37	19	342
157	chr8_5307	38	25	469
158	chr8_675	220	141	760
159	chr8_9158	40	29	348
160	chr9_3633	129	88	631
161	chr9_393	20	12	226
162	chr9_5190	85	68	719
163	chr9_5218	29	11	339
164	chr9_5246	49	38	738
165	chr9_6343	99	60	662
166	chr9_6414	74	43	671
167	chr9_7170	23	16	491
168	chrZ_11420	29	7	596
169	chrZ_11557	14	11	226
170	chrZ_11569	18	13	236
171	chrZ_3358	125	67	738
172	chrZ_3515	204	126	783
173	chrZ_4787	6	6	165
174	chrZ_495	53	25	341
175	chrZ_532	34	19	478
176	chrZ_5495	9	9	109
177	chrZ_6036	36	20	411
178	chrZ_6615	37	19	340
179	chrZ_6727	38	27	632
180	chrZ_6778	28	20	311
181	chrZ_72	47	20	328
182	chrZ_7924	44	20	551
183	chrZ_8059	7	5	198
Mean		44.0	28.4	408.5
Median		33	20	357

3.7. Uninformative loci yield peculiar gene trees

Relationships among orders of placental mammals have been difficult to differentiate owing to extreme rate heterogeneity among clades, several short internal branches that reflect rapid cladogenesis, and saturation problems that result from the temporal depth of the early placental radiation (Fig. 3; Springer et al., 2004; Meredith et al., 2011; dos Reis et al., 2012). Concatenations with thousands of phylogenetically informative sites ultimately were required to provide robust support for a family-level tree (Madsen et al., 2001; Murphy et al., 2001a, 2001b; Springer et al., 2003; Meredith et al., 2011). By contrast, the short loci of McCormack et al. (2012) would not be expected to yield accurate gene trees for this challenging and deep systematic problem.

Most loci analyzed by McCormack et al. (2012) have few parsimony informative characters and support unconventional relationships. The five least variable loci comprise a total of only ten informative characters for the ingroup (Table 1). Combined, these five loci are incapable of reconstructing phylogenetic relationships among 22 species – only one informative character for every two internal branches in the tree. Separately, data partitions with one to three informative characters have almost nothing to offer regarding relationships among placental orders (Fig. 8). In the coalescence approach to phylogenetic analysis, the c-gene is the basic unit of analysis, so each of these short loci is given *equal weight* in analysis relative to the very few loci in this dataset that are strongly congruent with published systematic work. Therefore, in coalescence analyses, a locus with one informative character (locus 29) is equivalent to an information-rich locus with 202 informative characters (locus 50) (Table 1). Some proponents of coalescence methods have asserted that the supermatrix approach mis-weights data from different loci (e.g., McVay and Carstens, 2013), but shortcut coalescence methods clearly are prone to analogous problems.

For the short and uninformative loci, multiple species, even distantly related species such as a bird and a dog that diverged >300 MY ago (Fig. 3), commonly have identical sequences (e.g., locus 22), and a variety of counterintuitive relationships are resolved (Fig. 8). The 13 loci from McCormack et al. (2012) with the median number of informative nucleotide positions, 20 sites (Table 1), likewise support odd relationships that have never been reported in any previous study of mammalian phylogeny, and sequences from distantly related species (e.g., locus 61: dog, horse, chimpanzee) again are identical (Fig. 9). This suggests that the sampling of genetic loci in this study is not adequate for the difficult task at hand – reconstruction of accurate gene trees that are required for successful implementation of shortcut coalescence methods (Liu et al., 2009b, 2010).

Given limited informative variation (Table 1), the inferred gene trees from McCormack et al. (2012) are characterized by numerous, very short, internal branches (Figs. 8–10). The total number of internal branches across the 183 gene trees is 4575, but the majority of these internal branches, 2458 (54%), are $\leqslant$0.000001 expected substitutions per site. These diminutive branches are the product of an ML search routine in PhyML (Guindon et al., 2010) that saves only one fully resolved tree per search (Simmons and Norton, 2013). Nevertheless, this routine was utilized in McCormack et al. (2012) and other coalescence analyses of placental phylogeny (Liu et al., 2009b, 2010; Song et al., 2012). There is no evidence of an

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

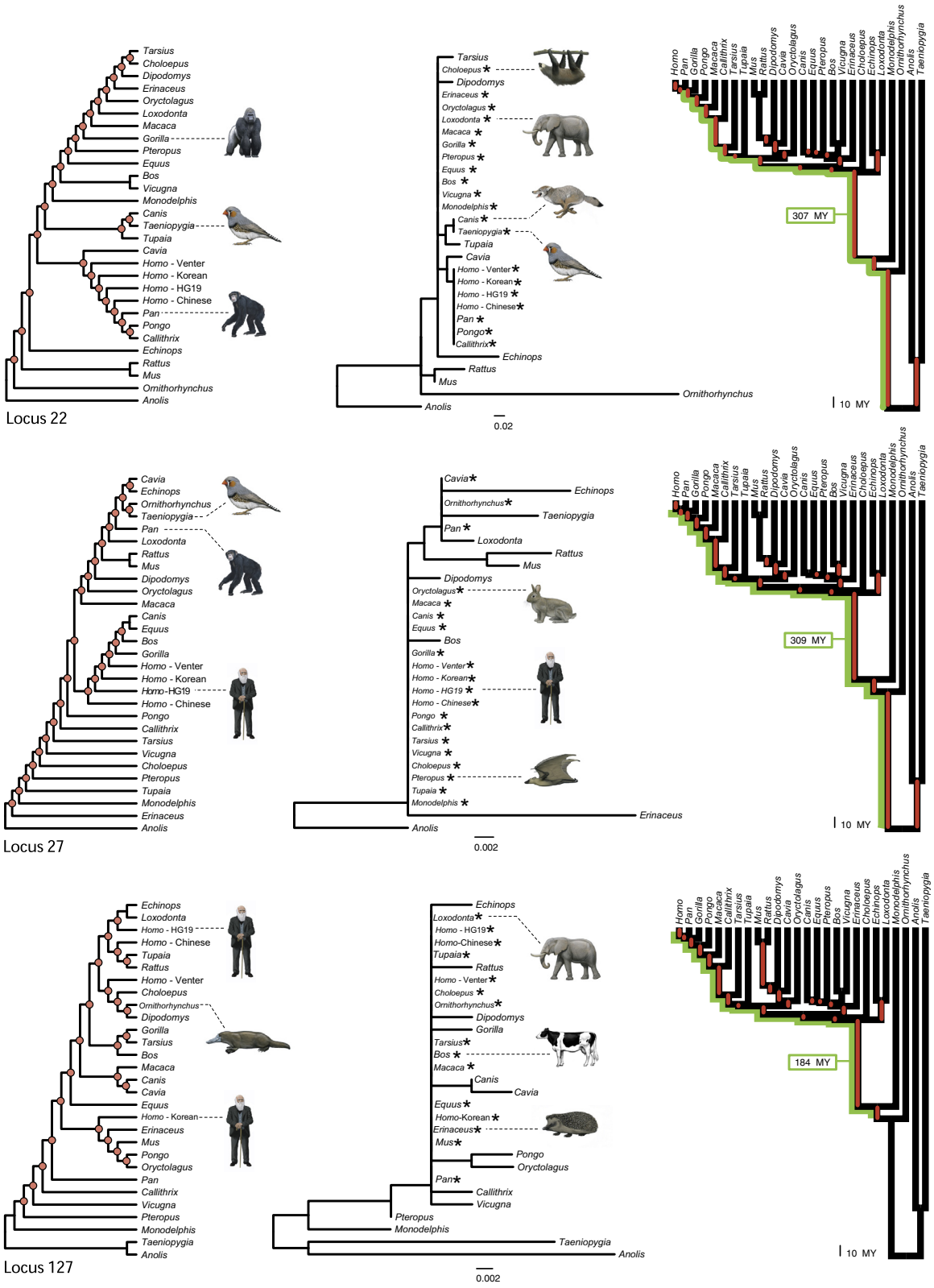

Fig. 8. Highly incongruent gene trees from McCormack et al. (2012) with unrealistic deep coalescences. ML trees for three loci with <10 parsimony-informative characters each are illustrated. The fully resolved gene trees that were used in STAR coalescence analysis are shown on the left with nodes that conflict with the inferred species tree (Fig. 3) marked by red circles. Especially strange relationships, that if true demand exceptionally long retentions of ancestral polymorphism, are indicated. The same gene trees are shown in the middle column but with branch lengths (expected substitutions per site). Taxa with identical sequences are denoted with asterisks (*). The trees on the right represent the inferred species tree based on STAR coalescence analysis of all 183 loci from McCormack et al. (2012) with branch lengths proportional to time and scale bar in millions of years (MY; see Fig. 3). For each locus, internodes in the species tree that conflict with the gene tree are colored red, and the longest deep coalescence is shown for each locus (green lineages). Although highly incongruent with the overall species tree, these gene trees are not outliers for the UCE dataset (McCormack et al., 2012). Within Placentalia, the ingroup, an average of only ~4.6 of 20 clades in the species tree (Fig. 3) were supported by each gene tree. Paintings are by C. Buell. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

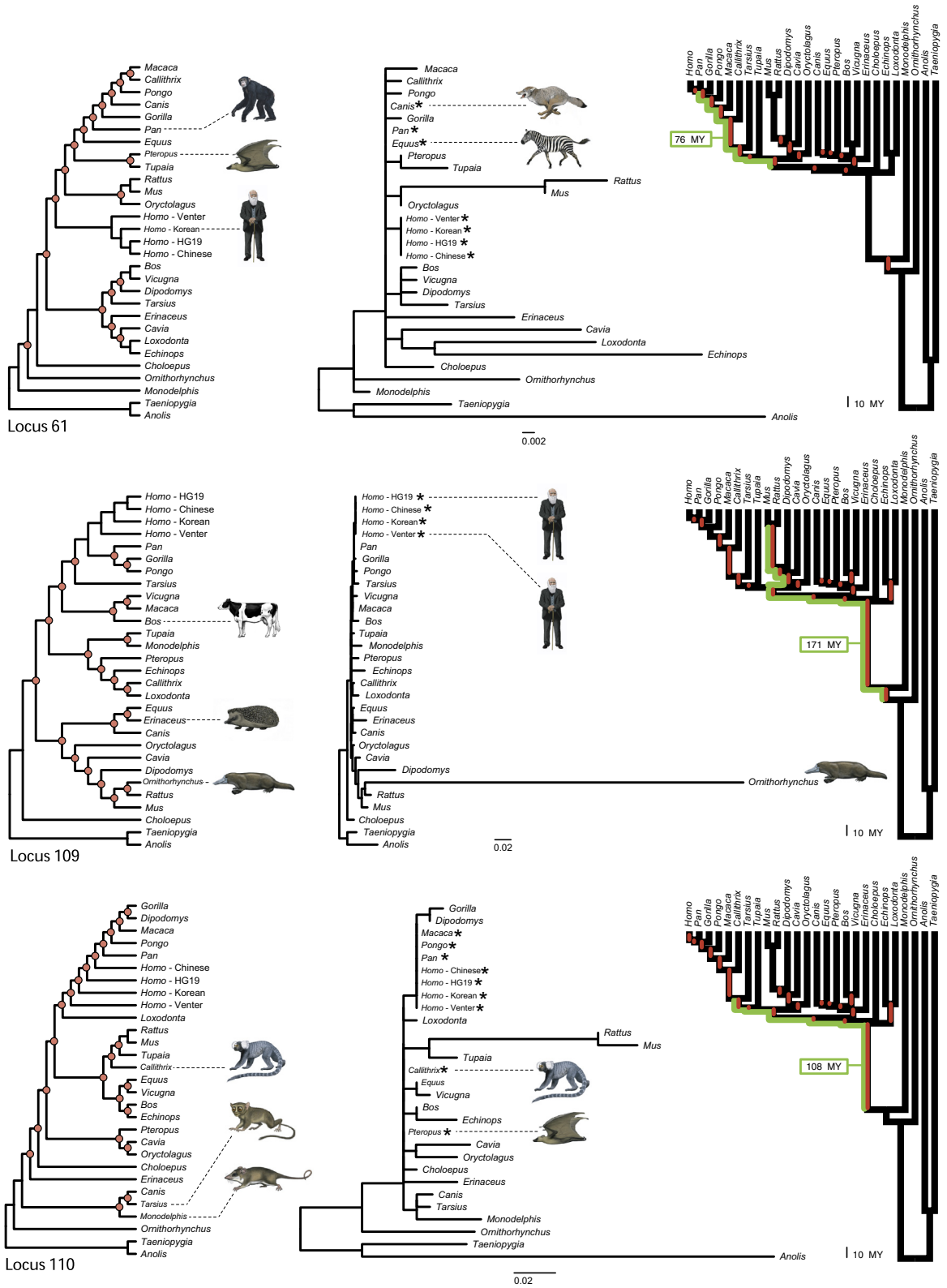

Fig. 9. Highly incongruent gene trees from McCormack et al. (2012) with unrealistic deep coalescences. ML trees for three loci with the median variation (20 parsimony-informative characters) are illustrated. The fully resolved ML gene trees that were used in STAR coalescence analysis are shown on the left with nodes that conflict with the inferred species tree (Fig. 3) marked by red circles. Especially bizarre relationships that if true, demand exceptionally long retentions of ancestral polymorphism, are indicated. The same gene trees are shown in the middle column but with branch lengths (substitutions per site). Taxa with identical sequences are denoted with asterisks (*). The trees on the right represent the inferred species tree based on STAR coalescence analysis of all 183 loci from McCormack et al. (2012) with branch lengths proportional to time and scale bar in millions of years (MY; see Fig. 3). For each locus, internodes in the species tree that conflict with the gene tree are colored red, and the longest deep coalescence is shown for each locus (green lineages). Although highly incongruent with the overall species tree, these gene trees are not outliers for the UCE dataset (McCormack et al., 2012). Within Placentalia, an average of only ~4.6 of 20 clades in the species tree (Fig. 3) were supported by each gene tree. Paintings are by C. Buell. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

actual nucleotide substitution at these short internodes, only 1/1000 of a substitution or less. Even when sequences from different species are identical for a particular locus, the algorithm still resolves a fully bifurcating gene tree (e.g., Fig. 8; locus 27); such internodes are predictably highly incongruent with all published phylogenetic hypotheses for Placentalia, including the preferred topology of McCormack et al. (2012) (Fig. 3).

For many commonly used shortcut coalescence methods – STAR, STEAC, MP-EST – only fully resolved (bifurcating) gene trees are permitted as input, so arbitrary resolution of gene trees is a prerequisite of these methods when analyzing loci with poor information content (Liu et al., 2009b, 2010). For STEM, the user manual warns against inputting zero length branches in gene trees, and ominously notes that zero length branches, "...may not result in an error" (Kubatko et al., 2009; also see DeGiorgio and Degnan, 2014). In addition to inadequate variation (Table 1), arbitrary resolutions of internal branches in c-gene trees for McCormack et al.'s (2012) mammalian dataset contribute to the numerous conflicts with established, uncontroversial relationships among placental mammals (Figs. 8–10).

Many gene trees from McCormack et al. (2012) additionally show long branches, extreme rate heterogeneity, and still more peculiar relationships (Figs. 9 and 10). These include: (a) a sister-group relationship between *Rattus* (rat) and *Ornithorhynchus* (duckbilled platypus – long branch) to the exclusion of *Mus* (mouse) (locus 109), (b) a clade composed of the xenarthran *Choloepus* (sloth – long branch) and the perissodactyl *Equus* (horse) (locus 129), (c) a grouping of the primate *Tarsius* (tarsier – long branch) with the afrothere *Echinops* (tenrec), with another primate – *Pongo* (orangutan) – closer to a bird (*Taeniopygia*) (locus 3), (d) the cetartiodactyl *Vicugna* (alpaca – long branch) sister to a peculiar clade that features murid rodents (*Rattus* and *Mus* – long branches) clustered within Afrotheria (elephant + tenrec), with a second cetartiodactyl (*Bos*) and two other rodents (*Cavia*, *Dipodomys*) more distant in this gene tree (locus 53). These are only a few examples, among many other unlikely groupings, that have never been suggested in the past 150 years of research on mammalian systematics (Figs. 9 and 10). Thus, extreme rate heterogeneity and long branches add extra layers of complication to the reconstruction of reliable gene trees in McCormack et al.'s (2012) study.

3.8. Strange gene trees imply impossibly deep coalescences

In recent phylogenetic analyses of placental mammals, rampant conflicts among gene trees were forwarded as a rationale for utilizing coalescence methods (Liu et al., 2009b, 2010; McCormack et al., 2012; Song et al., 2012; Kumar et al., 2013; but see Meredith et al., 2011). We instead hypothesize that most of these conflicts are due to more mundane problems that have been recognized for decades in the systematics literature (see Patel et al., 2013).

Due to these various problems and in the context of multiple molecular clock studies of Mammalia (e.g., Fig. 3), the 183 ML gene trees from McCormack et al. (2012) imply highly improbable deep coalescences and suggest that particular ancestral polymorphisms have been retained for over 50 MY of evolutionary history at 66% of the loci sampled (Figs. 8–10; Table 2). The deepest coalescence in each gene tree averages 101.2 MY (Table 2; median 78.7 MY). Given the extreme conservation of UCEs, long coalescence times such as these would not be expected because of strong purifying selection against new genetic variants at these loci (McCormack et al., 2012). Nonetheless, McCormack et al.'s (2012) gene trees suggest that ancestral polymorphisms have been maintained for >200 MY at 21 of 183 loci (Figs. 8 and 10; Table 2), a pattern that is inconsistent with previous empirical and theoretical work (Appendix A; Leffler et al., 2013). Even alleles that are maintained by balancing selection within Mammalia, such as at major histocompatibility (MHC) loci (Piertney and Oliver, 2006 and references therein), are not known to have deep coalescences of the magnitudes implied by *most* gene trees from McCormack et al. (2012).

As noted above, gross distortions of coalescence times are apparent by assaying gene tree conflicts at well-established, uncontroversial clades within Mammalia (Section 3.4). The inability to consistently recover clades with long stem lineages is a pattern that is widespread across the 183 gene trees (Fig. 3). For example, Cetartiodactyla (~11 MY stem lineage) was contradicted by 120 (66%) of the UCE gene trees. By contrast, this easy to resolve, traditional clade (Gatesy, 1998) was contradicted by only 4 of 26 genes (15%) sampled by Meredith et al. (2011) in their recent supermatrix analysis. In the concatalescence analysis of Song et al. (2012), a study that employed larger data partitions (~3100 bp), Cetartiodactyla was lacking in just 18% of the 447 gene trees (Supplementary Fig. 1). In McCormack et al. (2012), 50% of the individual gene trees contradicted Anthropoidea (~27 MY stem lineage), but none of the 26 gene trees from Meredith et al. (2011), and only 9% of the gene trees from Song et al. (2012) conflict at this node (Fig. 11). The much stronger congruence among gene trees in concatalescence (Song et al., 2012) relative to coalescence (McCormack et al., 2012) is striking and holds across all clades in the preferred tree of McCormack et al. (2012). On average McCormack et al.'s (2012) 183 gene trees based on short UCE loci exhibit 29% more conflict at each node relative to Song et al.'s 447 concatalescence trees (Supplementary Fig. 1). Within Primates, the contrast is even more striking; at each node within this clade, ~45% more of the gene trees conflict with the species tree in coalescence analysis relative to concatalescence analysis (Fig. 11). The concatenation of distantly linked exons in Song et al. (2012) yielded much higher congruence, consistent with early simulation work that indicated greater accuracy with an increase in locus length (Bull et al., 1993; Huelsenbeck and Hillis, 1993; Swofford et al., 2001).

For Anthropoidea, a clade with an ~27 MY stem lineage (Fig. 3), perfect congruence among gene trees (Meredith et al., 2011) or even 9% incongruence (Song et al., 2012) is much more reasonable than 50% conflict (McCormack et al., 2012) (Fig. 11). Instead of hypothesizing wholesale deep coalescences (Table 2), a much simpler explanation is that the short UCE loci of McCormack et al. (2012) are inadequate phylogenetic markers, and in isolation are unable to accurately reconstruct relationships among ancient mammalian lineages (Fig. 11; Supplementary Fig. 1). This alternative explanation is more biologically plausible than the retention of ancestral polymorphisms for >100 MY at 43% of the sampled loci (Table 2). For example, locus 3 (16 informative characters) groups the orangutan with a bird to the exclusion of other primates (Fig. 10). If just a few more homoplastic substitutions are accepted at this locus, however, traditional placement of the orangutan within Primates results – a much better explanation relative to retention of ancestral polymorphism across part of the Triassic, the Jurassic, the Cretaceous, and much of the Cenozoic (~300 MY). This example, and many others (Figs. 8–10), highlight the unrealistic assumptions of shortcut coalescence methods at deep phylogenetic scales.

3.9. Shortcut coalescence analysis of coalescence artifacts: making phylogenetic sausage"?

We suggest that the coalescence protocol utilized by McCormack et al. (2012) is the equivalent of making "phylogenetic sausage." In the messy construction of a sausage, a butcher grinds together chunks of undesirable cuts (e.g., lips, heart valves, assorted fatty bits, etc.), mixes this flotsam with spices, forcibly extrudes the concoction into a casing made of intestinal lining, and forms a product

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

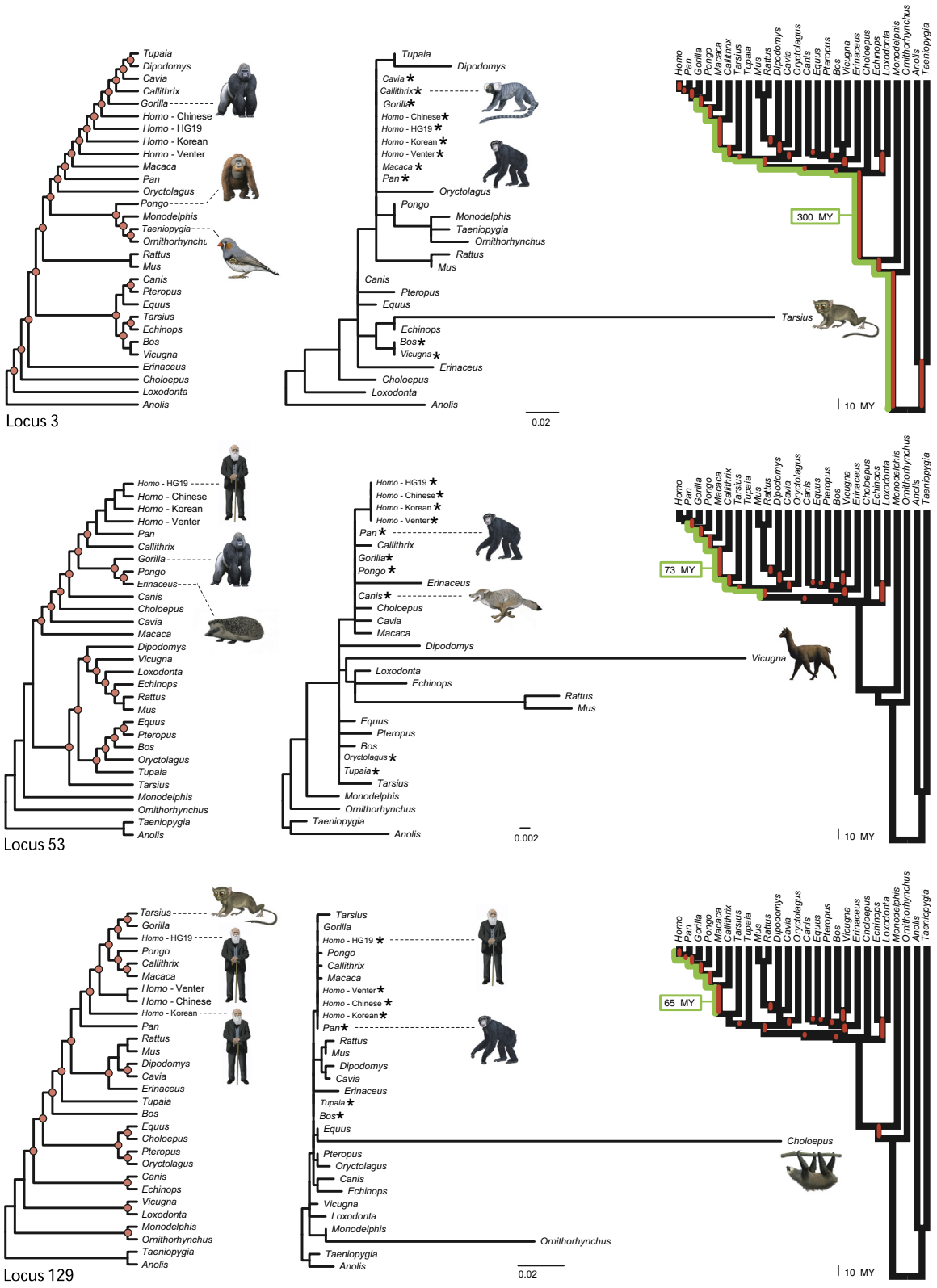

Fig. 10. Gene trees with exceptionally long branches from McCormack et al. (2012) that imply unrealistic deep coalescences. The fully resolved ML gene trees that were used in STAR coalescence analysis are shown on the left with nodes that conflict with the inferred species tree (Fig. 3) marked by red circles. Especially peculiar relationships that if true, demand exceptionally long retentions of ancestral polymorphism, are indicated. The same gene trees are shown in the middle column but with branch lengths (substitutions per site). Taxa with identical sequences are denoted with asterisks (*). The trees on the right represent the species tree based on STAR coalescence analysis of all 183 loci from McCormack et al. (2012) with branch lengths proportional to time and scale bar in millions of years (MY; see Fig. 3). For each locus, internodes in the species tree that conflict with the gene tree are colored red, and the longest deep coalescence is shown for each locus (green lineages). Although highly incongruent with the overall species tree, these gene trees are not outliers for the UCE dataset (McCormack et al., 2012). Within Placentalia, an average of only ~4.6 of 20 clades in the species tree (Fig. 3) were supported by each gene tree. Paintings are by C. Buell. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Table 2
Gene tree numbers, clades showing maximum deep coalescence, species tree dates, minimum gene tree dates implied by deep coalescence, and maximum coalescence values for 183 gene trees from McCormack et al. (2012). Dates are in millions of years. We used a divergence date of 0.2 million years for the most recent common ancestor of *Homo sapiens* (Underhill and Kivisild, 2007). All other divergence dates are from dos Reis et al. (2012).

Tree number	Clade with maximum deep coalescence	Species tree date	Minimum date on gene tree	Magnitude of deep coalescence
1	Hominoidea	17.4	89.1	71.7
2	Catarrhini	26.4	87.1	60.7
3	Hominoidea	17.4	317	299.6
4	Myomorpha	56	89.1	33.1
5	Theria	172.8	317	144.2
6	Myomorpha	56	87.1	31.1
7	Mammalia	184.6	317	132.4
8	Cetartiodactyla	61.4	89.1	27.7
9	Myomorpha	56	87.1	31.1
10	*Homo*	0.2	37.4	37.2
11	Cetartiodactyla	61.4	89.1	27.7
12	Theria	172.8	317	144.2
13	Myomorpha	56	89.1	33.1
14	Cetartiodactyla	61.4	184.6	123.2
15	Cetartiodactyla	61.4	89.1	27.7
16	Myomorpha	56	89.1	33.1
17	Theria	172.8	317	144.2
18	*Rattus* + *Mus*	14	56	42
19	Placentalia	89.1	317	227.9
20	African Hominoids	10.4	89.1	78.7
21	Myomorpha	56	89.1	33.1
22	African Hominoids	10.4	317	306.6
23	Myomorpha	56	89.1	33.1
24	Rodentia	61.3	89.1	27.8
25	Myomorpha	56	89.1	33.1
26	Theria	172.8	317	144.2
27	*Pan* + *Homo*	8.2	317	308.8
28	*Homo*	0.2	37.4	37.2
29	*Homo*	0.2	87.1	86.9
30	*Pan* + *Homo*	8.2	83.8	75.6
31	*Equus* + *Canis*	69.4	184.6	115.2
32	Theria	172.8	317	144.2
33	*Pan* + *Homo*	8.2	65	56.8
34	Rodentia	61.3	83.8	22.5
35	Myomorpha	56	317	261
36	Myomorpha	56	89.1	33.1
37	Mammalia	184.6	317	132.4
38	Mammalia	184.6	317	132.4
39	Mammalia	184.6	317	132.4
40	Anthropoidea	37.4	83.8	46.4
41	Theria	172.8	317	144.2
42	*Homo*	0.2	17.1	16.9
43	*Homo*	0.2	37.4	37.2
44	*Homo*	0.2	75.8	75.6
45	Myomorpha	56	89.1	33.1
46	Mammalia	184.6	317	132.4
47	Mammalia	184.6	317	132.4
48	Myomorpha	56	317	261
49	Rodentia	61.3	89.1	27.8
50	Catarrhini	26.4	37.4	11
51	Hominoidea	17.1	65	47.9
52	Catarrhini	26.4	89.1	62.7
53	African Hominoids	10.4	83.8	73.4
54	Myomorpha	56	89.1	33.1
55	*Rattus* + *Mus*	14	70.8	56.8
56	Myomorpha	54	184.6	130.6
57	Exafroplacentalia	87.1	317	229.9
58	Hominoidea	17.1	83.8	66.7
59	*Homo*	0.2	83.8	83.6
60	Afrotheria	70.4	89.1	18.7
61	African Hominoids	8.2	83.8	75.6
62	Mammalia	184.6	317	132.4
63	*Rattus* + *Mus*	14	87.1	73.1
64	Rodentia	61.3	89.1	27.8
65	Primates	65	317	252
66	Mammalia	184.6	317	132.4
67	Myomorpha	56	317	261
68	African Hominoids	10.4	75.8	65.4
69	*Homo*	0.2	317	316.8
70	Anthropoidea	37.4	317	279.6
71	Myomorpha	56	184.6	128.6
72	Myomorpha	56	184.6	128.6

(*continued on next page*)

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Table 2 (*continued*)

Tree number	Clade with maximum deep coalescence	Species tree date	Minimum date on gene tree	Magnitude of deep coalescence
73	Myomorpha	56	89.1	33.1
74	*Homo*	0.2	317	316.8
75	Mammalia	184.6	317	132.4
76	*Homo*	0.2	89.2	89
77	*Homo*	0.2	83.8	83.6
78	Myomorpha	56	83.8	27.8
79	Myomorpha	56	83.8	27.8
80	Myomorpha	56	89.1	33.1
81	Theria	172.8	317	144.2
82	Cetartiodactyla	61.4	172.8	111.4
83	*Homo*	0.2	37.4	37.2
84	Mammalia	184.6	317	132.4
85	Mammalia	184.6	317	132.4
86	*Homo*	0.2	65	64.8
87	Mammalia	184.6	317	132.4
88	*Equus* + *Canis*	69.9	89.1	19.2
89	Myomorpha	56	83.8	27.8
90	Primates + Scandentia	74.1	172.8	98.7
91	African Hominoids	10.4	83.8	73.4
92	African Hominoids	10.4	89.1	78.7
93	*Homo*	0.2	89.1	88.9
94	Myomorpha	56	89.1	33.1
95	Anthropoidea	37.4	75.8	38.4
96	Myomorpha	56	184.6	128.6
97	Mammalia	184.6	317	132.4
98	Mammalia	184.6	317	132.4
99	*Rattus* + *Mus*	14	83.8	69.8
100	*Homo*	0.2	37.4	37.2
101	Mammalia	184.6	317	132.4
102	Catarrhini	26.4	89.1	62.7
103	*Pan* + *Homo*	8.2	172.8	164.6
104	Mammalia	184.6	317	132.4
105	Catarrhini	26.4	83.8	57.4
106	*Equus* + *Canis*	69.9	89.1	19.2
107	Myomorpha	56	184.6	128.6
108	Theria	172.8	317	144.2
109	*Rattus* + *Mus*	14	184.6	170.6
110	Primates	65	172.8	107.8
111	Exafroplacentalia	87.1	317	229.9
112	*Homo*	0.2	89.1	88.9
113	Cetartiodactyla	61.4	89.1	27.7
114	*Homo*	0.2	83.8	83.6
115	Anthropoidea	37.4	75.8	38.4
116	Theria	172.8	317	144.2
117	Mammalia	184.6	317	132.4
118	*Homo*	0.2	75.8	75.6
119	Glires	70.8	83.8	13
120	Myomorpha	56	89.1	33.1
121	Myomorpha	56	89.1	33.1
122	*Pan* + *Homo*	8.2	65	56.8
123	*Homo*	0.2	87.1	86.9
124	*Pan* + *Homo*	8.2	89.1	80.9
125	Cetartiodactyla	61.4	317	255.6
126	Myomorpha	56	89.1	33.1
127	*Homo*	0.2	184.6	184.4
128	*Homo*	0.2	37.4	37.2
129	*Homo*	0.2	65	64.8
130	African Hominoids	10.4	83.8	73.4
131	Mammalia	184.6	317	132.4
132	Myomorpha	56	83.8	27.8
133	Anthropoidea	37.4	317	279.6
134	*Pan* + *Homo*	8.2	75.8	67.6
135	*Equus* + *Canis*	69.9	184.6	114.7
136	Myomorpha	56	172.8	116.8
137	Catarrhini	26.4	83.8	57.4
138	Rodentia	61.3	317	255.7
139	Theria	172.8	317	144.2
140	Theria	172.8	317	144.2
141	Laurasiatheria	76	317	241
142	Theria	172.8	317	144.2
143	Mammalia	184.6	317	132.4
144	Catarrhini	26.4	65	38.6
145	Myomorpha	56	89.1	33.1
146	Primates + Scandentia	74.1	89.1	15
147	*Homo*	0.2	83.8	83.6

Table 2 (*continued*)

Tree number	Clade with maximum deep coalescence	Species tree date	Minimum date on gene tree	Magnitude of deep coalescence
148	Myomorpha	56	83.8	27.8
149	*Homo*	0.2	17.1	16.9
150	Myomorpha	56	75.8	19.8
151	*Rattus* + *Mus*	14	317	303
152	*Homo*	0.2	37.4	37.2
153	Theria	172.8	317	144.2
154	African Hominoids	10.4	65	54.6
155	Theria	172.8	317	144.2
156	Hominoidea	17.1	87.1	70
157	Hominoidea	17.1	65	47.9
158	Theria	172.8	317	144.2
159	Theria	172.8	317	144.2
160	Rodentia	61.3	89.1	27.8
161	Myomorpha	56	89.1	33.1
162	Theria	172.8	317	144.2
163	*Homo*	0.2	89.1	88.9
164	Myomorpha	56	89.1	33.1
165	Myomorpha	56	89.1	33.1
166	Anthropoidea	37.4	89.1	51.7
167	Myomorpha	56	317	261
168	African Hominoids	10.4	75.8	65.4
169	Primates	65	317	252
170	Theria	172.8	317	144.2
171	Primates	65	89.1	24.1
172	Laurasiatheria	76	89.1	13.1
173	*Homo*	0.2	317	316.8
174	Mammalia	184.6	317	132.4
175	Catarrhini	26.4	83.8	57.4
176	Myomorpha	56	184.6	128.6
177	Mammalia	184.6	317	132.4
178	*Homo*	0.2	74.1	73.9
179	Cetartiodactyla	61.4	89.1	27.7
180	Theria	172.8	317	144.2
181	Mammalia	184.6	317	132.4
182	*Pan* + *Homo*	8.2	37.4	29.2
183	*Homo*	0.2	172.8	172.6
Mean				101.2
Median				78.7

of passable palatability. Likewise, in the first step of the shortcut coalescence procedure, undesirable gene trees were reconstructed from small, uninformative bits of the genome (Fig. 7; Table 1) that in some cases have wildly varying rates of evolution in different mammalian lineages (Fig. 10). The arbitrarily resolved, inaccurate gene trees were then forced into a final species tree casing under the assumption that all gene tree conflicts are due to deep coalescence. Utilizing this crude and messy process, a phylogenetic hypothesis for placental mammals that is fairly congruent with supermatrix results at easy to resolve nodes was produced (Fig. 3). However, it is difficult to argue that this tree, which is based on gene trees with impossibly deep coalescences (Figs. 8–10; Table 2), is a prime steak; i.e., that the inferred species tree is accurate at difficult to resolve polytomies (Fig. 3) and represents a significant improvement relative to previous supermatrix studies (e.g., Meredith et al., 2011).

In fact, reanalysis of the 183 ML gene trees from McCormack et al. (2012) using different shortcut coalescence methods gave conflicting results. Our STAR run yielded the STAR tree topology reported by McCormack et al. (2012), but analysis with MP-EST, the shortcut coalescence method utilized by Song et al. (2012), disagreed with the STAR topology at three of the 20 nodes (15%) within Placentalia (Figs. 3 and 12). Glires monophyly was not supported, relationships within Rodentia were rearranged, and the basal rooting of placental mammals shifted. Instead of rooting on Afrotheria, as in the STAR analysis of McCormack et al. (2012), the MP-EST analysis rooted on Xenarthra as in morphological cladistic analyses (O'Leary et al., 2013). Important conflicts between STAR and MP-EST also have been reported in a recent reanalysis of land plant origins (Springer and Gatesy, 2014), an even deeper phylogenetic question (500–900 MY divergences) that was thought to be resolved with consistent support from MP-EST analyses (Zhong et al., 2013). STAR analyses of the same phylogenomic datasets for green plants gave contradictory results, sometimes with higher bootstrap support (Springer and Gatesy, 2014). This conflict among methods has been blamed on the inability of STAR, a distance based procedure, to deal with missing data (Zhong et al., 2014), but "a lot of missing lineages may dramatically reduce the performance of the pseudo-likelihood approach [MP-EST]" as well (Liu et al., 2010; pp. 4–5). For the mammalian UCEs, incongruence between shortcut coalescence methods (Figs. 3 and 12) cannot be blamed on missing data.

STAR and MP-EST are thought to be statistically consistent methods when gene trees are reconstructed accurately (Liu et al., 2009b, 2010). For the placental dataset of 183 loci in which gene trees are highly inaccurate (Figs 8–10; and Table 2), the two coalescence methods disagree, and it is not clear which topology should be preferred. This result challenges the assertion that coalescence analysis of UCE loci robustly resolved the controversial basal split in placental mammals – the primary conclusion of McCormack et al. (2012). The preferred outcome of McCormack et al. (2012) using STAR is not even supported by re-analysis of their own data using MP-EST, and both results contradict the rooting position solidly supported by the concatalescence analysis of Song et al. (2012). Concatenation studies have documented similar patterns of conflict at this polytomy (Murphy et al., 2001a,b, 2007; Nikolaev et al., 2007; Wildman et al., 2007; Meredith et al., 2011; Morgan et al., 2013; Romiguier et al., 2013), and rare genomic changes (Nishihara et al., 2009) are nearly evenly divided among the three possible resolutions (Xenarthra versus Epitheria root = 25

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

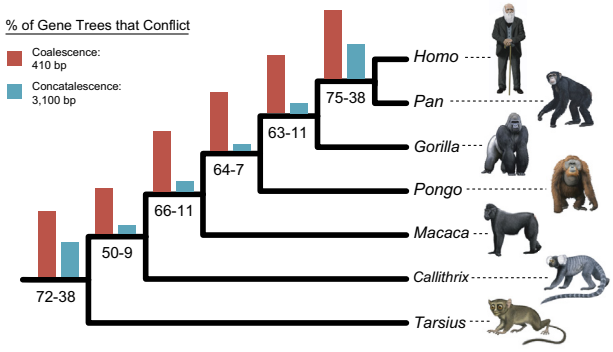

Fig. 11. The percentage of gene trees that conflict at particular primate nodes. Gene tree conflicts in coalescence analysis of ~410 bp data partitions (red bars; McCormack et al., 2012) versus gene tree conflicts in concatalescence analysis of ~3100 bp data partitions (blue bars; Song et al., 2012) are shown for primate clades supported by both studies. The percentage of gene trees that contradict a particular clade is shown below each internode (left = McCormack et al. (2012) and right = - Song et al. (2012)). For example, 50% of the gene trees from McCormack et al. (2012) conflict with Anthropoidea, while only 9% of the gene trees from Song et al. (2012) did not support this clade. On average, ~45% fewer of the gene trees conflict at a given node in concatalescence analysis compared to coalescence analysis, likely due to the fact that data were concatenated in the former. The protein-coding loci from Song et al. (2012) averaged 3100 bp, but the actual loci span up to >1,000,000 bp in mammalian genomes, because introns were trimmed away before analysis (Fig. 7). Thus, the concatalescence loci likely correspond to summations of many independent c-genes, each with different branching histories. Paintings 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.)

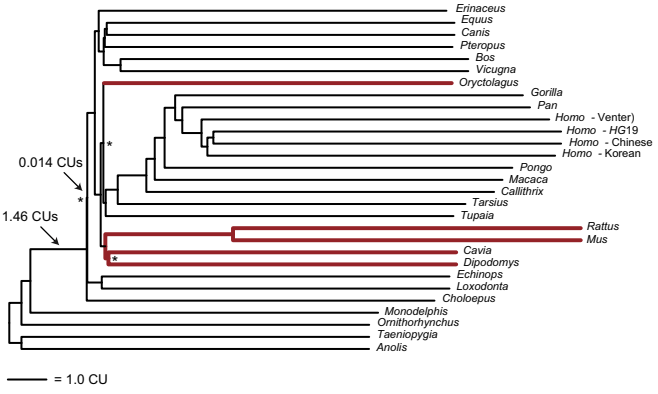

Fig. 12. Stunted MP-EST "bonsai tree" with branch lengths in coalescent units (CUs) based on the 183-locus UCE dataset. All internodes in the tree are extremely short. The top arrow points to the miniscule stem branch to Epitheria (all placentals except Xenarthra) that is barely perceptible at this scale; the lower arrow points to the compressed stem lineage of Placentalia. Asterisks (*) are at nodes in the MP-EST tree that conflict with STAR analysis of the same UCE dataset (also see Fig. 3). Red branches indicate members of Glires (rodents + rabbits), a traditionally recognized clade that is not supported by the MP-EST analysis. Terminal branch lengths are arbitrarily drawn to nine CUs. Scale bar corresponds to one CU. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

transposon insertions; Afrotheria versus Exafroplacentalia root = 22 transposon insertions; Atlantogenata versus Boreoeutheria root = 21 transposon insertions). In sum, these results reiterate the conclusion that the basal placental split (Fig. 3) remains unresolved, whether coalescence or concatenation is employed (see discussion in Meredith et al., 2011).

The inability of shortcut coalescence methods to discern traditional mammalian clades such as Glires based on phylogenomic datasets is more disconcerting (Figs. 3 and 12). Glires monophyly is delimited by multiple anatomical synapomorphies (Meng et al., 2003; O'Leary et al., 2013) and has garnered solid support even from relatively short concatenations of less than 30 genes (e.g., Madsen et al., 2001; Murphy et al., 2001a,b; Meredith et al., 2011). Furthermore, two unique insertions of transposons support this clade (Nishihara et al., 2006). Yet, MP-EST coalescence analysis of 183 UCE loci contradicts Glires, a clade with an ~5 MY stem lineage (Fig. 3; dos Reis et al., 2012). The inability of coalescence to recover uncontroversial clades such as Glires based on 74,750 bp is surprising, and from our perspective, it is awkward to then argue that shortcut coalescence analysis of the same misleading gene trees (Figs. 8–10) can somehow solve truly difficult problems in mammalian systematics (e.g., the placental root). Our exploration of hidden support revealed numerous, additional cases where traditionally recognized clades with long stem lineages were not recovered by shortcut coalescence analyses of >100 loci. By contrast, ML supermatrix analyses of these same loci recovered each of these clades with solid bootstrap support (Fig. 3).

3.10. Phylogenetic "bonsai trees"

For the UCE loci, inaccurately reconstructed gene trees imply highly improbable deep coalescences (Figs. 8–11; and Table 2). A repercussion of these errors is a species tree with branch lengths, in coalescent units, that are impossibly short (Springer and Gatesy, 2014). Patel et al. (2013) suggested that an N_e of 200,000 diploid individuals and a generation time of one year are reasonable approximations for many vertebrates (1 coalescent unit = 400,000 years). Given these values, the MP-EST species tree based on UCEs is stunted (Fig. 12). The only internal branch within crown Placentalia that exceeds 400,000 years is the stem lineage to *Mus* (mouse) plus *Rattus* (rat). In striking contrast to previous estimates, the stem lineage to Placentalia is only 584,000 years (1.46 coalescent units). Dates based on first fossil occurrences (O'Leary et al., 2013) suggest that the stem placental branch represents at least 65 MY of time, and this increases to 95 MY if *Juramaia* is recognized as the oldest crown therian (Luo et al., 2011). Similarly, dates based on relaxed molecular clocks (Meredith et al., 2011; dos Reis et al., 2012) imply a long branch length of 83.7–88.6 MY for this basal lineage (Fig. 3). Thus, relative to fossil and relaxed molecular clock dates, the MP-EST branch length for stem Placentalia is more than two orders of magnitude too small, and the entire MP-EST tree is severely compressed (Fig. 12).

The MP-EST species tree derived from UCEs (McCormack et al., 2012) is a phylogenetic "bonsai tree" with dwarfed branch lengths (Fig. 12), a phenomenon previously noted in phylogenetic analyses of green plants (Springer and Gatesy, 2014). As for the MP-EST analysis of plants (Zhong et al., 2013), in which one gene tree implied a deep coalescence >500 MY (Springer and Gatesy, 2014), inaccurate reconstruction of gene trees (Figs. 8–11) provides a more plausible explanation for the MP-EST bonsai effect than incomplete lineage sorting (Springer and Gatesy, 2014). Retention of ancestral polymorphism is not a viable hypothesis for gene trees that support a "bird-dog"clade (*Taeniopygia* + *Canis*, locus 22) and a "beaked tetrapod" clade that clusters zebra finch with duck-billed platypus (*Taeniopygia* + *Ornithorhynchus*, locus 27). Gene trees based on these loci imply deep coalescences >300 MY (Fig. 8).

MP-EST bonsai trees with artificially shrunken branch lengths should not be used as the basis for simulations that seek to discriminate between gene tree reconstruction errors and incomplete lineage sorting (Song et al., 2012; Zhong et al., 2013). When the underlying MP-EST tree, the input for the simulations, has conflated these sources of phylogenetic inaccuracy from the start, extensive conflicts among simulated gene trees cannot be used to justify application of coalescence methods (Song et al., 2012; Zhong et al., 2013, 2014), especially when gene tree reconstruction errors are ignored in the circular simulation routine (Springer and Gatesy, 2014).

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

3.11. Inaccuracy of shortcut coalescence in simulations relative to concatenation

At first glance, early simulation work appears inconsistent with our point of view and uniformly favors shortcut coalescence methods over concatenation (Kubatko et al., 2009; Liu et al., 2009b, 2010). The coalescence approach has been championed based on its success in the "anomaly zone." In this area of the overall tree space, at least two very short internodes are connected (Rosenberg, 2013), and assuming complete neutrality, the most likely gene tree does not match the species tree (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007). Under such conditions, simulations suggest that concatenation fails whereas several shortcut coalescence methods – STAR, STEAC, MP-EST – succeed (Liu et al., 2009b, 2010). In combination with another simulation study in which STEM outperformed concatenation (Kubatko et al., 2009), this early work was interpreted by some as a death knell for the supermatrix approach at all levels of divergence (McCormack et al., 2012; Song et al., 2012; Kumar et al., 2013; Zhong et al., 2014). Upon closer examination, however, it is apparent that this first set of simulations focused on a tiny fraction of all possible topologies. Only small (4–5 taxa), shallow, pectinate species trees were examined in each simulation (Fig. 13; Appendix B). Thus, it would be premature to assume that such restricted results are general, especially with reference to deep radiations in the Tree of Life (e.g., Fig. 3).

Subsequent simulations across a broader spectrum of parameter space strongly support the opposite result. Concatenation was more accurate than shortcut coalescence across nearly all of the parameter combinations that were assessed (Leaché and Rannala, 2011; Aguiar and Schrago, 2013; Bayzid and Warnow, 2013; Patel et al., 2013). The range of topologies in this more recent work dwarfed the set of species trees tested in the initial simulations (Fig. 13; Appendix B), but no evidence for the general superiority of shortcut coalescence was revealed. Leaché and Rannala (2011) compared simple concatenation approaches (Bayesian, parsimony) to the shortcut coalescence method STEM (5 taxa), and concatenation solidly outperformed STEM in nearly all comparisons, even though the data were simulated using coalescence models. Patel et al. (2013) obtained similar results in a second comparison between STEM and concatenation (20 taxa). Specifically, the supermatrix method was generally more accurate than STEM, and this discrepancy between the two approaches was more evident at greater divergences and with more loci. Bayzid and Warnow (2013) again analyzed DNA sequences simulated using neutral coalescence models (11–17 taxa), and for a third time, simple concatenation dominated shortcut coalescence methods (MP-EST, MDC), in this case, across every parameter set that was examined. Bayzid and Warnow (2013) is the first simulation study that also assessed a concatalescence approach. Although concatalescence fared better than coalescence, ML concatenation was more accurate than both of these methods. In half of the trials, simple ML concatenation even outperformed *BEAST, a parameter-rich coalescence method that co-estimates gene and species trees (Bayzid and Warnow, 2013). Aguiar and Schrago (2013) simulated data according to a concatenation model with no heterogeneity in gene tree topology and at very deep evolutionary scales (64 taxa). For a symmetrical species tree, all methods were accurate. In the remaining trials, however, supermatrix analysis out-competed shortcut coalescence (STAR and STEAC) in 75% of the parameter combinations tested. For a pectinate tree with high rate variation among genes, STEAC showed 0% accuracy, STAR yielded ~5% accuracy, whereas the supermatrix approach gave ~70% accuracy (Aguiar and Schrago, 2013).

A key insight from this second batch of simulation studies is that even when data were generated using a coalescence model that incorporates all major assumptions of the coalescence methods tested, simple concatenation was more accurate in the large majority of cases that were examined (Leaché and Rannala, 2011; Bayzid and Warnow, 2013; Patel et al., 2013). In Leaché and Rannala (2011), the simulation model assumed no selection, no recombination, correct delimitation of c-genes a priori, no gene flow post-speciation, true population size known a priori, and a molecular clock, all of which should have favored STEM. Yet, simple Bayesian concatenation out-competed coalescence across nearly all of the parameter combinations tested, and equally weighted parsimony also solidly outperformed STEM in most of these trials (Leaché and Rannala, 2011). STEM was less accurate than supermatrix methods even though *many* times more model parameters are required to generate a species tree with STEM (also see Patel et al., 2013). For example, in an analysis of 100 loci, only one substitution model plus one topology with branch lengths are required for simple ML concatenation. By contrast, 100 estimated gene tree topologies, each with estimated branch lengths, are necessary input for a STEM analysis, which also requires relative rate parameters for 100 loci and a population size parameter to then finally estimate a species tree topology.

Our survey of relevant simulations that have directly measured the performance of concatenation versus shortcut coalescence shows that shortcut coalescence has so far out-competed concatenation in only a small portion of parameter space – pectinate, shallow trees with sparse taxon sampling (4–5 taxa), most of which are in or near the "anomaly zone." In two of the simulations (Liu et al., 2009b, 2010), the same five-taxon, pectinate tree in the anomaly zone was modeled; in terms of tree depth, these two simulations extended to only 3.1 coalescent units (Fig. 13; Appendix B), the equivalent of only a few million years, at most, for typical mammalian species (Appendix A). In the only other simulation study where shortcut coalescence bested concatenation (Kubatko et al., 2009), four species were simulated in a similarly shallow pectinate tree (2.2–3.0 coalescent units), and supermatrix trees were midpoint rooted (Fig. 13; Appendix B). Although midpoint rooting is necessary for some shortcut coalescence methods, such as STEM, that require input of ultrametric gene trees, this procedure is rarely used by molecular systematists, and we agree with Hillis (1987; p. 33) that "the use of midpoint rooting algorithms should be consciously avoided".

For nearly every other tree shape and depth that has been tested, the supermatrix approach has dominated shortcut coalescence methods (STEM, MP-EST, STAR, STEAC, MDC) in head-to-head comparisons (Appendix B). Thus, the repeated assertion that coalescence outperforms concatenation in simulation (e.g., McCormack et al., 2012; Song et al., 2012; Zhong et al., 2013, 2014) is untrue for shortcut coalescence methods that can actually be applied to large phylogenetic datasets. A broader range of parameter combinations must be explored to arrive at any general conclusions regarding the performance of shortcut coalescence versus concatenation at deep divergences. However, c-gene size can only ratchet downward with increased taxon sampling, whereas the number of internodes that needs to be resolved in each gene tree ratchets ever upward with increased taxon sampling (Fig. 6). This is a debilitating biological and analytical reality for shortcut coalescence methods. We therefore predict that current methods will perform even more poorly in simulations as the number of taxa more closely approximates sampling seen in modern empirical systematic studies and as more realistic c-gene sizes are simulated (see Section 3.12).

To this point, simulations that have focused on ancient divergences in large clades have not favored shortcut coalescence over concatenation, likely due to the challenge of deriving accurate gene trees at this depth with limited sequence data (Leaché and Rannala, 2011). Patel et al. (2013) found that when true gene trees,

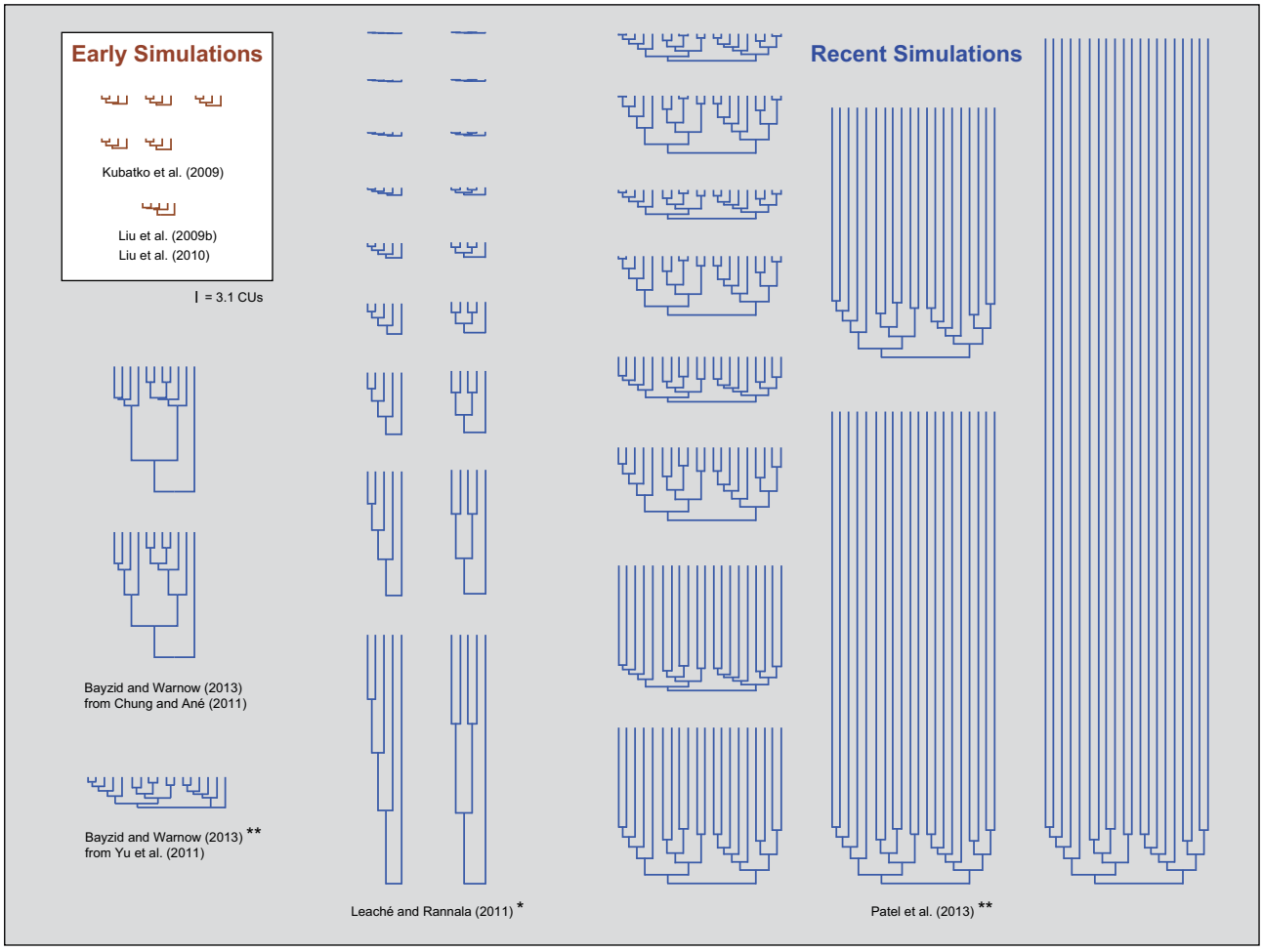

Fig. 13. In simulations, shortcut coalescence methods have out-competed concatenation over a tiny region of the overall species tree space. Species tree shapes and depths are shown from six simulation studies. In each, simple concatenation and shortcut coalescence methods were directly compared, head-to-head, and gene trees/DNA sequences were generated via neutral coalescence models. The three early simulation studies in the upper left (red species trees in white box) are those where shortcut coalescence methods (STEM, STAR, STEAC, MP-EST) yielded more accurate results, but all species trees are shallow, pectinate, and include just four or five taxa. Six combinations of topology, depth, and relative branch length were tested (Appendix B), and these are all shown. The three more recent simulation studies (blue species trees on gray background) compared shortcut coalescence to concatenation over a much broader range of species tree depths, topologies, and branch lengths, with >5 species sampled in two studies. This work nearly uniformly showed better accuracy for concatenation relative to shortcut coalescence methods (MP-EST, MDC, STEM) at deep timescales (Appendix B). For these more recent simulations, one asterisk (*) indicates that a wide range of species tree branch lengths was assessed for each topology/depth. Two asterisks (**) indicate that multiple species tree topologies and multiple sets of relative branch lengths were assessed for the species tree depth that is shown. Scale bar in coalescent units (CUs) is given (upper left), and corresponds to the total depth of the single, shallow species tree examined in Liu et al. (2009b) and also in Liu et al. (2010). Another study where shortcut coalescence (STAR, STEAC) was directly compared to concatenation (Aguiar and Schrago, 2013) is not shown, because data were simulated via a concatenation model (no heterogeneity among gene tree topologies). This work simulated trees with 64 taxa that correspond to 500 MY of evolution, and concatenation was more accurate than shortcut coalescence for the majority of parameter sets tested (Appendix B). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the gene trees directly generated via simulation, are utilized, the shortcut coalescence method STEM outperformed concatenation, but true gene trees are unknown outside of simulations and instead gene trees must be inferred from sequences. When gene trees were estimated via ML analysis of nucleotide sequences (as in Patel et al., 2013), STEM was nearly uniformly outperformed by concatenation. The simulations suggest that supermatrix analysis does better than shortcut coalescence methods at deep nodes in the Tree of Life, because inaccurately reconstructed gene trees are more problematic for coalescence methods than incomplete lineage sorting is for supermatrix analysis (Bayzid and Warnow, 2013; Patel et al., 2013; Springer and Gatesy, 2014). Thus, recent simulation work corroborates the patterns observed in our re-analysis of phylogenomic data from Mammalia.

Although the impact of these simulation results has been questioned (Zhong et al., 2014), to our knowledge, no study has shown the opposite pattern – higher accuracy at depth for shortcut coalescence methods relative to concatenation in head-to-head comparisons. Zhong et al. (2014) instead cited simulation studies where concatenation fails but was not directly compared to shortcut coalescence methods at depth (e.g., Kubatko and Degnan, 2007; Liu and Edwards, 2009), so it is unclear whether shortcut coalescence would have fared any better than concatenation in these trials. This hypothesis was not tested, and previously published simulations suggest otherwise (Fig. 13; Appendix B).

Several recent simulation studies have concluded that shortcut coalescence methods, in particular STEM, are robust to recombination (Lanier and Knowles, 2012) or extensive missing data (Hovmöller et al., 2013) and that using more informative loci can save STEM from failure (Lanier et al., 2014). Contrary to some interpretations (Wu et al., 2013; Zhong et al., 2014; Lanier et al., 2014), however, this work is not relevant to the shortcut coalescence versus concatenation debate. Shortcut coalescence and concatenation were not directly compared in these papers; instead, shortcut

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

coalescence was compared to itself (e.g., Lanier and Knowles, 2012; Hovmöller et al., 2013; DeGiorgio and Degnan, 2014; Lanier et al., 2014). When STEM has been compared directly to simple concatenation at depth, the supermatrix approach has been consistently dominant (Fig. 13; Appendix B).

For these simulations (Fig. 13) and for our analyses of emergent support (Figs. 3–5), why does concatenation, a method that does not account for the conflicting evolutionary histories of different genes and is considered naive (e.g., Edwards, 2009; Knowles, 2010; McVay and Carstens, 2013; Lanier et al., 2014), detect phylogenetic signals that are bypassed by shortcut coalescence methods designed *specifically* to accommodate gene tree heterogeneity (Liu et al., 2009a)? Ironically, the answer seems to be that shortcut coalescence methods do *not* properly accommodate gene tree heterogeneity at deep phylogenetic divergences, and have inherent flaws that override any advantage of attempting to account for retention of ancestral polymorphism across subsequent nodes in a tree. At deep divergences, the end result is more accurate representations of phylogenetic history for methods that completely ignore reconstructions of gene trees, at least for the simulations that have been completed thus far (Fig. 13).

3.12. The coalescence/concatalescence conundrum

The prerequisite of accurate gene trees in shortcut methods currently presents a critical conundrum for the coalescence approach to systematics at deep phylogenetic scales (Fig. 14). Empirical work suggests that for large, ancient, speciose clades, c-genes will be quite short. For example, in two recent phylogenomic analyses of primates (Hobolth et al., 2007, 2011), conflicting segments of the genome that do not match the inferred species tree were predominately <100 base pairs in length (Fig. 7); segments of the genome that matched the species tree were longer. These inferences are based on analyses of large genomic segments using Bayesian methods that simultaneously account for mutation, deep coalescence, and recombination (e.g., Dutheil et al., 2009). The primate data examined by Hobolth et al. (2007, 2011) are especially relevant to our analysis, because the taxa in these studies (*Homo*, *Pan*, *Gorilla*, *Pongo*, *Macaca*) were also included in McCormack et al.'s (2012) dataset. Given that conflicting regions of the genome averaged <100 bp in comparisons of only four species (Hobolth et al., 2007, 2011), the same conflicting c-genes could be no longer in a larger dataset that includes 24 mammalian species (Fig. 3), and presumably would be even shorter. These petite c-genes, 75% between 17 and 93 bp (Hobolth et al., 2011), are much closer in size to the basic unit of analysis in supermatrix studies, a single nucleotide position, than to the larger data partitions (410–1,000,000 bp) utilized in recent coalescence/concatalescence studies of Mammalia (Hallström et al., 2011; McCormack et al., 2012; Song et al., 2012; Kumar et al., 2013) (Fig. 7).

Topological incongruence is not the only source of disagreements among gene trees as "there can be significant and detectable heterogeneity in branch lengths, such that the gene trees are for practical purposes still heterogeneous" (Edwards, 2009; p. 3) even though they are topologically identical. In this context, branch length heterogeneity refers to branch lengths that exhibit gene-to-gene variation in coalescence times without deep coalescence that yields topological conflicts among gene trees (Ané, 2010). Branch length heterogeneity across genes is a natural consequence of recombination and may be the most common cause of gene tree heterogeneity. Unlike deep coalescence, branch length heterogeneity will occur even when the underlying species tree has branch lengths that are long in terms of coalescent units (Edwards, 2009). Thus, c-genes may be much smaller than implied by Hobolth et al.'s (2011) results that only accounted for deep coalescence.

In a supermatrix analysis, each nucleotide substitution is treated as an independent piece of phylogenetic evidence that is used to assess the relationships among species (Kluge, 1989; Nixon and Carpenter, 1996; de Queiroz and Gatesy, 2007). At the molecular level, each mutation and fixation of that mutant is considered the basis for systematic conclusions. In this more traditional interpretation, character independence in concatenation is likely overestimated because linkage of different sites is not taken into account (Fig. 14). However, depending on the pattern of recombination at a locus, as well as the temporal sequence of mutations and the fixation of those mutants, nucleotide substitutions in the same gene and on the same phylogenetic branch might be completely independent from each other in terms of both mutation and fixation. For example, on a very long branch, such as the stem lineage to Placentalia (84 MY; Fig. 3), a mutation and the subsequent fixation of that mutant very early, say 165 MY ago, might be considered nearly completely independent from a second mutation + fixation in the same gene and on the same branch at 100 MY. These events are separated by 65 MY. In this case, coalescence methods would underestimate the independence of evolutionary events, as well as the character support, by distilling all substitutions at this locus down to a single gene tree topology. This undercount of independent phylogenetic evidence is greatly amplified when a researcher assumes that c-genes are exceptionally long (e.g., 1,000,000 bp), when in reality c-genes are much shorter (e.g., <100 bp; Fig. 7). Many independent c-genes would be counted as only one, and this might greatly underestimate support for a particular clade. In combination with wholesale gene tree reconstruction errors (Figs. 8–10), the underestimation of character independence in coalescence analysis could explain the inability of coalescence methods to detect phylogenetic signals in the mammalian UCE data, whereas supermatrix analysis excels (Fig. 3). These properties also might explain, in part, the much weaker bootstrap support in shortcut coalescence analyses of mammals (Song et al., 2012) in comparison to supermatrix analyses of the same database (Meredith et al., 2011; also see Gatesy and Springer, 2013).

When speciation events are deep in time and tightly spaced, accurate reconstruction of gene trees cannot be expected for tiny segments of DNA (Figs. 8–10). An obvious response to short and uninformative c-genes (e.g., Table 1) is to concatenate, but this entails abandonment of the most basic rationale for doing a coalescence analysis in the first place (Fig. 6) and is an admission that the supermatrix approach is necessary for the coalescence approach to succeed – i.e., concatalescence (Gatesy and Springer, 2013). At deep divergences within Mammalia, concatalescence analyses (Hallström et al., 2011; Song et al., 2012; Kumar et al., 2013) have so far outnumbered honest attempts at coalescence analysis, such as McCormack et al. (2012). For ancient divergences, analysis of complete protein-coding regions has been standard practice because: (1) most introns have evolved rapidly and are difficult to align across deep mammalian clades plus outgroups, and (2) some systematists have wrongly focused on the gene (Hallström et al., 2011; Song et al., 2012; Kumar et al., 2013), not the more appropriate c-gene (Doyle, 1992), as the basic unit of coalescence analysis.

Sequence conservation due to strong purifying selection permits unambiguous sequence alignment among distantly related taxa and limits saturation due to multiple overlapping substitutions, but provides little evidence for reconstructing a reliable gene tree. In mammals and many other taxa, exons commonly are quite short and separated by long, highly divergent introns (O'Brien et al., 1999; Lindblad-Toh et al., 2011). The logical suggestion of Ané et al. (2007; p. 414) to "treat each exon as a different gene" in cases where recombination is suspected generally has been ignored in coalescence analyses at deep phylogenetic scales (but

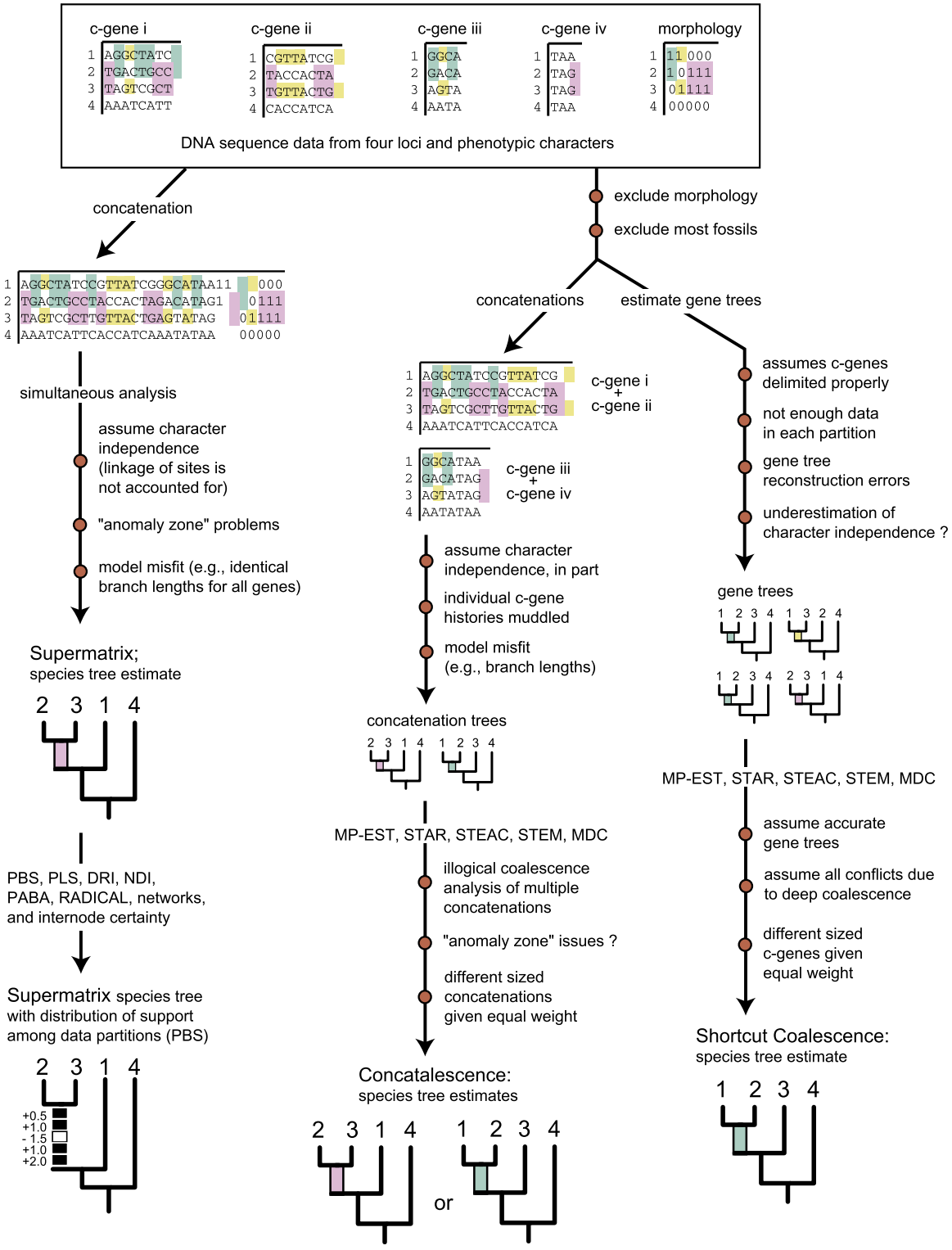

Fig. 14. A flow chart that compares the concatenation, concatalescence, and shortcut coalescence approaches to reconstructing species trees. Four c-genes (i–iv) and morphological character codes from four taxa (1–4) are shown. Some problems encountered with analysis of these data in each paradigm are marked by red circles. For simplicity, phylogenetic analyses of character data (supermatrix plus individual c-genes) are based on equally weighted parsimony. Parsimony informative characters that support 1 + 2 (green), 1 + 3 (yellow), and 2 + 3 (pink) are highlighted, and the same colors mark these clades in gene trees and species trees. Note that phenotypic evidence is excluded from the start in currently implemented shortcut coalescence and concatalescence procedures. Multiple methods have been developed over the last 15 years for assessing the distribution of support among genes in supermatrix analysis (PBS – Baker and DeSalle, 1997; NDI and DRI – Gatesy et al., 1999; PLS – Lee and Hugall, 2003; Gatesy and Baker, 2005; networks – Huson and Bryant, 2006; PABA – Struck et al., 2006; RADICAL – Narechania et al., 2012; internode certainty – Salichos and Rokas, 2013); these measures indicate whether the lion's share of support for a particular clade is restricted to a few loci or is more evenly distributed among many loci. For the contrived dataset in the figure, PBS (partitioned branch support) is shown for the five data partitions in the supermatrix (bottom left; black = positive scores, white = negative score). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

see Tsagkogeorga et al., 2013). Instead, to assemble informative loci for phylogenomic coalescence analysis, researchers have stripped away long intervening introns and concatenated short, conserved exons from any given gene, often over broad spans of the genome, and then applied coalescence methods (Hallström et al., 2011; Song et al., 2012; Kumar et al., 2013; Xi et al., 2013; Zhong et al., 2013).

If proponents of coalescence are willing to merge nucleotide positions that are >1,000,000 bp apart into a single "c-gene" (Fig. 7; Song et al., 2012), why not concatenate sites that are 10,000,000 or 100,000,000 bp apart? Why not concatenate all nucleotide positions in the genome and just get it over with? Given that so much has been made of the benefits of coalescence and the analysis of c-genes that have not recombined over the history of a clade (Maddison, 1997; Slowinski and Page, 1999; Edwards et al., 2009), analysis of 139,600 bp "pseudo c-genes" (Song et al., 2012) is difficult to justify. At least for the mammalian taxa analyzed here, these concatenated loci would surely include multiple, conflicting recombination units (Hobolth et al., 2007, 2011; Dutheil et al., 2009), especially in view of Edwards' (2009) assessment that gene tree discordance includes both topological incongruence and branch length heterogeneity.

Wu et al. (2013) attempted to defend Song et al. (2012) by arguing that, "...a recent study shows that recombination is not a primary factor influencing the accuracy of coalescent models" (p. E1180), but the cited work (Lanier and Knowles, 2012) assessed recombination in simulated loci that were only 1000 bp long, not up to >1,000,000 bp. Like their initial calculation of locus lengths (Song et al., 2012), Wu et al. (2013) were again off by several orders of magnitude (Fig. 7). Even if recombination does not drastically hurt the accuracy of shortcut coalescence methods (Lanier and Knowles, 2012), this is little consolation given the weak performance of these methods relative to simple concatenation in simulations (Fig. 13; Appendix B; Leaché and Rannala, 2011; Aguiar and Schrago, 2013; Bayzid and Warnow, 2013; Patel et al., 2013).

Wu et al. (2013) furthermore argued that, "because recombination only occurs within species, it is unlikely to be problematic for any but the most extreme cases of incomplete lineage sorting" (p. E1180). However, the most vexing empirical conflicts between concatenation and coalescence occur precisely where internodes are very short with "extreme cases of incomplete lineage sorting" (e.g., Fig. 3; Nishihara et al., 2006, 2009; Meredith et al., 2011; dos Reis et al., 2012; McCormack et al., 2012; Song et al., 2012). Coalescence methods were developed explicitly to resolve short internal branches such as this (Edwards, 2009). Unfortunately, reconstruction of accurate gene trees, a central concern for shortcut coalescence, is extremely challenging at such tight internodes given limited sequence data and extensive divergence among taxa. This explains the failures of shortcut coalescence methods in recent situations (Leaché and Rannala, 2011; Bayzid and Warnow, 2013; Patel et al., 2013).

At deep phylogenetic scales, the choice between shortcut coalescence and concatalescence therefore might be a choice between a rock and a hard place at this point – both methods entail distortions and model misfit (Fig. 14). With concatalescence, the most basic rationale for coalescence analysis, the independence provided by multiple c-gene tree histories, is abandoned. Instead, long segments of DNA that include sub-segments with contrasting evolutionary histories are treated as if they have a single history and are then processed by coalescence procedures that deny any history of recombination (Gatesy and Springer, 2013). The alternative to concatalescence, coalescence analysis of actual tiny c-genes (Fig. 7), is perhaps even more difficult to justify due to inaccurate c-gene reconstruction at deep phylogenetic scales (Figs. 8–10). The supermatrix approach also makes assumptions that are clearly violated (Fig. 14; i.e., assumes a single gene tree for all the data, or alternatively, assumes complete independence of all nucleotide substitutions), but thus far, concatenation has solidly out-competed shortcut coalescence in simulations due to the fact that inaccurate gene trees were more problematic for the coalescence methods than incomplete lineage sorting was for concatenation (Leaché and Rannala, 2011; Bayzid and Warnow, 2013; Patel et al., 2013).

The stark reality, at least in mammalian genomes that have been analyzed in detail so far (Hobolth et al., 2007, 2011), is that conflicting c-genes may be even smaller than the shortest loci that have been utilized in simulations and empirical systematic work of mammals (Fig. 7). Thus, a large percentage of published coalescence analyses at deep timescales are really concatalescence analyses in which multiple c-genes with conflicting histories have been merged and muddled. It will be critical, in the near future, to estimate average c-gene size for much larger sets of taxa from across the Tree of Life using models that simultaneously account for deep coalescence, recombination, and linkage as in the primate work (Hobolth et al., 2007, 2011; Dutheil et al., 2009). We predict that for large phylogenetic trees with multiple tight internodes (e.g., Meredith et al., 2011), the lengths of conflicting c-genes will be minute, because recombination events on short internodes will inexorably ratchet c-gene sizes downward with the addition of more and more taxa (Fig. 6).

At this time, methods that treat single nucleotide characters as the basic unit of analysis in systematics, using either concatenation (Kluge, 1989) or coalescence (Bryant et al., 2012), might be more appropriate than shortcut coalescence analysis of gene trees for resolving ancient relationships. For example, by randomly sampling individual homologous sites (SNPs) from across the genome and deriving a tree from the concatenation of these widely-spaced sites, any character correlations due to linkage would be minimized in a supermatrix analysis (Decker et al., 2009). Sets of randomly selected sites could be pooled based on overall rate to tailor the data to the particular time depth of a given phylogenetic problem, perhaps with a gamma parameter included to account for variation in evolutionary rates among sites. We see "gene-tree-free" approaches, such as this, that resample individual nucleotide positions, as a productive way forward with the rapid accumulation of genomic data. As noted by de Queiroz and Gatesy (2007, p. 36), at any particular node, "a strength of the supermatrix approach is that it might be expected to work well as long as one simple condition is met, namely that most [relevant informative] characters reflect any given branching event in the species tree." Proponents of the coalescence approach have focused on the fact that in the "anomaly zone," this condition may not prevail (Kubatko and Degnan, 2007; Kubatko et al., 2009; Liu et al., 2009b, 2010; Edwards, 2009), so it will be of interest in the future to determine how the supermatrix approach and shortcut coalescence methods behave in anomaly zone conditions at deep timescales (see Section 3.14 below).

3.13. Can concatenation save shortcut coalescence methods?

There is a growing concern that with the implementation of shortcut coalescence methods (Fig. 1D), a large segment of the systematics community has implicitly returned to faith in the analysis of small data partitions that are compromised by sampling error (Fig. 14; Meredith et al., 2011; Townsend et al., 2011; Chiari et al., 2012; Rosenfeld et al., 2012; Bayzid and Warnow, 2013; Gatesy and Springer, 2013; McCormack et al., 2013; Patel et al., 2013; Salichos and Rokas, 2013; Springer and Gatesy, 2014). The most general response to this concern has been to increase the information content of genes by: (1) sampling loci that evolve at high rates (e.g., Lanier et al., 2014) and/or (2) increasing the length of individual loci (e.g., Song et al., 2012).

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Given that DNA mutates only so fast, analysis of long sections of the genome has been advocated by some proponents of coalescence methods (Hallström et al., 2011; Song et al., 2012; Kumar et al., 2013; Xi et al., 2013; Zhong et al., 2013). For placental mammals, this concatalescence approach greatly increases congruence among gene trees relative to analysis of the shorter, more slowly evolving UCEs (Fig. 11; Supplementary Fig. 1). Even loci that on average include ~3100 bp (Song et al., 2012), however, yield gene trees with shockingly high error rates. For example, unique transposon insertions, characters that generally are not prone to reversal and convergence (Hartig et al., 2013), show 100% congruence in support of the traditional mammalian clade Haplorhini (Anthropoidea + Tarsiiformes) with ascertainment bias accounted for (Fig. 15). All of the 104 transposons that were screened group Anthropoidea (monkeys and apes) with Tarsiiformes (tarsiers) to the exclusion of Strepsirrhini (lemurs and kin). *None* of the 104 unique transposon insertions support the alternative conflicting topologies, and this implies a very low probability (perhaps zero?) of deep coalescence at this node (Hartig et al., 2013). Yet, for the 3100 bp loci employed by Song et al. (2012), 202 out of the 447 gene trees (45%) contradict Haplorhini (Fig. 15).

This example demonstrates that use of even highly unrealistic c-genes, in terms of overall length (Fig. 7), does not absolve shortcut coalescence methods from significant gene tree errors. Based on the >100 completely congruent transposon inserts, the gene tree reconstruction errors at the Haplorhini node approach 50% of the total (Fig. 15). If there is this much gene tree error, even with extensive concatalescence and at a node where 0% conflict is expected, it is worth pondering how much error of this type occurs at short, difficult to resolve internodes that are still deeper in the mammalian tree (Fig. 3). Zhong et al. (2014) contended that gene tree reconstruction errors have been grossly overestimated by critics of shortcut coalescence methods (Gatesy and Springer, 2013; Springer and Gatesy, 2014), but the above example provides a compelling counter to this claim (also see Figs. 8–11).

In addition to "local concatalescence," where adjacent c-genes (i.e., exons of the same gene) are grouped into much longer loci (Song et al., 2012), alternative "global concatalescence" procedures recently have been developed, but these admit concatenation of even more distantly linked regions of the genome to yield "pseudo-loci" for analysis by standard shortcut coalescence methods. Bayzid and Warnow (2013) offered two approaches. The first (naïve binning) entails concatenating randomly selected genes into sets (e.g. five genes in each set), and then applying shortcut coalescence methods to the various random concatenations of genes. The "pseudo-locus" in this concatalescence approach could include DNA segments from completely unlinked chromosomes that might have very different evolutionary histories due to incomplete lineage sorting. In the same paper, Bayzid and Warnow (2013) offered an alternate approach, statistical binning, wherein loci that do not strongly conflict with each other are binned (i.e., concatenated) and then analyzed by shortcut coalescence methods. However, this procedure also admits concatenation of distant linkage groups. All of the above procedures that are required to save shortcut coalescence from failure are hybrid methods that conflict with the main impetus for applying coalescence in the first place (Figs. 2 and 14), and in our view are concatenation methods as much as they are coalescence procedures.

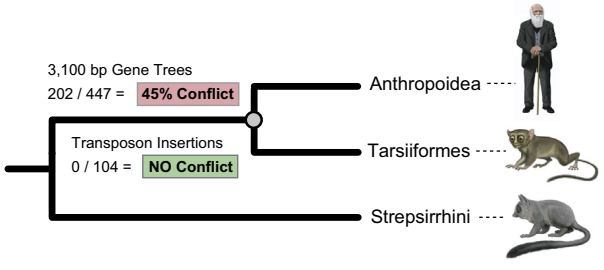

Fig. 15. Wholesale gene tree conflicts where none are expected. Genomic comparisons of primates indicate that 104 homoplasy-free transposon insertions support Haplorhini (Anthropoidea + Tarsiiformes; gray circle); 100% congruence implies a complete absence of deep coalescence at this node (Hartig et al., 2013). Yet, in the concatalescence analysis of Song et al. (2012), 202 of the 447 loci (45%) conflict at the same node, indicating an extremely high rate of gene tree reconstruction error despite analysis of loci that average ~3100 bp. Branch lengths are not proportional to time. Paintings are by C. Buell.

Concatalescence is a conflation of two distinct approaches, and we do not recognize any logical or biological rationale for its use. Instead, concatalescence requires the abandonment of a fundamental tenet of the coalescence approach to systematics – that recombination occurs between loci but not within loci. Nevertheless, competing methods also should be assessed by empirical performance through simulations and congruence. Employing simulations, Bayzid and Warnow (2013) showed that their first concatalescence procedure described above (naïve binning) was more accurate than standard shortcut coalescence using MP-EST. This result is expected. By incorporating concatenation into a standard shortcut coalescence method, results improved. However, results for the concatalescence method were less accurate than simple ML concatenation using a single model of evolution. This second outcome was not surprising either; unadulterated concatenation performed better than analysis of smaller concatenations that were tainted by subsequent processing via shortcut coalescence methods. In this study, supermatrix analysis was again more accurate than methods that required many more parameters to first derive separate gene trees and then a species tree – with no discernible payoff in accuracy (Bayzid and Warnow, 2013).

3.14. The anomaly zone at deep phylogenetic scales

Coalescence theory predicts a so-called "anomaly zone," a set of conditions wherein the most probable gene tree conflicts with the species tree. Gene trees that are more probable than the gene tree that matches the species tree topology are referred to as anomalous gene trees (AGTs) (Kubatko and Degnan, 2007). With complete neutrality, no AGTs are expected for three species, but AGTs may occur with four taxa when the species tree is pectinate and there are short internal branches (Fig. 16; Kubatko and Degnan, 2007; Degnan, 2013). With five or more taxa, AGTs may be present for any tree topology, depending on the lengths of internodes (Rosenberg and Tao, 2008).

Theoretical work and simulations suggest that with the accumulation of sequence data from many loci, concatenation may recover an incorrect species tree with high support in the anomaly zone (Kubatko and Degnan, 2007). Thus, the anomaly zone has been a primary motivation for the development of coalescence methods in phylogenetics (Edwards, 2009; Liu and Edwards, 2009). The coalescence approach originally was applied most commonly to shallow divergences (e.g., Carstens and Knowles, 2007) but more recently has been advocated for resolving ancient radiations (Song et al., 2012; McCormack et al., 2012; Xi et al., 2013) that in some cases extend into the Precambrian (Zhong et al., 2013). A critical question is whether the anomaly zone should be considered a major danger for systematic analysis at such timescales (Huang and Knowles, 2009).

Empirical estimates for population size and generation time provide a framework for visualizing anomaly zone conditions at contrasting temporal depths (Fig. 16). As noted above, for many vertebrates, one year and 200,000 individuals are fair approximations for generation time and population size, respectively (Patel et al., 2013). These values yield coalescent units equivalent to

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

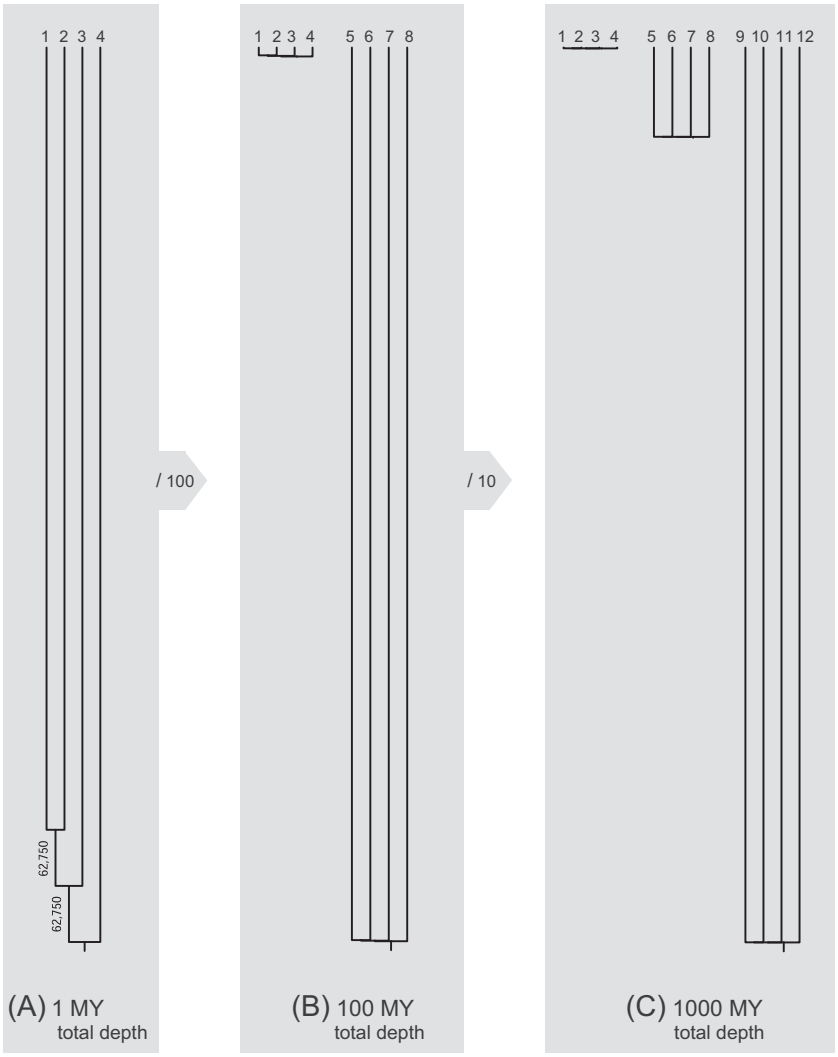

Fig. 16. Three trees that depict anomaly zone conditions at different time depths. All three trees have short internal branches that are each 62,720 years (0.1568 coalescent units per branch assuming a generation time of 1 year and population size of 200,000 diploid individuals – see text). The tree with taxa 1, 2, 3, and 4 has a total depth of 1 million years (MY) and is shown in the left (A), middle (B), and right (C) gray panels at three different scales. The tree with taxa 5, 6, 7, and 8 has a total depth of 100 MY, and is shown in panels (B) and (C) at two scales. The tree with taxa 9, 10, 11, and 12 has a total depth of 1000 MY and is shown in panel (C) at a single scale. In (A), the total depth of the tree is 1 MY, in (B), the total depth of the longest tree is 100 MY, and in (C), the total depth of the longest tree is 1000 MY. Branch lengths are in millions of years; the arrows pointing from left to right indicate the changes in scale between panels. Note that the short internodes in these trees are nearly invisible at two of the timescales. In the panel on the right, the entire tree that includes taxa 1–4 is nearly invisible. This miniscule tree (Kubatko and Degnan, 2007) approximates the tree depth and branch lengths used in the few simulations that suggest shortcut coalescence is superior to concatenation (Liu et al., 2009b, 2010; see Fig. 13).

400,000 years. With two consecutive internal branches of equal length, pectinate trees of four taxa fall into the anomaly zone when both branches are each $\leqslant$0.1568 coalescent units (Kubatko and Degnan, 2007). One of these two branches may be slightly longer, but at the expense of shortening the second branch (Kubatko and Degnan, 2007). Consecutive internal branches of 0.1568 coalescent units are discernible in a species tree with a total depth of only 1 MY (Fig. 16A), but these short internal branches of 0.1568 coalescent units are effectively invisible at 100 MY or 1 billion years (Fig. 16B–C).

Although shortcut coalescence methods have been hailed as saviors when anomaly zone conditions prevail (Liu et al., 2009a,b), this notion has only been corroborated in simulations at very shallow scales (Fig. 13; Appendix B; Liu et al., 2009b, 2010). Given the relative lengths of internal and terminal branches in the anomaly zone at 100–1000 MY, we would not expect shortcut coalescence methods to resolve such internodes – the internal branches cannot even be clearly visualized at this timescale (Fig. 16B–C). Problems due to inaccurate gene trees are expected to be most acute when internal branches are very short as in the "anomaly zone," and at depth, and we predict that both coalescence and concatenation approaches will fail to consistently recover the correct species tree in such cases (see Patel et al., 2013).

To reconstruct accurate gene trees at these depths, long sequences would be required, but long sequences likely represent concatenations of multiple c-genes, and the merits of coalescence methods in the anomaly zone are not guaranteed when conflicting c-genes are concatenated. To our knowledge, no simulations have shown superiority of shortcut coalescence methods in the anomaly zone at depth, despite inductive hopes (Zhong et al., 2013, 2014) based on species trees simulated at shallow depths (Liu et al., 2009b, 2010). In much less challenging situations at deep timescales, shortcut coalescence methods have been out-competed so far in direct comparisons to simple concatenation (Fig. 13; Appendix B). We therefore suspect that a perceived advantage of coalescence methods, success in anomaly zone conditions, is irrelevant at depth (Fig. 16B–C).

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

3.15. "Supermatrix trees" versus "coalescence trees": whose species tree is it anyway?

Theory predicts deep coalescence, due to retention of ancestral polymorphism, at all levels of divergence, and the resulting gene tree heterogeneity is not directly accounted for by the mechanics of supermatrix analysis (Fig. 14; Degnan and Rosenberg, 2009; Edwards, 2009; Oliver, 2013). The supermatrix approach has been characterized as a giant gene tree analysis in which the concatenated data are treated like a single "supergene." Lanier and Knowles (2012: p. 1) neatly summarized this view, "Concatenation assumes that all genes share a common underlying tree, a model violated by most multilocus data." At this point, a segment of the systematics community has concluded that estimation of a species tree *requires* analysis of multiple gene trees (e.g., Edwards et al., 2009; Liu et al., 2009a; Knowles, 2010; Knowles and Kubatko, 2010). For example, Edwards (2009; p. 12) noted that, "...I suggest we simply exercise a verbal substitution and reserve the term "phylogeny" to refer to species trees. Phylogenies as they have been built in the last few decades would then be called gene trees, which is generally what they are, *sensu stricto*." From this perspective, a supermatrix analysis presumably does not estimate phylogeny, and instead reconstructs something else or something less, and has been characterized as naive (Edwards, 2009).

A species tree is a representation of the evolutionary history of species, while a gene tree records the evolutionary history of a single c-gene (Maddison, 1997). Should a tree derived from a supermatrix that includes information from many loci be considered a "species tree," or a "gene tree" (Edwards et al., 2009; Knowles, 2010)? It is clear that the goal of authors who have concatenated many genes from multiple species to derive a tree is to assess the phylogenetic relationships among species (e.g., Meredith et al., 2011). With respect to recent simulations that yield more accurate species tree reconstructions for concatenation relative to shortcut coalescence (Fig. 13; Appendix B), it would be ironic to describe the more accurate supermatrix trees as giant "gene trees" while simultaneously accepting that the less accurate trees derived from coalescence methods are "species trees." Defining away the species tree methods of competitors by pejoratively noting that they yield "gene trees" due to lack of perfect match to certain biological observations (i.e., assume no deep coalescence; Edwards, 2009; Knowles, 2010) does not excuse shortcut coalescence methods from their own flaws that may be even more debilitating for the reconstruction of a species tree (i.e., assume gene trees can be reconstructed accurately, assume all incongruence among genes is due to deep coalescence, assume no recombination within a locus, etc.) (Fig. 14). More than 17 years after the initial development of shortcut coalescence methods (Maddison, 1997), it is not clear that this systematic approach produces more accurate species trees at deep timescales relative to simple supermatrix methods that are >30 years old (Felsenstein, 1981; Fig. 13).

The assertion that supermatrix trees are not estimates of "species trees" (Knowles, 2010) is even more problematic when applied to supermatrices that merge DNA sequences with phenotypic characters, behavioral traits, transposon insertions, chromosomal rearrangements, and fossils. Such phylogenetic studies seek to distill common phylogenetic signals from diverse data to reconstruct phylogenetic relationships among species (Eernisse and Kluge, 1993; Gatesy et al., 2004; Nylander et al., 2004; Deméré et al., 2008; Geisler et al., 2011; Ronquist et al., 2012). For groups such as Mammalia, extinct taxa represent the majority of described taxonomic diversity (McKenna and Bell, 1997); current coalescence methods cannot incorporate phenotypic characters and fossils into species tree estimates (Fig. 14), despite vague hopes to develop this capability some day (Liu et al., 2009a). Therefore, species trees derived from supermatrices will continue to be the standard for taxa with rich fossil records such as Mammalia (Simmons and Geisler, 1998; Horovitz, 1999; O'Leary, 1999, 2001; Asher et al., 2003, 2005; Bibi et al., 2012; O'Leary et al., 2004, 2013; Asher and Hofreiter, 2006; Deméré et al., 2008; O'Leary and Gatesy, 2008; Geisler and Theodor, 2009; Geisler et al., 2011; Spaulding et al., 2009; Pérez and Pol, 2012; Gatesy et al., 2013).

By directly accounting for sorting of ancestral polymorphism, linkage, recombination, gene flow, and gene tree heterogeneity, the coalescence approach potentially offers a seamless connection between population genetics (Hudson, 1990) and phylogenetics (Maddison, 1997). This is a noble goal, but a fear is that in seeking to achieve this endpoint, proponents of coalescence methods may have overplayed their hand and ignored defects in their own methods at very deep timescales, where population genetics models have not been applied in the past (Gatesy and Springer, 2013; Springer and Gatesy, 2014). With the analysis of large, complex datasets, the supermatrix approach still holds critical advantages over current shortcut coalescence methods for choosing among alternative relationships at ancient splits in the Tree of Life (e.g., incorporation of diverse data, computational speed). In particular, like previous systematic procedures that were based on isolated analyses of small data partitions (Fig. 1B–C), shortcut coalescence methods (Fig. 1D) bypass hidden support that is strongly expressed in supermatrix analysis (Fig. 1A) of the same data (Fig. 3). It will be informative to see how these systematic paradigms compare in future phylogenetic work at deep timescales where congruence across empirical studies and head-to-head success in simulations are assessed in much more detail.

4. Conclusions

1. Very large datasets are required to answer challenging phylogenetic questions at deep divergences. For even moderately sized datasets, however, coalescence co-estimation methods are computationally intractable and shortcut coalescence methods instead must be implemented.
2. For ancient clades, phylogenetic analysis of modern systematic datasets (many genes and many taxa) has taken two paths given currently available methods – supermatrix analysis or shortcut coalescence analysis. Each approach includes assumptions that are violated by empirical systematic data (Fig. 14).
3. For ten uncontroversial mammalian clades, extreme hidden support emerged in 10 of 10 trials with the supermatrix approach, but in only 8 of 20 of trials when two different shortcut coalescence methods were applied to the same UCE data from Mammalia (Fig. 3). STAR and MP-EST could not recover traditionally easy to resolve groups with up to 175 loci (Fig. 4).
4. In the shortcut coalescence examples, hidden support is synapomorphy-free, and in most cases is likely an artifact that is unrelated to the reconciliation of actual deep coalescences (e.g., Fig. 5).
5. Published coalescence analyses of mammalian phylogeny strongly conflict with each other. This is an empirical refutation of the persistent assertion that supermatrix analysis is prone to strong conflicts among analyses whereas coalescence methods are not.
6. Isolated analyses of individual UCE loci were largely uninformative and yield odd gene trees, in part due to arbitrary resolution in the fully bifurcating gene trees that are required input for STAR and MP-EST (Figs. 8–10). The dense mammalian fossil record, in combination with rigorous molecular clock studies, imply unrealistic retentions of ancestral

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

polymorphism, with the deepest coalescence in each locus averaging >100 MY (Table 2). Coalescence analysis of gross coalescence artifacts, such as this, should not be expected to solve difficult systematic problems that are deep in the Tree of Life.

7. Recent simulations indicate that concatenation outperforms shortcut coalescence methods at deep timescales (Fig. 13; Appendix B) because inaccurately reconstructed gene trees are more problematic for coalescence methods than incomplete lineage sorting is for concatenation. To our knowledge, no published simulation study has shown improved accuracy by choosing shortcut coalescence over concatenation at deep timescales. More simulation work is required to test the generality of the few direct comparisons between coalescence methods and concatenation at depth.
8. Based on genomic comparisons of only four primate species, conflicting c-genes are tiny, with most <100 bp (Fig. 7). C-gene size can only ratchet downward as taxonomic sampling increases, while the number of nodes in each gene tree ratchets inexorably higher with improved taxonomic sampling (Fig. 6). This double ratchet is a debilitating equation for shortcut coalescence methods that are predicated on accurate reconstruction of gene trees.
9. An obvious response to the inaccuracy of gene trees reconstructed from short loci is to concatenate. Indeed, analysis of longer loci greatly improves congruence among gene trees (Fig. 11). Unfortunately, "concatalescence," the concatenation of distantly linked stretches of the genome followed by application of coalescence methods, undermines the primary reason for applying coalescence methods to difficult systematic problems.
10. Recombination within a particular locus is expected to have an impact on shortcut coalescence methods primarily at short internodes, but these are the internodes that are difficult to resolve, where there are conflicts among different published studies, and where extensive gene tree reconstruction errors are expected (Fig. 3).
11. Even when loci are quite long, wholesale conflicts among gene trees are apparent. At a node within Primates that shows zero conflict among 104 transposon insertion characters, 202 of 447 gene trees (45%) contradict this clade, even though the underlying genes averaged ~3100 bp (Fig. 15). Interpreting such conflicts as due to deep coalescence and as a rationale for employing coalescence methods is naive. Gene tree reconstruction errors provide a simple explanation for stunted MP-EST "bonsai trees" that are inconsistent with the fossil record and molecular clocks (Fig. 12).
12. Simulations show that shortcut coalescence methods can resolve species trees in the anomaly zone at very shallow depths and with only a few taxa (Fig. 13; Appendix B), but for ancient radiations, internodes in the anomaly zone are nearly invisible (Fig. 16). To our knowledge, no published study has shown that shortcut coalescence methods (or supermatrix methods) can resolve such difficult problems using standard analyses of DNA sequence data.
13. Despite claims that shortcut coalescence studies have solved long-standing polytomies deep in the placental tree (Fig. 3), this work has generated more confusion than resolution. In particular, the complete 183-locus UCE dataset (McCormack et al., 2012) supports conflicting relationships at the root of placental mammals when analyzed with different shortcut coalescence methods (Figs. 3 and 12). Both resolutions favored by the UCE data disagree with an independent shortcut coalescence analysis that also claimed to solve this difficult systematic problem (Song et al., 2012).
14. History has repeated itself. Taxonomic congruence (Fig. 1B), the supertree approach (Fig. 1C), and shortcut coalescence methods (Fig. 1D) suffer from the same problems – isolated analyses of short loci, associated errors in gene tree reconstruction, and the resulting masking/distortion of phylogenetic signal. There is no problem, per se, with applying coalescence methods to deep divergences in the Tree of Life, but retention of ancestral polymorphism is not the only challenge to phylogenetic analysis at this scale. Our hope is that our discussion will inspire future work to improve computationally tractable coalescence methods that can be applied to complex data, eventually even including phenotypic and fossil data – as supermatrix methods have for several decades.

Acknowledgments

Two anonymous reviewers, R. Baker, A. de Queiroz, C. Hayashi, M. Simmons, and students in EEOB 282 (2013) at U. of California – Riverside gave helpful comments on various drafts of the paper. J. McCormack provided DNA sequence alignments and gene trees for 183 UCE loci. Simulation conditions in published studies were clarified by email discussions with C. Ané, E. Braun, R. Kimball, L. Kubatko, A. Leaché, B. Rannala, T. Warnow, and Y. Yu. This work was funded by NSF grant EF0629860 to M.S.S. and J.G.

Appendix A. Probabilities of deep coalescence and incongruence with the species tree

Probabilities that two alleles do not coalesce and are incongruent with the species tree for two different branches, stem Placentalia and stem Anthropoidea, given divergence times from dos Reis et al. (2012), a range of effective population sizes (N_e), and generation times of one or five years. Probabilities of incongruence (alleles not coalescing) were calculated using the equation $p(\text{incongruence}) = 2/3\exp(-t)$, where t is the time interval between two consecutive branching events (e.g., length of stem placental branch) in units of generations/$2N_e$ (Pollard et al., 2006; Hobolth et al., 2011). The number of coalescent units is given in parentheses. Patel et al. (2013) suggested that $2N_e$ = 400,000 and generation time = one year are appropriate for many vertebrates.

$2N_e$	83.7 MY stem placental branch with generation time = 1 year	83.7 MY stem placental branch with generation time = 5 years	27.6 MY stem anthropoid branch with generation time = 1 year	27.6 MY stem anthropoid branch with generation time = 5 years
40,000	0 (2092.5)	1.179419e−182 (418.5)	1.447825e−300 (690)	7.785207e−61 (138)
400,000	8.86724e−92 (209.25)	4.453327e−19 (41.85)	7.20426e−31 (69)	6.770873e−07 (13.8)
4,000,000	5.448747e−10 (20.925)	0.01014814 (4.185)	0.000671856 (6.9)	0.1677191 (1.38)

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Appendix B. Summary of simulation studies comparing coalescence to concatenation

Descriptions of the few simulation studies that have compared shortcut coalescence methods to concatenation in head-to-head trials are provided below (also see Fig. 13). To our knowledge, there are only seven relevant studies, and only four at deep timescales that have directly compared these approaches. Six studies simulated data via coalescence models (Kubatko et al., 2009; Liu et al., 2009b, 2010; Leaché and Rannala, 2011; Bayzid and Warnow, 2013; Patel et al., 2013), and one simulated data via a concatenation model (Aguiar and Schrago, 2013). All gene trees were estimated from the simulated DNA sequence data using ML. All loci analyzed in these simulations were longer than conflicting c-genes in empirical analyses of primate genomes (Hobolth et al., 2007, 2011), and recombination within loci was not modeled. To date, we know of no simulation study where direct comparison of concatenation to shortcut coalescence at deep timescales has found shortcut coalescence to be consistently more accurate.

Kubatko et al. (2009)

Methods compared: STEM versus ML concatenation
Number of taxa: 4
Depth of species tree: 2.2–3.0 CUs
Species tree shape: one pectinate topology with five sets of branch lengths. Only the basal internal branch was varied.
Number of loci: 10
Length of loci: 500 bp (L. Kubatko, pers. com.)
Model of DNA sequence evolution: Jukes and Cantor (molecular clock)
Measure of accuracy: tree topology is correct, but see "Notes."
Summary of results: STEM is more accurate than concatenation over all parameters that were tested but see "Notes."
Notes: This is the only simulation study that we know of wherein STEM was found to be more accurate than concatenation in a head-to-head comparison. This result might be attributable to two factors. First, there were only four taxa in the species tree, and the basal internal branch of the tree was the shortest internal branch. However, this internal branch does not exist for concatenation, unless the supermatrix tree is midpoint rooted. This simulation is therefore an exercise in resolving a very short branch at the base of a tree via midpoint rooting, a procedure that is rarely utilized by serious workers in the field. Second, only simulation runs where STEM produced a single tree were included in the results (L. Kubatko, pers. com.). Thus, cases where STEM gave an ambiguous result were not counted, but one of the primary goals of systematics is, of course, to choose among alternative trees.

Liu et al. (2009b)

Methods compared: STAR and STEAC versus Bayesian concatenation
Number of taxa: 5
Depth of species tree: 3.1 CUs
Species tree shape: one pectinate tree with one set of branch lengths in the anomaly zone
Number of loci: 100, 500, or 1000
Length of loci: 500 bp
Model of DNA sequence evolution: Jukes and Cantor (molecular clock)
Measure of accuracy: tree topology is correct
Summary of results: STAR and STEAC are more accurate than concatenation for 100, 500, and 1000 loci.
Notes: This study has been cited repeatedly as evidence that STAR and STEAC are superior to concatenation, but extrapolation from this shallow, pectinate tree (see Fig. 13) to much deeper radiations in the Tree of Life is speculative.

Liu et al. (2010)

Methods compared: MP-EST versus Bayesian concatenation
Number of taxa: 5
Depth of species tree: 3.1 CUs
Species tree shape: one pectinate tree with one set of branch lengths in the anomaly zone
Number of loci: 100, 500, 1000, 2000, or 2500
Length of loci: 500 bp
Model of DNA sequence evolution: Jukes and Cantor (molecular clock)
Measure of accuracy: tree topology is correct
Summary of results: MP-EST is more accurate than concatenation for 500, 1000, 1500, 2000, and 2500 loci; the two methods performed about equally well for 100 loci.
Notes: Simulated more loci than the other studies. This study has been cited repeatedly as evidence that MP-EST is superior to concatenation, but extrapolation from this shallow, pectinate tree (see Fig. 13) to much deeper radiations in the Tree of Life is speculative.

Leaché and Rannala (2011)

Methods compared: STEM versus Bayesian concatenation and concatenation with equally weighted parsimony
Number of taxa: 5
Depth of species tree: a range of depths from 0.25 to 64 CUs (see Fig. 13).
Species tree shape: Pectinate or maximally symmetrical for the ingroup of four taxa. Multiple branch lengths were generated via birth/death model with a variety of population size parameters.
Number of loci: 10 or 100
Length of loci: 1000 bp
Model of DNA sequence evolution: Jukes and Cantor (molecular clock)
Measure of accuracy: tree topology is correct, but dealt with ties by scoring partial success if correct tree was among the set of equally optimal topologies.
Summary of results: Bayesian concatenation was always more accurate than STEM, and concatenation with equally weighted parsimony was better than STEM for most parameter sets.

Bayzid and Warnow (2013)

Methods compared: MDC and MP-EST versus ML concatenation
Number of taxa: 11 or 17
Depth of species tree: 32 CUs for 11 taxon trees; 8 CUs for 17 taxon trees
Species tree shape: 1 topology and 2 sets of branch lengths for 11 taxon trees (see Fig. 13); Yule uniform speciation model generated multiple topologies and branch lengths for 17 taxon trees
Number of loci: 5, 10, 25, 50, or 100 for 11-taxon trees; 8 or 32 for 17-taxon trees
Length of loci: 500 bp for 11-taxon trees; 2000 bp for 17-taxon trees
Model of DNA sequence evolution: Jukes and Cantor (no

molecular clock) for 11 taxon species trees, and Jukes and Cantor (molecular clock) for 17-taxon species trees
Measure of accuracy: missing branch rate (i.e., false negative rate)
Summary of results: ML concatenation uniformly outperformed MDC and MP-EST. Arbitrary concatalescence ("naïve binning" followed by coalescence) fared better but was still out-competed by simple concatenation.
Notes: Bayzid and Warnow (2013) used simulated data from two previously published papers (Chung and Ané, 2011; Yu et al., 2011). *BEAST, a coalescence co-estimation procedure, was more accurate than the shortcut coalescence methods, but was bested by simple concatenation for half of the parameter combinations that were assessed.

Patel et al. (2013)

Methods compared: STEM versus ML concatenation
Number of taxa: 20
Depth of species tree: initial radiation (see below) placed at various depths (0.5, 1, 5, 25, 50, 100, or 200 CUs; see Fig. 13)
Species tree shape: radiations were examined by using a Yule process to generate the initial diversification at several speciation rates. The initial diversification was then positioned at various depths (see above and Fig. 13).
Number of loci: 5 or 50
Length of loci: 1000 bp
Model of DNA sequence evolution: HKY model with rates based on empirical intron data from tetrapods (birds and mammals)
Measure of accuracy: Robinson-Foulds difference relative to the true species tree.
Summary of results: ML concatenation out-performed STEM for 5 loci and badly out-competed STEM with 50 loci. All methods performed poorly for very rapid speciation rates at depth.
Notes: Adding 45 loci did not help STEM very much, but there was a large increase in accuracy for concatenation.

Aguiar and Schrago (2013)

Methods compared: STAR and STEAC versus ML concatenation
Number of taxa: 64
Depth of species tree: based on empirical data; 500 MY depth and data generated via a concatenation model (no variation among gene tree topologies) so branch lengths were not in CUs
Species tree shape: 3 topologies: maximally symmetrical, pectinate, and a third topology based on the animal Tree of Life from Dunn et al. (2008). One set of branch lengths for each tree.
Number of loci: 50
Length of loci: up to ~10,000 bp but most between 500 bp and 2000 bp; based on empirical gene sizes that were stripped of introns.
Model of DNA sequence evolution: HKY + gamma; model parameters based on empirical data with molecular clock, but rates for different genes were not the same. Tested high rate variation and lower rate variation sets of loci.
Measure of accuracy: Penny and Hendy distance
Summary of results: All methods were accurate for the symmetrical tree. Concatenation outperformed STAR and STEAC consistently for the pectinate tree. For the tree based on empirical data, results were mixed, depending on the rate variation among loci (low rate variation favored concatenation). For the two topologies that showed mixed results for concatenation versus shortcut coalescence (pectinate tree and tree based on empirical data), concatenation was more accurate for 75% of the parameter sets tested.
Notes: Aguiar and Schrago (2013) noted that generating data using a concatenation model might favor concatenation in simulations at depth, so they suggested that data should be generated via coalescence models. However, in previous (Leaché and Rannala, 2011) and subsequent (Bayzid and Warnow, 2013; Patel et al., 2013) simulations that generated data via coalescence models, concatenation consistently out-competed shortcut coalescence at depth (Fig. 13; see above). Aguiar and Schrago (2013) also compared shortcut coalescence methods to concatention methods using empirical data from mammalian genomes and by assessing match to an uncontroversial mammalian tree. Concatenation consistently out-competed STEAC and was matched by STAR in these trials based on actual genomic data.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.08.013.

References

Aguiar, B.O., Schrago, C.G., 2013. Conventional simulation of biological sequences leads to a biased assessment of multi-loci phylogenetic analysis. Evol. Bioinform. 9, 317–325.
Anderson, F.E., Swofford, D.L., 2004. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. Mol. Phylogenet. Evol. 33, 440–451.
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, New Jersey, pp. 35–52.
Ané, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24, 412–426.
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.
Asher, R.J., Emry, R.J., McKenna, M.C., 2005. New material of *Centetodon* (Mammalia, Lipotyphla) and the importance of (missing) DNA sequences in systematic paleontology. J. Vertebr. Paleontol. 25, 911–923.
Asher, R.J., Hofreiter, M., 2006. Tenrecid phylogeny and the noninvasive extraction of nuclear DNA. Syst. Biol. 55, 181–194.
Avise, J.D., Robinson, T., 2008. Hemiplasy: a new term in the lexicon of phylogenetics. Syst. Biol. 57, 503–507.
Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46, 654–673.
Barrett, M., Donoghue, M.J., Sober, E., 1991. Against consensus. Syst. Zool. 40, 486–493.
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.
Bibi, F., Fack, F., Vrba, E., 2012. A new fossil caprin and a combined molecular and morphological Bayesian analysis of Caprini (Mammalia: Bovidae). J. Evol. Biol. 25, 1843–1854.
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.
Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42, 384–397.
Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from melanoplus grasshoppers. Syst. Biol. 56, 400–411.

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.

Chippindale, P.T., Wiens, J.J., 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43, 278–287.

Chung, Y., Ané, C., 2011. Comparing two Bayesian methods for gene tree/species tree reconstruction: a simulation with incomplete lineage sorting and horizontal gene transfer. Syst. Biol. 60, 261–275.

Collins, T.M., Fedrigo, O., Naylor, G.J.P., 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenies. Syst. Biol. 54, 493–500.

Decker, J.E., Pires, J.C., Conant, G.C., McKay, S.D., Heaton, M.P., Chen, K., Cooper, A., Vilkki, J., Seabury, C.M., Caetano, A.R., et al., 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proc. Natl. Acad. Sci. USA 106, 18644–18649.

DeGiorgio, M., Degnan, J.H., 2014. Robustness to divergence time underestimation when inferring species trees from estimated gene trees. Syst. Biol. 63, 66–82.

Degnan, J.H., 2013. Anomalous unrooted gene trees. Syst. Biol. 62, 574–590.

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

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.

Degnan, J.H., Salter, L.A., 2005. Gene tree distributions under the coalescent process. Evolution 59, 24–37.

Deméré, T.A., McGowen, M.R., Berta, A., Gatesy, J., 2008. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. Syst. Biol. 57, 15–37.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., et al., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–749.

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

de Queiroz, A., Donoghue, M.J., Kim, J., 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26, 657–681.

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. Roy. 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., 1992. Gene trees and species trees: molecular systematics as one character taxonomy. Syst. Bot. 17, 144–163.

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

Dutheil, J.Y., Ganapathy, G., Hobolth, A., Mailund, T., Uyenoyama, M.K., Schierup, M.H., 2009. Ancestral population genomics: the coalescent hidden Markov model approach. Genetics 183, 259–274.

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

Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104, 5936–5941.

Eernisse, D.J., Kluge, A.G., 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol. Biol. Evol. 10, 1170–1195.

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

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Gatesy, J., 1998. Molecular evidence for the phylogenetic affinities of Cetacea. In: Thewissen, J.G.M. (Ed.), The Emergence of Whales. Plenum, New York, pp. 63–111.

Gatesy, J., 2000. Linked branch support and tree stability. Syst. Biol. 49, 800–807.

Gatesy, J., Arctander, P., 2000. Hidden morphological support for the phylogenetic placement of *Pseudoryx nghetinhensis* with bovine bovids: a combined analysis of gross anatomical evidence and DNA sequences from five genes. Syst. Biol. 49, 515–538.

Gatesy, J., Baker, R.H., 2005. Hidden likelihood support in genomic data: can forty-five wrongs make a right? Syst. Biol. 54, 483–492.

Gatesy, J., Springer, M.S., 2004. A critique of matrix representation with parsimony supertrees. In: Bininda-Emonds, O. (Ed.), Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life, Computational Biology, vol. 4. Kluwer Academic Publishers, pp. 369–388.

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

Gatesy, J., O'Grady, P., Baker, R.H., 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15, 271–313.

Gatesy, J., Matthee, C., DeSalle, R., Hayashi, C., 2002. Resolution of a supertree/supermatrix paradox. Syst. Biol. 51, 652–664.

Gatesy, J., Amato, G., Norell, M., DeSalle, R., Hayashi, C., 2003. Combined support for wholesale taxic atavism in gavialine crocodylians. Syst. Biol. 52, 403–422.

Gatesy, J., Baker, R.H., Hayashi, C., 2004. Inconsistencies in arguments for the supertree approach: supermatrices versus supertrees of Crocodylia. Syst. Biol. 53, 342–355.

Gatesy, J., DeSalle, R., Wahlberg, N., 2007. How many genes should a systematist sample? conflicting insights from a phylogenomic matrix characterized by replicated incongruence. Syst. Biol. 56, 355–363.

Gatesy, J., Geisler, J.H., Chang, J., Buell, C., Berta, A., Meredith, R.W., Springer, M.S., McGowen, M.R., 2013. A phylogenetic blueprint for a modern whale. Mol. Phylogenet. Evol. 66, 479–506.

Geisler, J.H., Theodor, J.M., 2009. Brief communications arising: hippo and whale phylogeny. Nature 458, E1–E4.

Geisler, J.H., McGowen, M.R., Yang, G., Gatesy, J., 2011. A supermatrix analysis of genomic, morphological, and paleontological data from crown Cetacea. BMC Evol. Biol. 11, 112.

Goloboff, P., Pol, D., 2002. Semi-strict supertrees. Cladistics 18, 514–525.

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.

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.

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.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., van Vuuren, B.J., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T.T., Couloux, A., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. CR. Biol. 335, 32–50.

Hedtke, S.M., Townsend, T.M., Hillis, D.M., 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55, 522–529.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Hillis, D.M., 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Syst. 18, 23–42.

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.

Horovitz, I., 1999. A phylogenetic study of living and fossil platyrrhines. Am. Mus. Novit. 3269, 1–40.

Hovmöller, R., Knowles, L.L., Kubatko, L.S., 2013. Effects of missing data on species tree estimation under the coalescent. Mol. Phylogenet. Evol. 69, 1057–1062.

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.

Hudson, R.R., 1990. Gene genealogies and the coalescent process. Ox. Surv. Evol. Biol. 7, 1–44.

Huelsenbeck, J.P., Hillis, D.M., 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. 42, 247–264.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Janies, D.A., Studer, J., Handelman, S.K., Linchangco, G., 2013. A comparison of supermatrix and supertree methods for multilocus phylogenetics using organismal datasets. Cladistics 29, 560–566.

Kimball, R.T., Wang, N., Heimer-McGinn, V., Ferguson, C., Braun, E.L., 2013. Identifying localized biases in large datasets: a case study using the avian tree of life. Mol. Phylogenet. Evol. 69, 1021–1032.

Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38, 7–25.

Knowles, L.L., Kubatko, L.S. (Eds.), 2010. Estimating Species Trees: Practical and Theoretical Aspects. Wiley & Sons, Hoboken, New Jersey.

Knowles, L.L., 2010. Sampling strategies for species tree estimation. In: Knowles, L.L., Kubatko, L.S. (Eds.), Estimating Species Trees: Practical and Theoretical Aspects. Wiley-Blackwell, Hoboken, New Jersey, pp. 163–173.

Knowles, L.L., Lanier, H.C., Kilmov, P., He, Q., 2012. Full modeling versus summarizing phylogenetic uncertainty: method choice and species-tree accuracy. Mol. Phylogenet. Evol. 65, 501–509.

Kubatko, L.S., 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58, 478–488.

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

Kubatko, L.S., Carstens, B.C., Knowles, L.L., 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25, 971–973.

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., 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.

Lambkin, C., 2004. Partitioned Bremer support localizes significant conflict in bee flies (Diptera: Bombyliidae: Anthracinae). Invert. Syst. 18, 351–360.

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

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Lee, M., Hugall, A., 2003. Partitioned likelihood support and the evaluation of data set conflict. Syst. Biol. 52, 15–22.
Leffler, E.M., Gao, Z., Pfeifer, S., Ségurel, L., Auton, A., Venn, O., Bowden, R., Bontrop, R., Wall, J.D., Sella, G., et al., 2013. Multiple instances of ancient balancing selection shared between humans and chimpanzees. Science 339, 1578–1582.
Lemmon, A.R., Brown, J.M., Stanger-Hall, K., Lemmon, E.M., 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Syst. Biol. 58, 130–145.
Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al., 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478, 476–482.
Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542–2543.
Liu, L., Edwards, S.V., 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58, 452–460.
Liu, L., Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56, 504–514.
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.
Luo, Z.-X., Yuan, C.-X., Meng, Q.-J., Ji, Q., 2011. A Jurassic eutherian mammal and divergence of marsupials and placentals. Nature 476, 442–445.
Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.
Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21–30.
Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, 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.
McCormack, J., Harvey, M., Faircloth, B., Crawford, N., Glenn, T., et al., 2013. A phylogeny of birds based on over 1500 loci collected by target enrichment and high-throughput sequencing. PLoS ONE 8, e54848.
McKenna, M.C., Bell, S.K., 1997. Classification of Mammals Above the Species Level. Columbia University Press, New York.
McVay, J., Carstens, B., 2013. Phylogenetic model choice: justifying a species tree or concatenation analysis. J. Phylogenet. Evol. Biol. 1, 114.
Meng, J., Hu, Y., Li, C., 2003. The osteology of *Rhombomylus* (Mammalia: Glires): implications for the phylogeny and evolution of Glires. Bull. Am. Mus. Nat. Hist. 275, 1–247.
Meredith, R.W., Janečka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L.L., Stadler, T., et al., 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. Gateway Computing Environments Workshop, 1–8.
Miyamoto, M., 1985. Consensus cladograms and general classifications. Cladistics 1, 186–189.
Miyamoto, M.M., Fitch, W.M., 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44, 64–76.
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.
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., et al., 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.
Narechania, A., Baker, R.H., Sit, R., Kolokotronis, S., DeSalle, R., Planet, P.J., 2012. Random addition concatenation analysis: a novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. Genome Biol. Evol. 4, 30–43.
Nelson, G., 1979. Cladistic analysis and synthesis: principles and definitions, with a historical note on Adanson's Familles des Plantes (1763–1764). Syst. Zool. 28, 1–21.
Nikolaev, S., Montoya-Burgos, J.I., Margulies, E.H., NISC Comparative Sequencing Program, , Rougemont, J., Nyffeler, B., Antonarakis, S.E., 2007. Early history of mammals is elucidated with the ENCODE multiple species sequencing data. PLoS Genet. 3, e2.
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.
Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. Cladistics 12, 221–241.
Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.
Oaks, J.R., 2011. A time-calibrated species tree of Crocodylia reveals a recent radiation of the true Crocodiles. Evolution 65, 3285–3297.
O'Brien, S.J., Menotti-Raymond, M., Murphy, W.J., Nash, W.G., Wienberg, J., Stanyon, R., Copeland, N.G., Jenkins, N.A., Womack, J.E., Marshall Graves, J.A., 1999. The promise of comparative genomics in mammals. Science 286, 458–480.
O'Leary, M.A., 1999. Parsimony analysis of total evidence from extinct and extant taxa, and the cetacean-artiodactyl question. Cladistics 15, 315–330.
O'Leary, M.A., 2001. The phylogenetic position of cetaceans: further combined data analyses, comparisons with the stratigraphic record and a discussion of character optimization. Am. Zool. 41, 487–506.
O'Leary, M.A., Allard, M., Novacek, M.J., Meng, J., Gatesy, J., 2004. Building the mammalian sector of the tree of life. In: Cracraft, J., Donoghue, M.J. (Eds.), Assembling the Tree of Life. Oxford University Press, New York, pp. 490–516.
O'Leary, M.A., Gatesy, J., 2008. Impact of increased character sampling on the phylogeny of Cetartiodactyla (Mammalia): combined analysis including fossils. Cladistics 24, 397–442.
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., et al., 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. Science 339, 662–667.
Oliver, J.C., 2013. Microevolutionary processes generate phylogenomic discordance at ancient divergences. Evolution 67, 1823–1830.
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.
Pérez, M.E., Pol, D., 2012. Major radiations in the evolution of caviid rodents: reconciling fossils, ghost lineages, and relaxed molecular clocks. PLoS ONE 7, e48380.
Piertney, S.B., Oliver, M.K., 2006. The evolutionary ecology of the major histocompatibility complex. Heredity 96, 7–21.
Pisani, D., Wilkinson, M., 2002. Matrix representation with parsimony, taxonomic congruence, and total evidence. Syst. Biol. 51, 151–155.
Pollard, D., Iyer, V.N., Moses, A.M., Eisen, M.B., 2006. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS Genet. 2, e173.
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.
Ren, F., Tanaka, H., Yang, Z., 2005. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. Syst. Biol. 54, 808–818.
Rodrigo, A.G., 1993. A comment on Baum's method for combining phylogenetic trees. Taxon 42, 631–636.
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.
Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D.L., Rasnitsyn, A.P., 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst. Biol. 61, 973–999.
Rosenberg, N.A., 2013. Discordance of species trees with their most likely gene trees: a unifying principle. Mol. Biol. Evol. 30, 2709–2713.
Rosenberg, N.A., Tao, R., 2008. Discordance of species trees with their most likely gene trees: the case of five taxa. Syst. Biol. 57, 131–140.
Rosenfeld, J.A., Payne, A., DeSalle, R., 2012. Random roots and lineage sorting. Mol. Phylogenet. Evol. 64, 12–20.
Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331.
Sanderson, M.J., 1989. Confidence limits on phylogenies: the bootstrap revisited. Cladistics 5, 113–129.
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.
Simmons, M.P., 2012. Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. Cladistics 28, 208–222.
Simmons, M.P., Goloboff, P.A., 2013. An artifact caused by undersampling optimal trees in supermatrix analyses of locally sampled characters. Mol. Phylogenet. Evol. 69, 265–275.
Simmons, M.P., Norton, A.P., 2013. Quantification and relative severity of inflated branch-support values generated by alternative methods: an empirical example. Mol. Phylogenet. Evol. 67, 277–296.
Simmons, N.B., Geisler, J.H., 1998. Phylogenetic relationships of *Icaronycteris*, *Archaeonycteris*, *Hassianycteris*, and *Palaeochiropteryx* to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. Bull. Am. Mus. Nat. Hist. 235, 4–182.
Slowinski, J.B., Page, R.D.M., 1999. How should species phylogenies be inferred from sequence data? Syst. Biol. 48, 814–825.
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.
Spaulding, M., O'Leary, M.A., Gatesy, J., 2009. Relationships of Cetacea (Artiodactyla) among mammals: increased taxon sampling alters interpretations of key fossils and character evolution. PLoS ONE 4, e7062.

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013

Springer, M.S., de Jong, W.W., 2001. Which mammalian supertree to bark up? Science 291, 1709–1711.
Springer, M.S., Gatesy, J., 2014. Land plant origins and coalescence confusion. Trends Plant Sci. 19, 267–269.
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., 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.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Struck, T.H., Purschke, G., Halanych, K.M., 2006. Phylogeny of Eunicida (Annelida) and exploring data congruence using a partition addition bootstrap alteration (PABA) approach. Syst. Biol. 55, 1–20.
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.
Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50, 525–539.
Townsend, T.M., Mulcahy, D.G., Noonan, B.P., Sites, J.W., Kuczynski, C.A., Wiens, J.J., Reeder, T.W., 2011. Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. Mol. Phylogenet. Evol. 61, 363–380.
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.
Underhill, P.A., Kivisild, T., 2007. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu. Rev. Genet. 41, 539–564.
Wahlberg, N., Braby, M.F., Brower, A.V.Z., de Jong, R., Lee, M.-M., Nylin, S., Pierce, N.E., Sperling, F.A.H., Vila, R., Warren, A.D., Zakharov, E., 2005. Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. Proc. Biol. Sci. 272, 1577–1586.
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.
Wolsan, M., Sato, J.J., 2010. Effects of data incompleteness on the relative performance of parsimony and Bayesian approaches in a supermatrix phylogenetic reconstruction of Mustelidae and Procyonidae (Carnivora). Cladistics 26, 168–194.
Wu, S., Song, S., Liu, L., Edwards, S., 2013. Reply to Gatesy and Springer: the multispecies coalescent model can effectively handle recombination and gene tree heterogeneity. Proc. Natl. Acad. Sci. USA 110, E1180.
Xi, Z., Rest, J.S., Davis, C.C., 2013. Phylogenomics and coalescent analyses resolve extant seed plant relationships. PLoS ONE 8, e80870.
Yu, Y., Warnow, T., Nakhleh, L., 2011. Algorithms for MDC-based multi-locus phylogeny inference: beyond rooted binary gene trees on single alleles. J. Comput. Biol. 18, 1543–1559.
Zhong, B., Liu, L., Yan, Z., Penny, D., 2013. Origin of land plants using the multispecies coalescent model. Trends Plant Sci. 18, 492–495.
Zhong, B., Liu, L., Penny, D., 2014. The multispecies coalescent model and land plant origins: a reply to Springer and Gatesy. Trends Plant Sci. 19, 270–272.
Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion. Ph.D. Dissertation, The University of Texas at Austin.
Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588–598.

Glossary

anomaly zone: a set of conditions wherein the most probable gene tree conflicts with the species tree

coalescence gene (c-gene): a segment of the genome that has not recombined over the phylogenetic history of a clade

Concatalescence: concatenation of multiple 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)

hidden support: increased support for a clade in combined analysis of all data partitions relative to the sum of support evident in separate analyses of the individual data partitions

shortcut coalescence method: a coalescence method that does not co-estimate gene trees with the species tree

Please cite this article in press as: Gatesy, J., Springer, M.S. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. (2014), http://dx.doi.org/10.1016/j.ympev.2014.08.013