Syst. Biol. 70(4):803–821, 2021

DOI:10.1093/sysbio/syaa097
Advance Access publication December 26, 2020

Gene Tree Estimation Error with Ultraconserved Elements: An Empirical Study on *Pseudapis* Bees

SILAS BOSSERT[1,2,3,*], ELIZABETH A. MURRAY[2,3], ALAIN PAULY[4], KYRYLO CHERNYSHOV[5], SEÁN G. BRADY[2], AND BRYAN N. DANFORTH[1]

[1]*Department of Entomology, Cornell University, Comstock Hall, Ithaca, NY 14853, USA;* [2]*Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA;* [3]*Department of Entomology, Washington State University, Pullman, Washington 99164, USA;* [4]*O.D. Taxonomy and Phylogeny, Royal Belgian Institute of Natural Sciences, Rue Vautier 29, 1000 Brussels, Belgium;* [5]*College of Arts and Sciences, Cornell University, Ithaca, NY 14853, USA*

**Correspondence to be sent to: Department of Entomology, Cornell University, Comstock Hall, Ithaca, NY 14853, USA; E-mail: sb2346@cornell.edu*

Received 5 November 2019; reviews returned 18 November 2020; accepted 2 December 2020
Associate Editor: Claudia Solis-Lemus

Abstract.—Summarizing individual gene trees to species phylogenies using two-step coalescent methods is now a standard strategy in the field of phylogenomics. However, practical implementations of summary methods suffer from gene tree estimation error, which is caused by various biological and analytical factors. Greatly understudied is the choice of gene tree inference method and downstream effects on species tree estimation for empirical data sets. To better understand the impact of this method choice on gene and species tree accuracy, we compare gene trees estimated through four widely used programs under different model-selection criteria: PhyloBayes, MrBayes, IQ-Tree, and RAxML. We study their performance in the phylogenomic framework of >800 ultraconserved elements from the bee subfamily Nomiinae (Halictidae). Our taxon sampling focuses on the genus *Pseudapis*, a distinct lineage with diverse morphological features, but contentious morphology-based taxonomic classifications and no molecular phylogenetic guidance. We approximate topological accuracy of gene trees by assessing their ability to recover two uncontroversial, monophyletic groups, and compare branch lengths of individual trees using the stemminess metric (the relative length of internal branches). We further examine different strategies of removing uninformative loci and the collapsing of weakly supported nodes into polytomies. We then summarize gene trees with ASTRAL and compare resulting species phylogenies, including comparisons to concatenation-based estimates. Gene trees obtained with the reversible jump model search in MrBayes were most concordant on average and all Bayesian methods yielded gene trees with better stemminess values. The only gene tree estimation approach whose ASTRAL summary trees consistently produced the most likely correct topology, however, was IQ-Tree with automated model designation (ModelFinder program). We discuss these findings and provide practical advice on gene tree estimation for summary methods. Lastly, we establish the first phylogeny-informed classification for *Pseudapis* s. l. and map the distribution of distinct morphological features of the group. [ASTRAL; Bees; concordance; gene tree estimation error; IQ-Tree; MrBayes, Nomiinae; PhyloBayes; RAxML; phylogenomics; stemminess]

Accurate gene trees are critical for species tree reconstruction through gene tree summary methods. Also known as two-step coalescent approaches, summary methods infer species trees under the multispecies coalescent model (MSC; Rannala and Yang 2003), relying on *a priori* generated gene trees as input. In contrast to the traditionally used concatenation-approach, for which multiple gene sequences are concatenated and analyzed in a supermatrix, coalescent-based methods incorporate information on the individual evolutionary history of each locus and can therefore specifically model discordance between gene trees and species trees caused by incomplete lineage sorting (ILS; e.g., Maddison 1997; Kubatko and Degnan 2007; Degnan and Rosenberg 2009; Mirarab 2019, and references therein).

Gene trees are treated as independent but fixed observations by contemporary summary methods (Xu and Yang 2016). Methods to estimate gene trees, information about the confidence of topological placements (branch support), or branch lengths are generally not accounted for. In principle, gene tree topology provides the sole source of information, rendering gene trees the "Achilles' Heel" of coalescent-based summary methods by critical voices (Springer and Gatesy 2016). With coalescent approaches being regarded by some in the field as paradigm shifting (e.g., Edwards 2009; Edwards et al. 2016; Bravo et al. 2019), and being emphatically challenged by others (e.g., Gatesy and Springer 2013, 2014; Springer and Gatesy 2016), the need to understand the effects of gene tree accuracy on species tree inference has sparked significant research over the past years.

A greatly understudied aspect of summary approaches is the way gene trees are estimated before summarizing them. While an extensive body of theoretical studies (e.g., Bayzid and Warnow 2013; Patel et al. 2013; Mirarab et al. 2014b; Bayzid et al. 2015; Roch and Steel 2015) and empirical research (Blom et al. 2016; Hosner et al. 2016; Meiklejohn et al. 2016; Sayyari et al. 2017) showed that gene tree estimation error (GTEE) decreases the accuracy of summary methods, we have limited knowledge of how different methods of gene tree inference compare. Few studies used more than one method of gene tree reconstruction and even fewer compared their performance. Two of these (Xi et al. 2015; Sayyari et al. 2017) found maximum likelihood (ML) estimates of RAxML (Stamatakis 2014) more accurate than those of PhyML (Guindon et al. 2010) or the speed-optimized FastTree

(Price et al. 2010), respectively. Similarly, Zhang et al. (2018) found that IQ-Tree gene tree estimates had consistently better likelihood values over RAxML/ExaML, PhyML, and FastTree (in decreasing order) in an extensive assessment of empirical data sets. Another empirical study on birds assessed four tree-building methods (RAxML, GARLI, PhyML, and MrBayes) in their ability to infer a predefined clade, and found Bayesian gene trees to be slightly more accurate topologically than trees obtained by ML (Meiklejohn et al. 2016; which is line with observations of Mirarab 2019). However, differences in branch lengths remain unstudied, and all these studies applied site-homogeneous substitution models partitioned by entire gene alignments.

In order to better understand the impacts of method choice on gene and species tree estimation, our study compares a set of widely used algorithms that significantly differ in their underlying statistical frameworks, implementation of substitution models, and computational demands. Specifically, we compare the ML implementations RAxML (ver. 8, Stamatakis 2014) and IQ-Tree (ver. 2, Minh et al. 2020) under different model-selection strategies, as well as the Bayesian programs MrBayes (Ronquist et al. 2012) and PhyloBayes (Lartillot et al. 2009; Lartillot et al. 2013). The latter program is particular because it implements the site-heterogenous mixture model CAT-GTR (Lartillot and Philippe 2004; Lartillot et al. 2009). In contrast to previous research, we assess and compare both differences in topology and branch lengths of gene trees inferred with different methods, and further examine the resulting species tree inferences obtained through the popular summary method ASTRAL.

As a study system, we investigate the challenging and yet unexplored phylogenomic landscape of ultraconserved elements (UCEs; Faircloth et al. 2012) from the subfamily Nomiinae (Halictidae). These bees represent a major group in the Old World tropics and comprise over 600 described species (Ascher and Pickering 2020), yet their phylogeny has never been the focus of a comprehensive molecular phylogenetic study (Danforth et al. 2012). A thoroughly developed phylogenetic hypothesis is needed to evaluate the competing morphology-based classifications of the group and would greatly inform the process of settling on a rank-based taxonomy of monophyletic groups. UCEs render our research applicable to contemporary studies, as UCEs are increasingly applied all across the metazoan Tree of Life, and best practices are being progressively refined (e.g., Portik and Wiens 2020; Van Dam et al. 2020). They further provide a particularly interesting framework for examining GTEE, as they are short with few informative sites (see also Molloy and Warnow 2018), and nucleotide substitution rates are heterogeneously distributed with conserved core and variable flanking regions (Faircloth et al. 2012; Smith et al. 2014). This could impact model fit, since model parameters are estimated over very differently conserved nucleotide sequences (Zhang et al. 2018; Tagliacollo and Lanfear 2018). The research presented herein provides insights on choosing different gene tree estimation methods and consequences for gene tree topology and branch lengths. Ultimately, we show that gene trees inferred with Bayesian methods perform better in recovering a set of uncontroversial, predefined clades, and have greater relative length of internal branches than gene trees inferred with maximum likelihood methods. This greater topological concordance, however, does not always translate into more accurate summary trees (i.e., similarity to the species tree that we deem correct), and summary tree accuracy improves for all methods when loci with poor information content are removed.

Materials and Methods

Taxon Sampling

In order to establish a test data set with closely and distantly related lineages of nomiine bees, we focused on a readily identifiable group, the genus *Pseudapis*. Specifically, we compiled a comprehensive taxon sampling of 20 species of the *Pseudapis* group, including all six lineages that have been considered genera or subgenera in previous classifications of this group (Table 1). This represents about one-third of all described taxa of *Pseudapis* s. l. We further included 10 lineages of major nomiine groups from four continents, among them *Austronomia* (Australia, Africa, Asia), *Acunomia* (North America, Africa, Asia), *Lipotriches* (Africa, Asia), and *Macronomia* (Africa, Asia). Additionally, we included UCE sequence data from the currently available genomes of Halictidae: *Lasioglossum albipes*, *Dufourea novaeangliae*, and *Acunomia melanderi* (Kocher et al. 2013; Kapheim et al. 2015; Kapheim et al. 2019).

Acquisition and Processing of UCEs

A detailed version of the applied molecular methods and the bioinformatic processing, including a step-by-step documentation of all used commands, is available in the Supplementary Material available on Dryad at http://dx.doi.org/10.5061/dryad.z08kprrb6. Briefly, we followed a standard Proteinase-K based phenol/chloroform protocol (from Saghai-Maroof et al. 1984, modified) to extract DNA from the hindlegs of specimens or by soaking the whole specimen in extraction buffer. DNA was quantified, sheared, and prepared with dual-indexed libraries and TruSeq adapters (Glenn et al. 2019). We followed the protocol of Faircloth et al. (2015) and the modifications from Blaimer et al. (2016a,b). Target enrichment for 29 samples was conducted with the enhanced HymV2-ant UCE probe-set or the principal HymV2 set (Branstetter et al. 2017b), both manufactured by MYcroarray, Inc. (see Supplementary Material available on Dryad). Enrichment success was assessed with RT-qPCR. After

TABLE 1. Taxon sampling

Species	Collection locality	No. of captured UCE loci	NCBI SRA ID
Acunomia melanderi (Cockerell 1906)	USA: WA, Walla Walla Co., Touchet; see Kapheim et al. (2019)	2359	n/a
Afronomia circumnitens (Cockerell 1946)	South Africa: Mpumalanga Province, 7 km S. Graskop	1518	SRR7970542
Austronomia australica (Smith 1875)	Australia: South Australia, Jervois Co., Cowell	1538	SRR7970543
Curvinomia chalybeata (Smith 1875)	Vietnam: Cát Bà Island, Cat Ba National Park	1560	SRR7970540
Dieunomia (Dieunomia) heteropoda (Say 1824)	USA: Nebraska, Cherry CO.	1922	SRR7970541
Dieunomia (Epinomia) triangulifera (Vachal 1897)	USA: Kansas, Douglas Co., Lawrence	1784	SRR7970538
Dufourea novaeangliae (Robertson 1897)	USA: NY, Cayuga Co., Fair Haven Beach; Kapheim et al. (2015)	2343	n/a
Hoplonomia elliotii (Smith 1875)	Cambodia: Siem Reap Province	1632	SRR7970539
Lasioglossum albipes (Fabricius 1781)	France: Mt. Ventoux; see Kocher et al. (2013)	2420	n/a
Lipotriches (Patellotriches) collaris (Vachal 1903)	South Africa: Limpopo Prov. 75 km SW Thabazimbi	1621	SRR7970536
Lipotriches (Stellotriches) justiciae (Pauly 2014)	South Africa: Limpopo Prov., 8.5 km N. Vivo	1723	SRR7970537
Macronomia clavisetis (Vachal 1910)	Ethiopia: Oromia Region, Koka	1615	SRR7970534
Nomiapis bispinosa (Brullé 1832)	Spain: Almeria Province	1639	SRR7970535
Nomiapis diversipes (Latreille 1806)	Spain: Granada Province	1585	SRR7970530
Pachynomia amoenula (Gerstäcker 1870)	South Africa: KwaZulu-Natal, Kwangwanase	1736	SRR7970531
Pachynomia flavicarpa (Vachal 1903)	Cameroon: Adamawa Province, Meiganga	1094	SRR7970528
Pachynomia tshibindica (Cockerell 1935)	Burundi: Kayanza Province, Kibira	1642	SRR7970529
Pseudapis nilotica (Smith 1875)	UAE: Dubai, Nakhalai	967	SRR7970526
Pseudapis cinerea (Friese 1930)	South Africa: N. Cape Province, 4 km NW Hotazel	1499	SRR7970527
Pseudapis flavolobata (Cockerell 1911)	India: Rajasthan, Jaisalmer District	1614	SRR7970524
Pseudapis interstitinervis (Strand 1912)	Kenya: Great Rift Valley, Kajiado Co., Olorgesailie	1297	SRR7970525
Pseudapis kenyensis (Pauly 1990)	Kenya: Great Rift Valley, Kajiado Co., Olorgesailie	1444	SRR7970532
Pseudapis oxybeloides (Smith 1875)	Pakistan: Punjab, University of Agriculture Faisalabad	1526	SRR7970533
Pseudapis pandeana (Strand 1914)	Kenya: Great Rift Valley, Baringo Co., Mogotio	1738	SRR7970552
Pseudapis riftensis (Pauly 1990)	Kenya: Great Rift Valley, Kajiado Co., Olorgesailie	1454	SRR7970551
Pseudapis siamensis (Cockerell 1929)	Laos: Vientiane Pref., Tat Mun Waterfall	1695	SRR7970550
Ruginomia rugiventris (Friese 1930)	South Africa: KwaZulu-Natal, Maputaland Tembe Elephant Park	863	SRR7970549
Steganomus ennediensis (Pauly 1990)	Niger: Zinder Region, 49 km NW Tanout	1142	SRR7970548
Steganomus junodi (Gribodo 1895)	Ghana: Cape Coast, Univ. Cape Coast campus	1403	SRR7970547
Stictonomia aliceae (Cockerell 1935)	South Africa: Limpopo Province, 29 km NW Waterpoort	1449	SRR7970546
Stictonomia sangaensis (Pauly 1990)	Cameroon: Adamawa Province, Meiganga	1516	SRR7970545
Stictonomia schubotzi (Strand 1911)	Gabon: Ogcoué-Ivindo Dist.	550	SRR7970544

The scientific names of the included species and the respective collection localities. The classification follows the work of Pauly (1990–2014). Assembled UCE sequence data and the nucleotide matrices are available from our Dryad repository (http://dx.doi.org/10.5061/dryad.z08kprrb6).

inferring concentrations for each pool, they were combined at equimolar concentrations and size-selected for fragments between 250 bp and 800 bp. Sequencing was conducted with an Illumina HiSeq 2500 device and 150 bp paired-end reads at the Cornell Biotechnology Resource Center (Cornell BRC).

Raw reads were trimmed with the Trimmomatic (Bolger et al. 2014) wrapper Illumiprocessor (Faircloth 2013), followed by read assessments with FastQC (Andrews 2019). Reads were assembled with Trinity (Grabherr et al. 2011), as called from Phyluce (Faircloth 2016) under the default settings. We used LASTZ (Harris 2007) to query the three included genomes for regions that correspond to the UCE probe-set and used Phyluce to parse the results into a sqlite database. We then extracted the matches with 850 bp of flanking regions up- and downstream of the identified UCE core and treated the extracted sequences as Trinity assemblies.

The genome extracts and the assemblies of the de novo sequenced UCEs were then matched against the probe set. In order to exclude potentially contaminating sequences (Bossert and Danforth 2018), we required a minimum sequence overlap of 80% with at least 85% sequence identity. UCE matches were then aligned per locus, that is, each locus representing an individual alignment, using the L-INS-i algorithm of MAFFT v7.31 (Katoh and Standley 2013). We allowed loci to be incomplete and trimmed each locus with Gblocks (Castresana 2000) according to the "relaxed" conditions of Talavera and Castresana (2007). After trimming, we finalized a sequence matrix with 80% completeness, ensuring that every locus is represented by at least 80% of all taxa (=25 taxa). Lastly, individual alignments were examined by eye to identify and remove potential misalignments.

Phylogenetic Reconstruction

Phylogenetic reconstructions were carried out using Bayesian and maximum likelihood methods on the concatenated matrices, as well as summarizing gene trees under the multispecies coalescent model (Rannala and Yang 2003).

Gene Tree Inferences and the PhyloBayes Wrapper EZ-PB.— We generated six different sets of gene tree estimates, three using maximum likelihood methods and three using Bayesian inference. Each method was carried

out on the exact same individual alignments, which were partitioned by locus. For the maximum likelihood estimates, we used RAxML (ver. 8, Stamatakis 2014) to find the best-scoring ML tree under the GTR+G model and 200 bootstrap replicates. We then used IQ-Tree (ver. 2, Minh et al. 2020) under the same substitution model, discrete gamma, and 1000 bootstrap approximations (UFBoot2; Hoang et al. 2018). In order to compare the effects of choosing an automated approach to select substitution models, we further generated a set of gene trees by using IQ-Tree and the ModelFinder program ("MFP", Kalyaanamoorthy et al. 2017), to search for the best-fitting substitution models, including tests for optimal rate heterogeneity across sites (FreeRate model). To compare these sets of gene trees to Bayesian estimates, we used MrBayes (ver. 3.2.7, Ronquist et al. 2012) on the individual alignments using GTR+G. We executed two runs with each two chains and used the option to automatically stop analyses once the average standard deviations of split frequencies (ASDSF) were $\leq$0.01. We then repeated the MrBayes analyses with the reversible jump (Huelsenbeck et al. 2004) option, which is a model-averaging approach that allows the chains to sample the parameter space across all possible models that MrBayes can accommodate. The respective MrBayes blocks that were used to estimate the MrBayes gene trees are provided in the Supplementary Material available on Dryad. Lastly, in order to infer UCE gene trees with a substitution model that allows for among-site heterogeneity of substitution rates, we employed the Bayesian implementation PhyloBayes (Lartillot et al. 2009; Lartillot 2013), which implements infinite site-heterogeneous mixture-models (CAT/CAT-GTR; Lartillot and Philippe 2004). However, PhyloBayes is difficult to execute on large numbers of separate alignments, such as individual UCE loci. To facilitate this task, we developed the wrapping software EZ-PB, which is an easy-to-use Python wrapper around the PhyloBayes package. As such, the script itself does not reconstruct phylogenies but executes the parallel version of PhyloBayes and its diagnostic tools according to specified parameters. Briefly, the script executes the following tasks sequentially on each alignment in a set folder: (i) execute a desired number of chains, (ii) automatically check for sufficient sampling and convergence between chains until the values fall below the specified thresholds or if the specified maximum number of cycles is reached; then terminate the chains, (iii) organize and name "good" (i.e., chains have converged) and "bad" (i.e., chains have not converged) consensus trees and associated analyses files based on convergence criteria of the respective chains, and (iv) summarize the results in a spreadsheet. Parameters under which PhyloBayes is executed through EZ-PB are adjustable with a configuration file. The default parameters were designated according to the PhyloBayes authors' recommendation for very good runs: maximum discrepancies between bipartitions (maxdiff) of <0.1 (using *bpcomp*), effective sample sizes (ESS) of the log-likelihood of >300, and a relative log-likelihood differences of <0.1 (using *tracecomp*). To generate gene trees for the present study, we used the default settings and two chains.

Concatenation-Based Methods.—First, we calculated an ML tree using IQ-Tree 2 (Minh et al. 2020) and the concatenated 80% matrix. We partitioned the matrix by locus, used the ModelFinder search to assign substitution models, designated edges to be linked, and calculated 1,000 bootstrap approximations. Using the previously estimated IQ-Tree gene trees (MFP), we estimated gene concordance factors (gCF) and site concordance factors (sCF), and mapped them onto the species tree. We also calculated the best-scoring ML tree using RAxML and 200 rapid bootstrap replicates ("-f a" option). Lastly, we executed three independent PhyloBayes chains on the 80% matrix using the parallel version of the program. We used the CAT-GTR model, ran at least 10,000 cycles for each chain, discarded 100 cycles as burn-in, and sampled every other tree. We considered convergence to be sufficient when the maximum and mean differences between bipartitions reached 0.0 and the relative differences of the log-likelihoods were $\leq$0.25 with ESS values being >250. We failed to achieve parameter convergence when running MrBayes on the partitioned concatenated supermatrix.

Comparing and Evaluating UCE Gene Trees

In order to assess different properties of the calculated gene trees, we applied a range of different measures (summarized in Fig. 1, red box). First, we approximated GTEE by using a measure of *concordance* for each individual gene tree, as implemented in the R package treespace (Jombart et al. 2017). The main advantage of this approach is that it allows us to compare gene trees without the same set of tips, as not a single one of the examined gene trees includes all taxa. This measure of gene tree concordance is not to be confused with the concordance factors that we estimated with IQ-Tree, which we only used to provide additional measures of branch support for the concatenation analyses. Instead, this measure of gene tree concordance, which is described in detail in Kendall et al. (2018), assesses how well a given gene tree fulfills a set of predefined, monophyletic groups, which are treated as reference categories. The measure is on a scale from 0 to 1, where 1 indicates a gene tree that is fully compatible with a reference tree, which consists only of the tips representing the test categories. As categories, we designated two uncontroversial clades: the genus *Dieunomia*, which represents the earliest branching lineage of Nomiinae, and the remaining Nomiinae. Monophyly of these nomiine lineages has been recovered in every species tree analyses in the present study and in virtually all previously conducted molecular research that involved these groups (e.g., Danforth et al. 2004; Hedtke et al. 2013; Cardinal et al. 2018). Internode branch

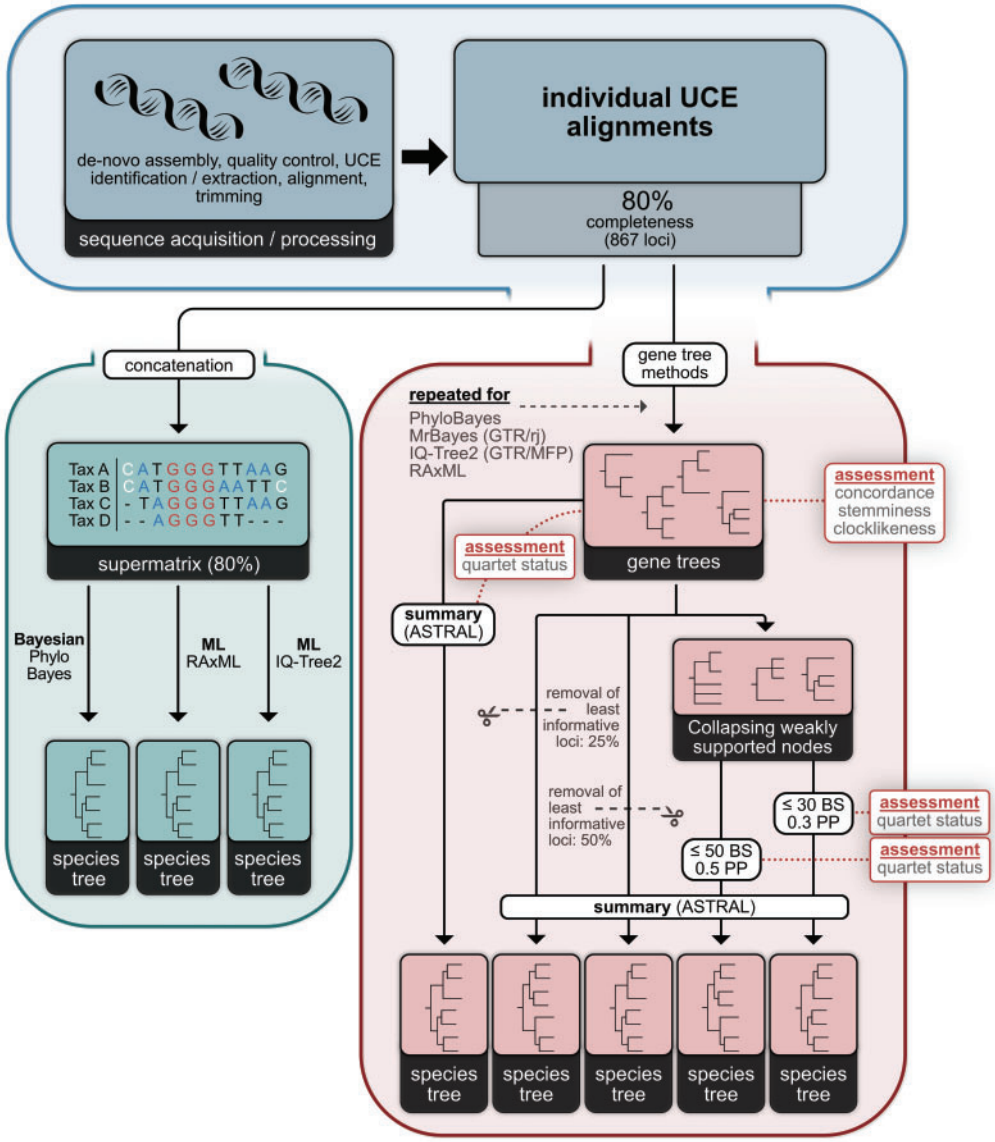

FIGURE 1. Summary of the developed workflow.

lengths leading to these clades are generally relatively long, indicating that these nodes are likely not affected by ILS. We also included the outgroup as third category and rooted every gene tree on both outgroup taxa, if present. We rooted on only one outgroup taxon if only one was present and rooted on *Lasioglossum albipes* if the outgroups were not sistergroups.

Second, we assessed the quartet status of the individual gene trees in respect to the summary tree inferred from all gene trees of a reconstruction method. This means, for example, that we first summarized all RAxML gene trees with ASTRAL-III (Zhang et al. 2018) and then assessed the quartet status of each individual RAxML gene tree in respect to this summary tree. Quantifying quartets of gene trees allowed us to measure (i) quartets that are identical in both trees, (ii) quartets that are conflicting, and (iii) quartets that are not resolved in the gene tree. Assessing unresolved quartets allows us to better understand the relationship of gene tree accuracy and resolution and represents a major advantage over other measures of tree similarity when evaluating gene trees with polytomies (see Smith 2019 for further discussion). We quantified quartets for individual gene trees using the R package Quartet (Smith 2020) and normalized over the sum of all quartets in each given gene tree. This assessment of quartet status was repeated for every gene tree estimation method and under three regimes of collapsing weakly supported nodes: (i) no nodes being collapsed to polytomies, (ii) retracting nodes of ≤30 bootstrap support or ≤0.3

posterior probability, and (iii) retracting nodes of ≤50 BS or ≤0.5 PP (Fig. 1).

In order to understand differences in branch lengths of different gene tree estimation methods, we calculated the stemminess metric (Fiala and Sokal 1985; Rohlf et al. 1990). Stemminess is a measure of branch lengths that quantifies the relative length of internal branches of a phylogeny. Low stemminess means that trees have a greater proportion of overall length concentrated along with the tips. Long terminal branches can be indicative of saturation and can suffer from long-branch attraction (Longhorn et al. 2010), which is why high values for stemminess are favorable. Stemminess has previously been used as a proxy for tree quality, assuming that higher stemminess confers a better signal-to-noise ratio (Kück et al. 2012) and that reconstruction artifacts are reduced. The stemminess metric used here is a variation of the cumulative stemminess function (from Fiala and Sokal 1985) that has been more widely used recently (e.g., Longhorn et al. 2010; Kück et al. 2012; Tong et al. 2018). We report stemminess as the proportion, from 0 to 1, of the total tree length represented by internal (nonterminal) branches. Stemminess values for all gene trees from the same estimation method were compared and the performance of each method was ranked from 1 (most stemmy) to 6 (least stemmy) for every individual UCE. We also tested the molecular clock hypothesis for each individual gene tree using an R script from the repository of Borowiec et al. (2015, see Supplementary Material available on Dryad).

Lastly, we ranked gene trees of each method based on a set of tree and alignment characteristics of the underlying UCEs: average node support values (bootstrap values or posterior probabilities), alignment length, proportion of variable sites, and % missing data. We then generated subsets of the 25 and 100 UCEs that performed best in each category and explicitly examined the performance of these subsets in comparison to all loci.

Gene Tree Summary Analyses

Summary analyses under the multispecies coalescent model were carried out with ASTRAL-III (Zhang et al. 2018) and default settings. In order to examine topological differences of summary trees calculated with different sets of informative loci, we calculated ASTRAL trees for every method and three sets of loci with different degrees of informative sites. To this end, we filtered alignments for the 25% and 50% of loci with the lowest proportion of variable sites and summarized these subsets of loci into species trees, as well as summarizing all loci. The resulting gene trees were compared pairwise using the Kendall and Colijn metric (Kendall and Colijn 2016) and resulting values were hierarchically clustered using Ward's criterion (D2; see Murtagh and Legendre 2014).

Results

UCE Capture Success

We sequenced a total of ~54 million reads for 29 samples, with an average of ~1,869,572 (min. 165,217–max. 3,868,765) reads per taxon. Assembling these reads yielded Trinity assemblies with an average of 58,834 (2407–176,652) contigs per sample. We extracted an average of 2374 (2343–2420) UCE loci from the three genomes and captured between 550 and 1922 UCEs from our de novo sequenced samples (average of 1474; Table 1). The generated 80% matrix had the following properties: 867 loci, 576,041 bp length, 26.8% missing nucleotides.

Phylogeny of Pseudapis *s.l.*

Species trees inferred in this study are generally congruent except for three areas on the tree (Fig. 2, nodes A–C). The maximum likelihood estimates and Bayesian inference of the concatenated supermatrix converged on the same topology. The concatenation species trees are unambiguously supported by very high bootstrap values or posterior probabilities, but the three problematic areas have very low gene concordance factors (gCF). Particularly node A, which is further characterized by very short internodes, has gene concordance factors of less than 20.

Topological conflict is caused by different gene tree summary analyses. Two of the conflicting nodes (Fig. 2, nodes B–C) are parents to terminals with very low information content, whereas the third node (A) involves a deeper split with very short internodes and short coalescent times (Supplementary Figs. S1 and S2 available on Dryad), but good sequence representation. ASTRAL summaries solve these nodes with five different alternative topologies summarized in Figure 3.

Gene Tree Inferences and Gene Tree Properties

For 867 UCE alignments, PhyloBayes chains of 853 runs sampled the parameter space sufficiently and converged in less than 30,000 cycles. We therefore used only those 853 gene trees for the comparisons among all gene tree inference approaches and gene tree summaries and discarded the remaining 14 loci for downstream comparisons and analyses across methods.

Concordances for solving the three predefined clades are similar among methods but are greater for gene trees inferred with Bayesian methods (Fig. 4, Supplementary Fig. S3A available on Dryad). In 495 of 853 cases, concordance values of all methods were identical, but at least one method differed in the remaining 358 examined sets of gene trees. Among Bayesian methods, gene trees inferred through MrBayes with the reversible jump model search are most concordant on average, and rank most often as first, closely followed by MrBayes (GTR+G) and PhyloBayes (Fig. 4, Supplementary Fig. S3A available on Dryad). Among ML methods, IQ-Tree with automated model selection

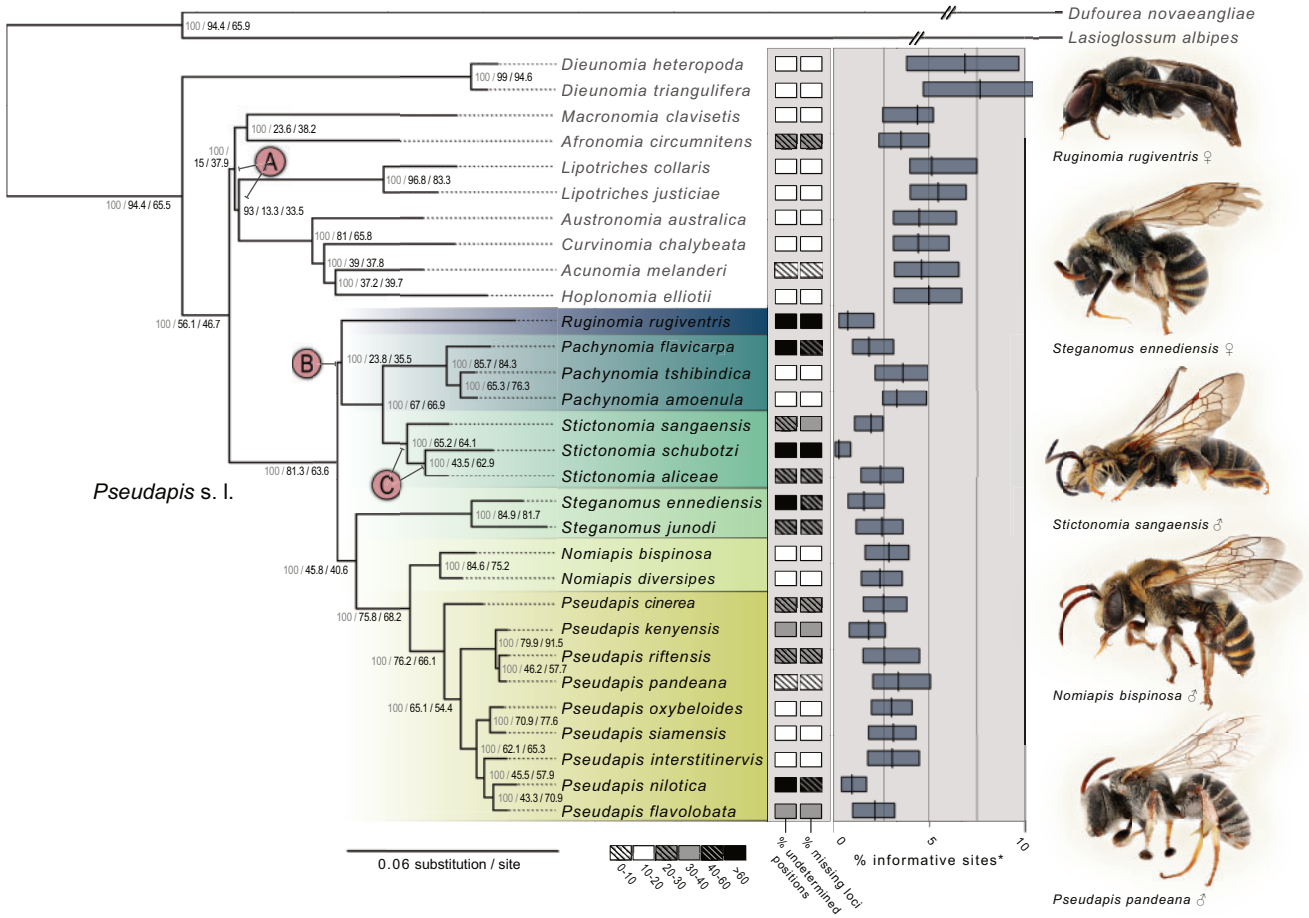

FIGURE 2. PhyloBayes consensus tree of *Pseudapis* s. l. and closely related Nomiinae based on 867 concatenated ultraconserved elements. The tree is topologically equivalent to the consensus tree of the IQ-Tree 2 ML analysis, the best-scoring tree of the RAxML run. Node support corresponds to 1.0 posterior probability on all nodes for PhyloBayes, while support values shown are (from left to right): UFBootstrap values, gene concordance factors (gCF) and site concordance factors (sCF). Circled letters A–C indicate nodes that are conflicting with at least one summary analysis conducted in this study. Percentage of informative sites (indicated by *) was calculated with DIVEIN (Deng et al. 2010) using all loci represented by at least 29 taxa (=90%).

(MFP) performed best on average. This makes both approaches of automated model selection (reversible jump for MrBayes and MFP for IQ-Tree, respectively) preferable over their GTR+G counterparts. Gene trees inferred with RAxML were the least concordant and rank lower than gene trees calculated with other methods.

The quartet assessment of the uncollapsed gene trees shows very similar quartet status for the different gene tree methods and reveals the greatest topological dissimilarity between gene trees and their summary trees in UCE loci with few informative sites (Fig. 5, Supplementary Figs. S4 and S5 available on Dryad). This trend is most clear for the first 20% of the least informative loci, whereas the remaining bins are similar in their overall quartet status. The different methods, however, respond differently to collapsing bifurcations with low node support to polytomies. Gene trees inferred with IQ-Tree have lower proportions of unresolved quartets and have relatively more shared and conflicting quartets than other methods (Supplementary Figs. S4 and S5 available on Dryad). This shows that nodes of IQ-Tree gene trees are less often collapsed and therefore have generally higher node confidence values than other methods. For the two bins of least informative loci, collapsing poorly supported nodes reduces the amount of conflicting quartets slightly more than the amount of shared quartets. This means that more conflicting quartets are transformed into unresolved quartets than shared quartets, which is desirable for subsequent summary analyses.

Branch lengths of gene trees inferred with different methods differ considerably. Generally, Bayesian gene trees are substantially stemmier than their ML counterparts (Fig. 6, Table 2, Supplementary Fig. S3B available on Dryad). Their stemminess is higher on average and they rank higher compared to ML gene trees when comparing gene trees inferred from the same alignments. In only 6 out of 853 sets of gene trees was

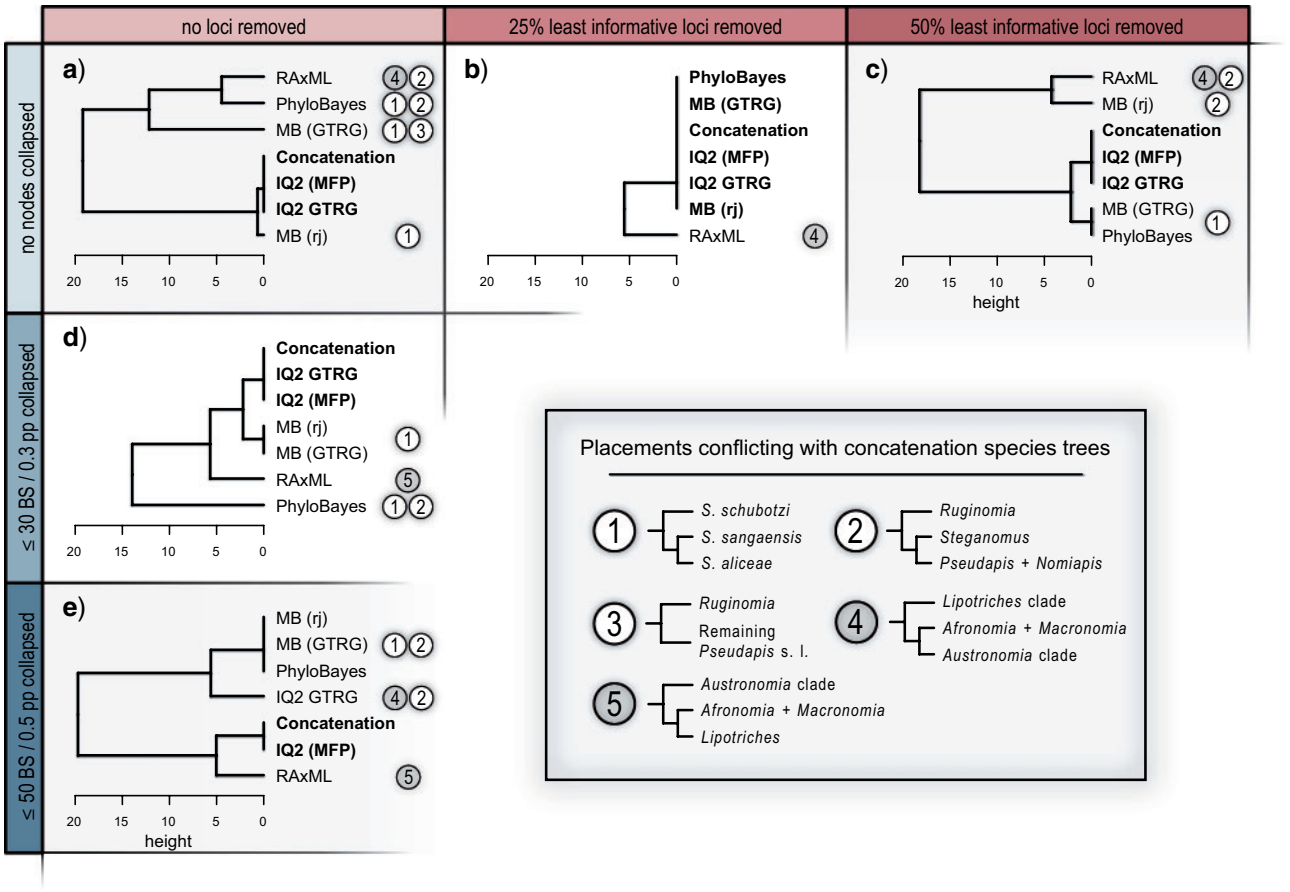

FIGURE 3. Topological differences of the 30 ASTRAL summary analyses conducted in this study. Dendrograms of species tree topologies are based on pairwise Kendall & Colijn distances, hierarchically clustered using Ward's criterion (D2). Leaf labels (except concatenation) correspond to the method used to estimate the gene trees that were given as input trees for ASTRAL summary analyses. For example, subfigure b) shows that the concatenation analyses and all but one of the ASTRAL summary trees (the one estimated from RAxML gene trees) recovered identical topologies (i.e., Kendall–Colijn distances of 0); the ASTRAL tree estimated from RAxML gene trees differs from the other trees by a different placement of the *Lipotriches* clade, as indicated by placement (4). The concatenation tree topology corresponds to Figure 2 and was not included in the locus removal and node collapsing experiments. Analyses in bold produced the most likely true species tree topology. Topological differences indicated by white circles involve samples with low sequencing success, whereas gray circles indicate conflict in the resolution of a deeper node with good sequence representation. rj = reversible jump; MFP = ModelFinder Plus; GTRG, generalized time-reversible model with discrete gamma.

the stemmiest gene tree inferred with an ML method. While there are substantial differences between Bayesian and ML methods, differences among only Bayesian or only ML methods are insignificant on the continuous scale from 0 to 1 (Fig. 6). Counting how different gene trees rank according to their stemminess on a locus-by-locus basis, however, shows that PhyloBayes gene trees are most stemmy in ~60% of all cases (509 out of 853, Supplementary Fig. S3B available on Dryad), followed by MrBayes (rj) and MrBayes (GTR+G). Testing the molecular clock hypothesis for the individual gene trees revealed that the Bayesian gene trees behave more similar to an ultrametric tree than the ML estimates, with MrBayes (rj) gene trees being most clock-like on average (Fig. 6, Table 2).

The 100 and 25 loci with the highest average bootstrap support or posterior probabilities have a greater concordance, are stemmier, and behave more clock-like than the average across all loci (Fig. 6, Table 2). They further perform better than the other examined subsets of loci, such as the longest, the least gappy, and the most variable UCEs. While the values of the different node support subsets are generally very similar, they usually rank in decreasing order, with the 25 highest scoring loci being slightly favorable over the 100 loci subsets. They further perform better than the longest, the least gappy, and the most variable UCEs.

The total computational wall times differ substantially between different approaches of gene tree estimation, and PhyloBayes is by far the most time-intensive strategy (Fig. 4). For inferring the PhyloBayes gene trees with EZ-PB on a local workstation (Dell Precision with Xeon® E5-2687W v4 @ 3.00 GHz; 64 GB DDR4 SDRAM), the total wall time summed up to 754.5 h (~31 days),

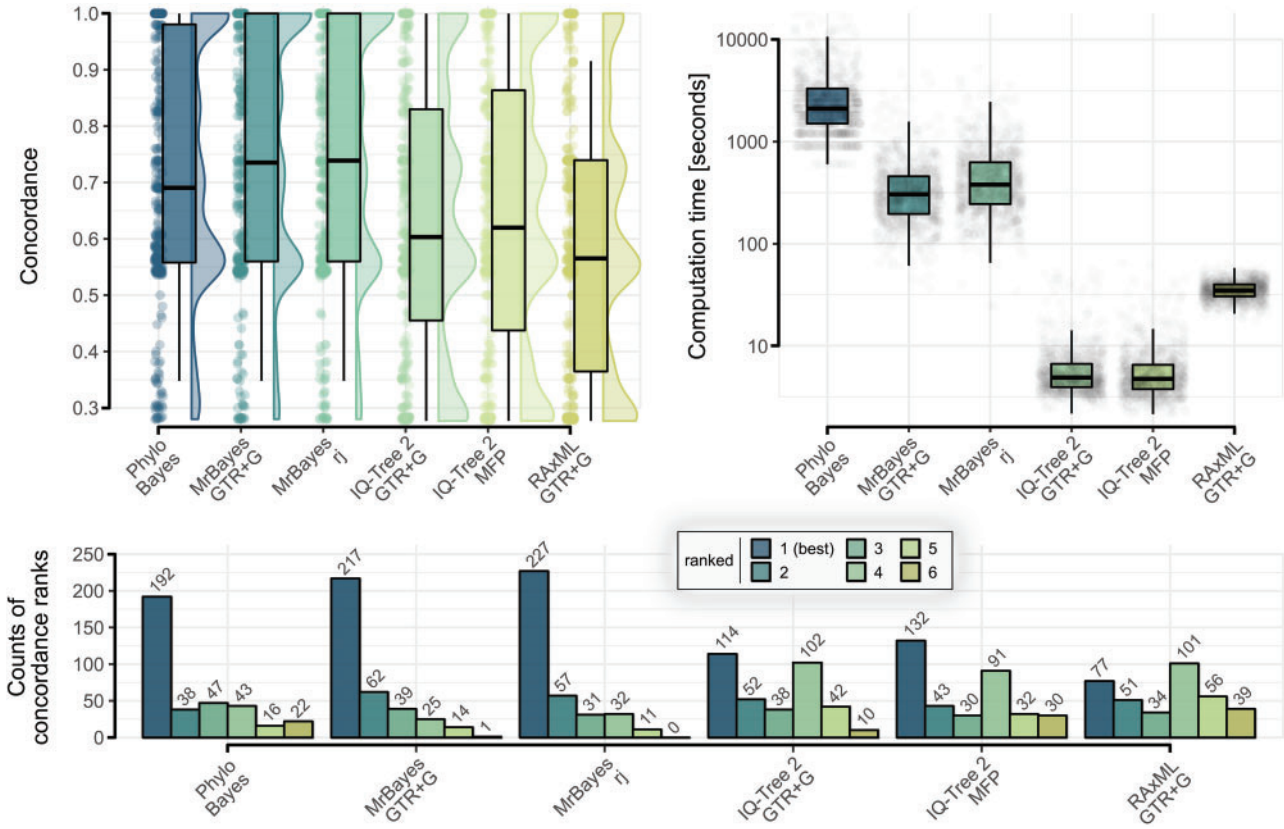

FIGURE 4. Concordance and computational time for the six different gene tree estimation approaches compared in this study. Display of concordance is reduced to only those 358 comparatively examined loci in which concordance values differed between analyses (see Supplementary Material available on Dryad for full version), whereas computational time reflects all 867 loci. Raincloud plots (Allen et al. 2019) of concordance show whiskers extending up to 0.5×IQR. Ranks of concordance can be tied. Computational time is per locus and sums up as follow: 31.4 days (PhyloBayes), 4 and 5.3 days (MrBayes and GTR+G / rj, respectively), 1.6 and 1.5 h (IQ-Tree 2 GTR+G / MFP, respectively), and 8.5 h (RAxML).

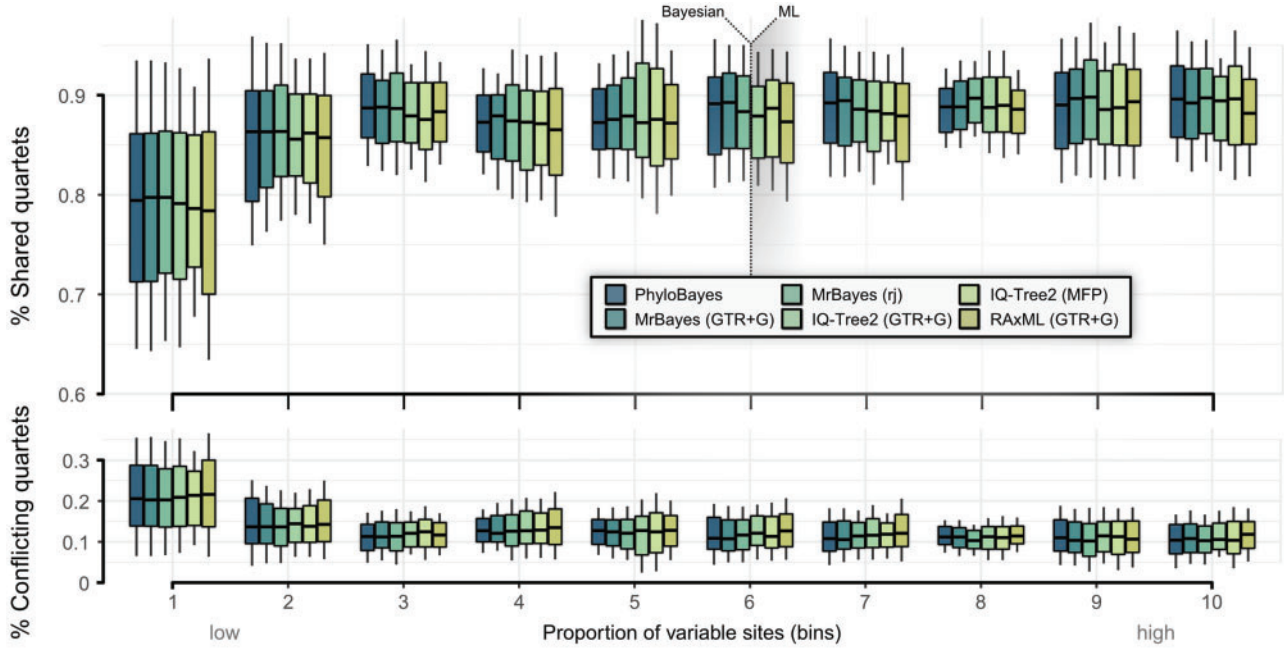

FIGURE 5. Quartet status of gene trees from loci with increasing proportion of variable sites. Each bin contains 85 individual loci, except of bin 10 with 88. Point of reference for quantifying shared and conflicting quartets of individual gene trees is the summary tree of the same method (i.e., each RAxML gene tree gets scored against the ASTRAL summary tree of all RAxML gene trees). Whiskers extend up to 0.5×IQR.

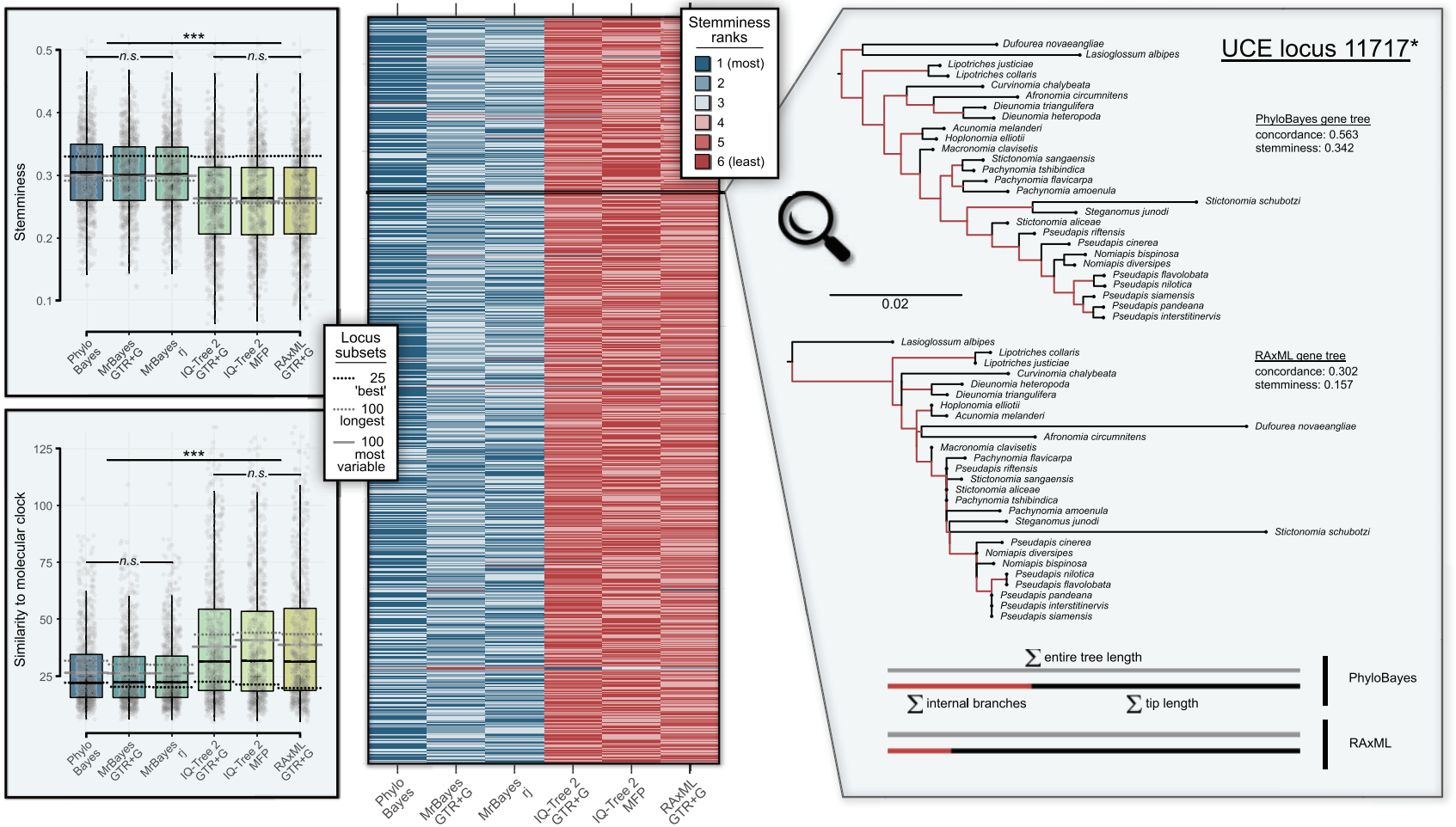

FIGURE 6. Comparison of gene tree branch lengths using the stemminess metric. Displayed is a locus-by-locus comparison of 853 gene trees inferred from the same alignments using six different gene tree estimation approaches. UCE locus 11717 lies within a primarily exonic region of uncharacterized gene locus GB42920 of the *Apis mellifera* reference genome (assembly Amel_4.5, annotation Ensembl-Metazoa release 41). Lines overlaying the boxplots indicate subsets of UCEs that were filtered according to specific alignment or gene tree properties: the 25 or 100 loci with the highest average node support ("best" loci according to average bootstrap values or posterior probabilities), the 100 longest, and the 100 most variable loci. Gene trees with high stemminess and clocklike behavior closer to 0 are preferable. Box plot whiskers extend up to 1.5×IQR. *** indicate $P < 0.0001$ for one-way ANOVA tests to compare differences in means among all six groups: stemminess ($F(5,5112) = 98.29$, $P <$ 2e-16) and similarity to the molecular clock ($F(5,5112) = 81.52$, $P <$ 2e-16).

TABLE 2. Statistics of UCE subsets with different characteristics

Category [loci]	Metric	PhyloBayes (mean ± SD)	MrBayes GTRG (mean ± SD)	MrBayes rj (mean ± SD)	IQ-Tree GTRG (mean ± SD)	IQ-Tree MFP (mean ± SD)	RAxML (mean ± SD)
"Best" 25	Stemminess	0.326 (±0.053)	**0.330 (±0.048)**	0.326 (±0.048)	**0.324 (±0.051)**	**0.329 (±0.046)**	**0.329 (±0.044)**
"Best" 100		**0.329 (±0.057)**	0.328 (±0.053)	**0.328 (±0.053)**	0.306 (±0.060)	0.309 (±0.054)	0.312 (±0.054)
100 longest		0.291 (±0.068)	0.290 (±0.064)	0.290 (±0.063)	0.256 (±0.071)	0.254 (±0.070)	0.256 (±0.071)
100 most variable		0.299 (±0.063)	0.299 (±0.062)	0.299 (±0.062)	0.265 (±0.073)	0.260 (±0.076)	0.264 (±0.074)
100 least gappy		0.311 (±0.062)	0.308 (±0.058)	0.309 (±0.058)	0.272 (±0.068)	0.271 (±0.070)	0.272 (±0.068)
All loci		0.305 (±0.064)	0.302 (±0.061)	0.302 (±0.061)	0.260 (±0.075)	0.260 (±0.074)	0.260 (±0.075)
"Best" 25	Clocklikeness	22.607 (±14.703)	**20.791 (±12.853)**	**20.639 (±12.304)**	**23.264 (±13.024)**	**21.879 (±13.198)**	**20.164 (±14.879)**
"Best" 100		**20.981 (±11.765)**	21.133 (±11.192)	20.805 (±10.754)	24.849 (±15.811)	23.956 (±14.850)	22.872 (±14.263)
100 longest		32.365 (±20.885)	30.995 (±19.272)	30.828 (±19.019)	43.164 (±30.230)	44.017 (±30.840)	43.221 (±30.040)
100 most variable		26.518 (±16.624)	26.087 (±16.896)	26.121 (±16.921)	37.832 (±28.522)	40.501 (±31.111)	38.684 (±29.741)
100 least gappy		25.878 (±14.106)	25.629 (±13.583)	25.617 (±13.600)	35.608 (±22.451)	36.210 (±24.144)	36.030 (±22.692)
All loci		27.579 (±16.461)	27.039 (±15.762)	26.941 (±15.682)	40.418 (±29.632)	40.128 (±28.343)	40.890 (±29.796)
"Best" 25	Concordance	**0.948 (±0.166)**	**0.914 (±0.183)**	**0.933 (±0.166)**	**0.938 (±0.142)**	0.891 (±0.187)	0.829 (±0.232)
"Best" 100		0.884 (±0.197)	0.891 (±0.187)	0.892 (±0.188)	0.885 (±0.188)	**0.904 (±0.178)**	**0.855 (±0.211)**
100 longest		0.788 (±0.246)	0.824 (±0.237)	0.825 (±0.238)	0.794 (±0.253)	0.7865 (±0.265)	0.769 (±0.265)
100 most variable		0.774 (±0.217)	0.779 (±0.223)	0.786 (±0.224)	0.725 (±0.248)	0.721 (±0.251)	0.708 (±0.250)
100 least gappy		0.794 (±0.225)	0.809 (±0.222)	0.807 (±0.229)	0.783 (±0.248)	0.771 (±0.258)	0.764 (±0.260)
All loci		0.784 (±0.236)	0.794 (±0.234)	0.796 (±0.234)	0.748 (±0.256)	0.753 (±0.256)	0.732 (±0.261)

with an average computation time of 52.3 min per locus. Bayesian approaches under site-homogeneous models with MrBayes performed considerably faster with approximately 4 days (MrBayes GTR+G; average 6.72 min per locus) and 5.3 days (MrBayes rj; average 8.86 min per locus) total wall time, using the same alignments and the same computational setup. In contrast, the fastest ML tree searches were achieved with IQ-Tree and MFP model selection (1.5 h total, 6.27 s per locus), which is nearly two magnitudes faster than the MrBayes runs and about 500× faster than PhyloBayes.

DISCUSSION

Reducing GTEE through Method Choice

Studying GTEE is critical for two-step summary approaches under the multispecies coalescent model. Regardless of the general debate of *concatenation* vs. *coalescence* (e.g., Springer and Gatesy 2016 vs. Edwards et al. 2016), there is little controversy over the need to provide summary methods with accurate gene trees and reduce GTEE.

Recent research led to a number of strategies to reduce the effects of GTEE on coalescent analyses. Most prominent approaches involve statistical binning (Bayzid and Warnow 2013; Mirarab et al. 2014a; Bayzid et al. 2015; but see Streicher et al. 2018, Adams and Castoe 2019), identification and removal of outlier loci or taxa (e.g., Wickett et al. 2014; Mai and Mirarab 2018; Leebens-Mack et al. 2019), removal of fragmentary data (Sayyari et al. 2017), testing predefined clades through gene genealogy interrogation (GGI; Arcila et al. 2017), or collapsing very poorly resolved gene tree nodes to polytomies (Zhang et al. 2018). Surprisingly little attention has been paid to explore if GTEE could be reduced by using more accurate gene tree estimation methods. Our results reveal substantial differences in gene tree topology and branch lengths between trees inferred by different methods. These findings show great potential to reduce GTEE by carefully choosing the most appropriate estimation method. They further indicate that significant amounts of gene tree/species tree incongruence in current phylogenomic data sets are not caused by true genealogical discordance, but by incorrect estimation.

We approximate gene tree accuracy by assessing concordance for a set of three deep-branching, uncontroversial clades, which have been recovered as monophyletic by every species tree analyses in the present study and previously published research. This topology-based measure shows higher concordance for the Bayesian methods, indicating that GTEE is slightly lower for Bayesian gene trees over maximum likelihood estimates. Meiklejohn et al. (2016) show a favorable performance of MrBayes in comparison to some ML methods and Mirarab (2019) anecdotally reports the potential that MrBayes may perform slightly better than RAxML. Our results are in line with this and reveal MrBayes and the reversible jump model selection as the preferred strategy to optimize concordance. Using MrBayes and GTR+G allows us to compare gene tree estimates with those of RAxML and IQ-Tree in the same modeling framework, and shows that gene trees of MrBayes are more concordant than the ML estimates. This means that differences in concordance cannot be attributed to different substitution models, but to the underlying statistical framework of the programs.

Species Tree Estimates and Assessment of Conflicting Nodes

Our study is based on empirical data and the "true" species tree is not known. All species tree estimates, however, produce very similar topologies,

supporting previous taxonomic classifications based on morphology. Topological differences of ASTRAL species trees are restricted to three nodes (Fig. 2). Two of these nodes involve the two samples with the worst sequencing success, *Ruginomia rugiventris* and *Stictonomia schubotzi*. Both these taxa have over 75% undetermined positions in the concatenated matrix and are present in less than half of all loci. This translates into significant amounts of missing data types 1 and 2 (following Hosner et al. 2016): sample sequence data is entirely missing in an individual alignment (type 1) and sequence data of the respective samples is present in individual alignments, but is (highly) fragmented (type 2). Type 1 missing data can bias summary analyses under the MSC (Xi et al. 2016) and individual gene trees are differentially weighted based on their taxon representation, effectively giving gene trees with fewer samples less weight (Gatesy et al. 2019). Type 2 is problematic as well, as it increases GTEE and can translate into inaccurate species trees (Hosner et al. 2016; Sayyari et al. 2017). Estimation error of gene trees involving *R. rugiventris* and *S. schubotzi* is therefore high, as they come with great proportions of missing data type 2: several individual alignments do not have a single unique site pattern in the sequences of these two samples. Concatenation approaches are statistically inconsistent under the MSC model (Roch and Warnow 2015), but can produce more accurate species trees when GTEE is very high (e.g., Mirarab and Warnow 2015; Molloy and Warnow 2018; Mirarab 2019). In this light, it is most likely that the topological placements of these samples in the concatenated analyses (Fig. 2), which are topologically corroborated by about half of all ASTRAL summary trees (Fig. 3), reflect the true placement of these samples in the species tree. Interestingly, most summary trees of Bayesian gene tree methods fail to converge on the topology that we deem correct, and ASTRAL summaries of IQ-Tree gene trees resolve these nodes favorably in nearly every analysis.

The third node that creates topological conflict between species tree analyses is a deeper split in the tree, characterized by short internodes and short coalescent times (Fig. 2, Supplementary Figs. S1 and S2 available on Dryad). In contrast to the other two conflicting nodes, sequence representation is very good for all involved tips, implying that missing data is not the cause for inconsistent results. Because of the short branching times, it is probable that incongruence due to ILS renders this node problematic. Most ASTRAL summary trees converge on the same topology at this node, which is topologically equivalent to the concatenated species tree. In contrast, species tree summaries of RAxML gene trees consistently fail to recover this topology (Fig. 3). With RAxML gene trees having the lowest average concordance (Fig. 4), we attribute this to GTEE of the RAxML trees.

The only gene tree estimation method whose summary trees consistently produced the same and the correct topology is IQ-Tree with automated model selection (MFP).

Data Filtering for Summary Analyses

In order to reduce the distorting effects of inaccurately estimated gene trees, recent studies using summary methods implemented a range of different filtering strategies. The two most prominent approaches involve the removal of entire gene trees based on an approximation for its quality, such as informative sites or bootstrap support, or the collapsing of poorly supported nodes of the input gene trees into polytomies.

Removal of Entire Gene Trees.—This strategy aims at removing gene trees that have high GTEE and hence introduce noise into summary analyses. This approach has not been accepted as a reasonable strategy altogether, as it can also remove an honest signal. The important study by Molloy and Warnow (2018) simulated GTEE on a test data set with varying degrees of ILS and found positive effects of gene tree removal on species tree accuracy when GTEE is high and ILS is low. However, for virtually all other modeled conditions, such as increased levels of ILS or various degrees of missing data, species tree accuracy was weakly or negatively affected. Assessing the effects of this filtering approach in empirical data sets is challenging because GTEE cannot be measured (Mirarab 2019) but needs to be approximated through an alignment or tree metric. Some empirical studies found little reason to exclude gene trees with low information content (e.g., Blom et al. 2016), but other studies showed more consistent and/or better-supported results after removing gene trees inferred from alignments with few informative sites (Hosner et al. 2016; Meiklejohn et al. 2016; Longo et al. 2017). The results from our study are in line with this and confirm positive effects on species tree accuracy when removing gene trees estimated from low information alignments. The quartet status assessment shows the greatest topological discordance for the two bins of gene trees (of 10 total) with the least informative loci (Fig. 5), which is consistent for all gene tree estimation methods. Strikingly, summary trees of all tested gene tree methods except RAxML converge on the same topology when the 25% least informative loci are removed, whereas they produce five different topologies when all loci are included (Fig. 3). Quartet status in the remaining bins (bins 3–10) are similarly high, showing that additional removal of gene trees does not bring the desired disproportional removal of noisy loci, but an unnecessary reduction of the total sample size. In line with his, summary trees that include only the 50% most informative loci are less accurate than the larger data set (Fig. 3), which leads us to conclude that aggressive filtering leads to adverse effects on species tree accuracy in our study.

Interestingly, all empirical studies which showed beneficial effects of removing noisy gene trees generated their trees using ultraconserved elements (i.e., Hosner et al. 2016; Meiklejohn et al. 2016; Longo et al. 2017). These markers are characterized by relatively short, slowly evolving DNA segments, with the most informative

sites located in the adjacent flanking regions (Faircloth et al. 2012; McCormack et al. 2012). This conservative nature gives UCE gene trees a somewhat particular status, as they have very few informative sites with great potential for GTEE. All above-mentioned studies, however, reduced their pool of input UCE gene trees to a small fraction and favored results from only the 25% most informative loci or less (Hosner et al. 2016; Meiklejohn et al. 2016; Longo et al. 2017). For our study, this would imply the removal of a large proportion of informative gene trees, which seems excessive. An explanation for this lies in the different degrees of conservation of vertebrate and Hymenopteran UCEs: elements targeted by the amniote probe set show substantially lower frequencies of substitution compared to UCEs of bees (compare Faircloth et al. 2012, Fig. 3 vs. Bossert et al. 2017, Fig. 3). For best practices on UCE data sets, our results show that removing uninformative loci is beneficial for species tree summary analyses, but the threshold for including or excluding gene trees needs to be assessed based on the individual data. Our results suggest that the best strategy for limiting the impact of uninformative loci is to examine the quartet status of each locus along a gradient and exclude only those loci which fall at the extreme end of the distribution (Fig. 5).

Collapsing Weakly Supported Nodes of Gene Trees.—Currently, available summary methods treat individual gene trees as observations and assume that they are correctly inferred. ASTRAL, for example, is relatively robust towards GTEE (Roch and Warnow 2015; Sayyari et al. 2017) but does not take node support of input gene trees into account. In the end, it does not matter if a clade is highly supported based on bootstrap values, or if it is arbitrarily solved with negligible confidence. It therefore seems intuitive to collapse weakly supported nodes into polytomies, in order to avoid the introduction of noise into the summary step. A recent simulation study examined this and found positive effects on species tree accuracy when weakly supported nodes are collapsed (Zhou et al. 2017). Very stringent collapsing, that is, of nodes that are supported by bootstrap values beyond 30 (or even 20, depending on the specific conditions), decreases species tree accuracy and is discouraged (Zhou et al. 2017). Our summary trees confirm these simulation results, as the summary trees become more similar to the best species tree topology when nodes below 30% (i.e., bootstrap values of $\leq$30/$\leq$0.3 PP) are collapsed (Fig. 3). Species trees are most in conflict with each other and the best topology when input gene trees are very aggressively trimmed ($\leq$50%). This overly strict trimming is the only condition that causes summary trees of IQ-Tree to deviate from the best topology.

While our results show moderately positive effects of collapsing weakly supported nodes of input gene trees, our quartet score assessment also reveals that most node retractions occur in loci with very low information (Supplementary Figs. S4 and S5 available on Dryad). The two bins of least informative loci show both the greatest decrease in shared and conflicting quartets, as well as the directly related increase in unresolved quartets. In this light, it seems imprecise to indiscriminately collapse nodes of all loci, if only those with low information content are obviously problematic. A future inroad into optimizing summary analyses could be the development of a flexible procedure to collapse weakly supported nodes only of those gene trees that have the greatest potential for GTEE (i.e., few informative sites) or those with unusually large amounts of conflicting quartets. For our specific data set, however, it ultimately proved best to discard gene trees of the least informative loci altogether (Fig. 3). An alternative future direction is the further development of site-based coalescent methods such as SVDQuartets (Chifman and Kubatko 2014), in order to circumvent the gene tree estimation step altogether.

Gene Tree Branch Lengths and Stemminess

As current summary methods consider only the topology of input gene trees, differences in branch lengths do not impact the estimation of species trees under the MSC. Nonetheless, we found significant differences in branch length and stemminess between tree estimation methods. The stemminess metric has been linked to saturation and a reduction thereof is reflected by higher, favorable stemminess values (Longhorn et al. 2010). Saturation in turn has long been recognized to facilitate long-branch attraction artifacts (e.g., Philippe and Laurent 1998; Brinkmann et al. 2005). The higher stemminess of Bayesian gene trees presented herein indicates better accounting for sequence saturation and reduced risk for long-branch attraction. While this does not necessarily improve the topology of every individual gene tree, it may reduce the number of individual trees that are affected by such reconstruction artifacts. The exemplified RAxML gene tree of UCE locus 11717 has *Dufourea* deeply nested within Nomiinae, likely deriving from a reconstruction artifact involving long branches, whereas the PhyloBayes gene tree has more reasonable topology and branch lengths (Fig. 6). Interestingly, we found Bayesian gene trees to be consistently more stemmy and concordant over their ML counterparts. However, the method which produced the stemmiest gene trees is PhyloBayes, which is the least concordant of the three Bayesian approaches. This shows that stemminess is not necessarily linked to concordance.

The differences in branch lengths between gene trees of different methods cannot be simply explained by different substitution models. Stemminess is higher for gene trees of MrBayes with GTR+G over its RAxML and IQ-Tree counterparts under the same model. In fact, our data show that the underlying statistical framework of tree estimation (Bayesian vs. ML) is the main driver of branch length differences. Among the different Bayesian

analyses, however, gene trees inferred using the site-heterogeneous CAT-GTR model are stemmier than those estimated under site-homogenous models.

Implications for Gene Tree Summary Methods

We summarized six sets of different gene trees, estimated through four different programs, and found five different species tree topologies when using the unaltered gene trees as input (Fig. 3). This illustrates the great sensitivity of coalescent-based summary analyses towards the quality of the input gene trees and underlines the need to optimize gene tree estimation for minimizing GTEE. Our assessment of concordance and resulting species tree topologies renders RAxML as the least preferred program to estimate gene trees, but practical recommendations are less obvious for the remaining methods. To this end, we deem three factors as important when considering the best approach for inferring gene trees: concordance, fragmentation of individual alignments (missing data type 2 following Hosner et al. 2016), and the size of the data set to be analyzed. First, concordance for solving the three predefined clades is greatest for gene trees inferred through MrBayes. This shows that GTEE at deep nodes is lower for these trees than for those of other estimation methods. The Bayesian gene trees, however, are more severely affected by type 2 missing data. Our results show IQ-Tree to be more robust towards this kind of missing data, and render summary trees of IQ-Tree and automated model-selection (MFP) as the only approach that consistently produces the best topology, despite lower concordance for deeper splits in the tree. This indicates IQ-Tree as the best choice for estimating gene trees using fragmentary sequence data (missing data type 2), a finding which should ideally be reassessed and confirmed by simulation experiments. A previous study found IQ-Tree to be preferable over other tested ML methods in empirical data sets, but this study did not compare gene trees inferred through Bayesian methods (Zhou et al. 2017). Lastly, estimation times between methods differ greatly, and Bayesian analyses could become unfeasible for larger data sets. The computation of MrBayes (rj) gene trees was over 5× faster than those of PhyloBayes, but it was still > 80× slower than the very efficient IQ-Tree (MFP), which was the fastest of all tested methods. For our moderately sized data set of 32 taxa, this translated into a difference of 5 days, but larger matrices will inevitably face greater computational challenges when choosing Bayesian methods.

For practical consideration, we suggest inferring gene trees with MrBayes if individual alignment fragmentation is moderate to low, and when the computational demands allow its use in a reasonable timeframe. With a greater degree of type 2 missing data (sensu Hosner et al. 2016) and when Bayesian inference becomes impracticable, we strongly recommend the use of IQ-Tree. Further, both these methods slightly benefited from an automated selection of substitution models, which in the case of IQ-Tree even reduced the overall computational time.

The "Best" Trees?

The wealth of sequence data produced in phylogenomic data sets brings particular challenges for computationally demanding analyses such as divergence time estimates. Downsizing the amount of sequence data for downstream analyses by "Gene Shopping" (Smith et al. 2018) is common practice in the field of molecular phylogenetics. UCE-based studies often apply a filtering approach in which loci are ranked by their average bootstrap support and only a subset of the 100, 50, or 25 "best" performing loci are used for dating analyses (e.g., Branstetter et al. 2017a; Ješovnik et al. 2017; Blaimer et al. 2018). However, it has never been evaluated if UCEs with the highest support have favorable properties that deem them particularly suitable for subsequent analyses, such as molecular dating. Specifically, it has never been examined how other locus selection criteria, such as filtering for particularly long or variable UCEs, compare.

The comparative results presented in this study show that loci with high average support perform, on average, better than any other subset in every assessed gene tree quality approximation (Fig. 6, Table 2). For all subsets, the "best" gene trees are more concordant (Table 2) and have favorable branch lengths (Fig. 6). Specifically, they perform better than the subsets of the longest, the most variable, and least gappy loci (lowest % of missing data). Strikingly, they behave more clock-like than the average gene tree, underlining their potential for divergence time estimates.

Phylogeny and Classification of Pseudapis *s.l.*

Higher-level bee classification has generally been guided or corroborated by molecular phylogenies. However, many species groups below the family level, such as the subfamily Nomiinae, remain unstudied from a phylogenetic perspective (Danforth et al. 2012). Herein, we focus on the genus *Pseudapis* and establish the first phylogeny-informed classification for this distinct group of Nomiinae.

The *Pseudapis* group (Fig. 2, all colored clades) comprises 81 species from the Old World. Previous classifications conflict in their use of genera and subgenera, and *Pseudapis* s. l. has been alternatively split into varying conformations, comprising up to six (Pauly 1990; Baker 2002) or just two genera (Michener 2007; Ascher and Pickering 2020) (summarized in Fig. 7). Michener's (2007) popular bee classification synonymizes several supraspecific names with *Pseudapis* (s. str.) to prevent potential paraphyletic taxa (*Nomiapis*, *Stictonomia*, *Ruginomia*), but *Pachynomia* remained a subgenus of *Pseudapis*. The unusual *Steganomus* is recognized as a separate genus and not part of *Pseudapis* (Michener 2007). Similarly, the classification of the Discover Life database (Ascher and Pickering 2020) regards *Nomiapis* and *Pseudapis* s. str. as subgenera, and synonymizes *Stictonomia* and *Ruginomia* with *Pseudapis* s. str. Both these classifications contrast Pauly (1990, 2009),

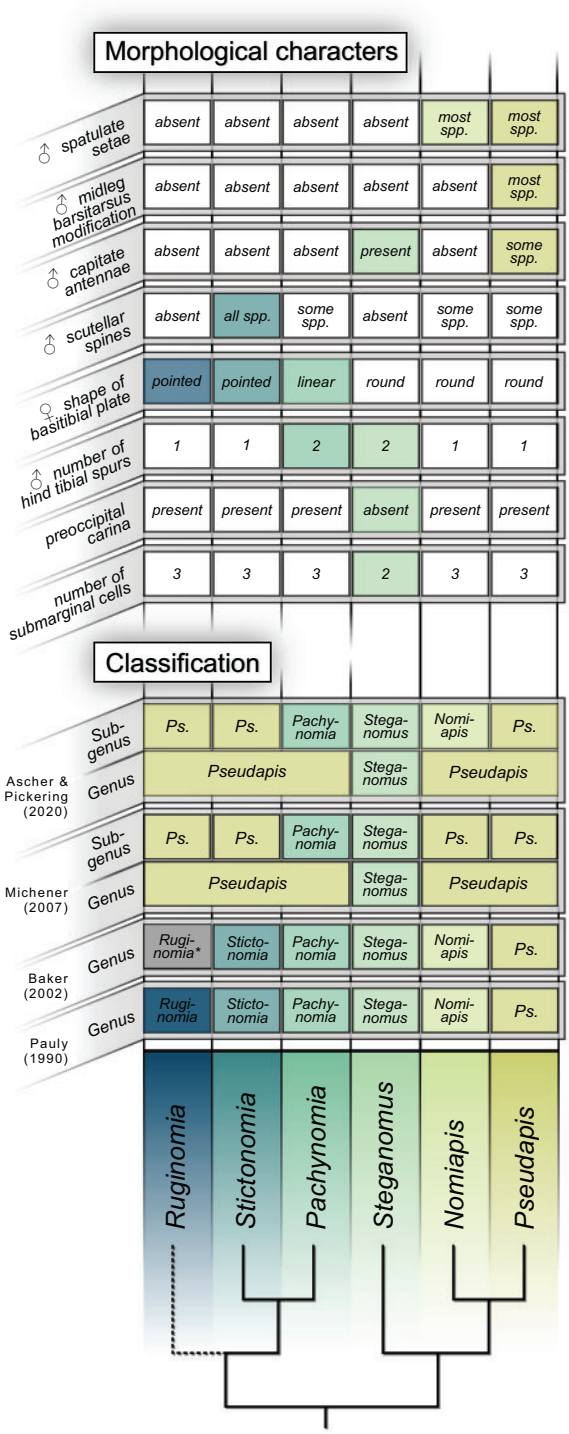

FIGURE 7. Reciprocal illumination of taxonomic classifications and selected morphological characters of *Pseudapis* and closely related groups. The only rank-based classification that yields monophyletic groups is the one of Pauly (1990). Baker's (2002) writing is inconclusive on the placement of *Ruginomia*, but he indicates that generic status may not be adequate. Morphological characters show a high degree of homoplasy, yet all groups can be distinguished by unique combinations of nonunique characters.

who recognizes all above-mentioned names as genera. Baker (2002) revised *Pseudapis* and *Nomiapis* from the Palearctic and the Oriental region and inferred them as separate monophyletic groups in morphology-based cladistic analyses. He follows Pauly's (1990) classification but concludes that the generic status of *Ruginomia* may have been overstated.

Our phylogenomic results show that the classification of Pauly (1990) is the most appropriate for *Pseudapis* s. l. in that it is the only one that results in monophyletic groups (Fig. 7). Comprising six genera, this classification brings the highest degree of taxonomic splitting, yet it is preferred to ensure a rank-based taxonomy of monophyletic groups. Only recognizing *Steganomus* and *Pseudapis* as genera (Michener 2007; Ascher and Pickering 2020) renders the latter paraphyletic and should be avoided. Alternatively, *Steganomus* could be regarded as a subgenus within a large genus *Pseudapis* that includes all lineages. However, this would also require the acceptance of all six lineages as subgenera of *Pseudapis*. As is, the classification of Pauly (1990) does not require taxonomic changes.

Morphological characters that have both been used to argue for and against synonymizations show a high degree of homoplasy. The mapped characters in Figure 7 are not exhaustive, yet they show that readily recognized features are often present in multiple lineages. The exception is *Steganomus* and the unique absence of a third submarginal wing cell. Reciprocally illuminating the phylogeny of *Pseudapis* s. l. proves Pauly's (1990) approach of designating genera based on unique combinations of nonunique characters as effective, and all groups should be identifiable with the respective keys (Pauly 1990, Pauly 2009; Baker 2002; Bossert and Pauly 2019).

Concatenation and most coalescent-based analyses favor a sistergroup relationship of *Ruginomia* to *Stictonomia* + *Pachynomia* (Figs. 2 and 3). Some summary analyses of Bayesian gene trees, however, recover a clade comprising *Ruginomia* + (*Steganomus* + (*Nomiapis* + *Pseudapis*)), but with weak support (Supplementary Fig. S1 available on Dryad). If this would represent the true tree, *Ruginomia* would still require generic status, and no species tree analysis recovers *Ruginomia* as part of *Pseudapis*, or as its sistergroup. Therefore, we need to regard *Ruginomia* as separate from *Pseudapis* s. str. in any case.

CONCLUSION

Our case study shows that gene trees estimated with different estimation methods substantially differ in topology and branch lengths, thereby decisively impacting the accuracy of summary methods under the multispecies coalescent model. In part, we can attribute these differences to GTEE, which can be reduced by choosing the most appropriate gene tree estimation method. Even with the increasing availability

of genomic sequence data, an important objective of future research should be the optimization of tree inferences of individual loci, in order to summarize the thoroughly inferred genealogical history of thousands of loci into single species trees.

Supplementary Material

This repository contains the Trinity-assemblies of the de-novo sequenced UCEs, extracted UCE sequences from the included genomes, and the concatenated 80% completeness matrix. We further provide all species trees and all 853 gene trees inferred through the different gene tree estimation methods. Lastly, we provide the R code that was used to infer stemminess. Trimmed Illumina reads associated with this study were deposited in the Sequence Read Archive (SRA accession PRJNA494583). The EZ-PB script developed in the course of this study is available on GitHub (https://github.com/Bluefire2/EZPB).

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.z08kprrb6.

Funding

This work was supported by a U.S. National Science Foundation grant (DEB-1555905 to B.N.D., S.G.B., J.P. Pitts, and R. Ross) and by Peter Buck fellowships at the Smithsonian Institution to S.B. and E.A.M.

Acknowledgments

We thank Jason Dombroskie (Cornell University Insect Collection) and Matthew Buffington (USDA-ARS) for access to the imaging systems. We further thank Martin Hauser (California Department of Food and Agriculture) for contributing specimens to this study, and Doug Yanega for providing loans from the Entomology Research Museum of the University of California Riverside. The laboratory work for this study was conducted in the L.A.B. facilities of the National Museum of Natural History, Smithsonian Institution.

References

Adams R.H., Castoe T.A. 2019. Statistical binning leads to profound model violation due to gene tree error incurred by trying to avoid gene tree error. Mol. Phylogenet. Evol. 134:164–171.

Allen M., Poggiali D., Whitaker K., Marshall T., Kievit R. 2019. Raincloud plots: a multi-platform tool for robust data visualization. Wellcome Open Res. 4:63.

Andrews S. 2019. FastQC: A quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (September 2018).

Arcila D., Ortí G., Vari R., Armbruster J.W., Stiassny M.L.J., Ko K.D., Sabaj M.H., Lundberg J., Revell L.J., Betancur-R R. 2017. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. Nat. Ecol. Evol. 1:0020.

Ascher J.S., Pickering J. 2020. Discover life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila). Available from: http://www.discoverlife.org/mp/20q?guide=Apoidea_species.

Baker D.B. 2002. On Palaearctic and oriental species of the genera *Pseudapis* W.F. Kirby, 1900, and *Nomiapis* Cockerell, 1919. Beitr. Entomol. 52:1–83.

Bayzid M.S., Mirarab S., Boussau B., Warnow T. 2015. Weighted statistical binning: enabling statistically consistent genome-scale phylogenetic analyses. PLoS One 10:e0129183.

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

Blaimer B.B., LaPolla J.S., Branstetter M.G., Lloyd M.W., Brady S.G. 2016a. Phylogenomics, biogeography and diversification of obligate mealybug-tending ants in the genus *Acropyga*. Mol. Phylogenet. Evol. 102:20–29.

Blaimer B.B., Lloyd M.W., Guillory W.X., Brady S.G. 2016b. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PLoS One 11:e0161531.

Blaimer B.B., Ward P.S., Schultz T.R., Fisher B.L., Brady S.G. 2018. Paleotropical diversification dominates the evolution of the hyperdiverse ant tribe Crematogastrini (Hymenoptera: Formicidae). Insect Syst. Div. 2:1–14.

Blom M.P.K., Bragg J.G., Potter S., Moritz C. 2016. Accounting for uncertainty in gene tree estimation: summary-coalescent species tree inference in a challenging radiation of Australian Lizards. Syst. Biol. 66:352–366.

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

Borowiec M.L., Lee E.K., Chiu J.C., Plachetzki D.C. 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. BMC Genom. 16:1–15.

Bossert S., Danforth B.N. 2018. On the universality of target enrichment baits for phylogenomic research. Methods Ecol. Evol. 9:1453–1460.

Bossert S., Murray E.A., Blaimer B.B., Danforth B.N. 2017. The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. Mol. Phylogenet. Evol. 111:149–157.

Bossert S., Pauly A. 2019. Two new species of *Pseudapis* Kirby, 1900 (Hymenoptera: Halictidae: Nomiinae) from Africa. Zootaxa 4608:517–530.

Branstetter M.G., Danforth B.N., Pitts J.P., Faircloth B.C., Ward P.S., Buffington M.L., Gates M.W., Kula R.R., Brady S.G. 2017a. Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. Curr. Biol. 27:1019–1025.

Branstetter M.G., Longino J.T., Ward P.S., Faircloth B.C. 2017b. Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. Methods Ecol. Evol. 8:768–776.

Bravo G.A., Antonelli A., Bacon C.D., Bartoszek K., Blom M.P.K., Huynh S., Jones G., Knowles L.L., Lamichhaney S., Marcussen T., Morlon H., Nakhleh L.K., Oxelman B., Pfeil B., Schliep A., Wahlberg N., Werneck F.P., Wiedenhoeft J., Willows-Munro S., Edwards S.V. 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ 7:e6399.

Brinkmann H., van der Giezen M., Zhou Y., de Raucourt G.P., Philippe H. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst. Biol. 54:743–757.

Cardinal S., Buchmann S.L., Russell A.L. 2018. The evolution of floral sonication, a pollen foraging behavior used by bees (Anthophila). Evolution 72:590–600.

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

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

Danforth B., Brady S., Sipes S., Pearson A. 2004. Single copy nuclear genes recover Cretaceous age divergences in bees. Syst. Biol. 53:309 - 326.

Danforth B.N., Cardinal S., Praz C., Almeida E.A.B., Michez D. 2012. The impact of molecular data on our understanding of bee phylogeny and evolution. Annu. Rev. Entomol. 58:57–78.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.
Deng W., Maust B.S., Nickle D.C., Learn G.H., Liu Y., Heath L., Pond S.L.K., Mullins J.I. 2010. DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. BioTechniques 48:405–408.
Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.
Edwards S.V., Xi Z., Janke A., Faircloth B.C., McCormack J.E., Glenn T.C., Zhong B., Wu S., Lemmon E.M., Lemmon A.R., Leaché A.D., Liu L., Davis C.C. 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94:447–462.
Faircloth B.C. 2013. illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. doi: 10.6079/J9ILL.
Faircloth B.C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786–788.
Faircloth B.C., Branstetter M.G., White N.D., Brady S.G. 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol. Ecol. Resour. 15:489–501.
Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717–726.
Fiala K.L., Sokal R.R. 1985. Factors determining the accuracy of cladogram estimation: evaluation using computer simulation. Evolution 39:609–622.
Gatesy J., Springer M.S. 2013. Concatenation versus coalescence versus "concatalescence". Proc. Natl. Acad. Sci. USA 110:E1179.
Gatesy J., Springer M.S. 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80:231–266.
Gatesy J., Sloan D.B., Warren J.M., Baker R.H., Simmons M.P, Springer M.S. 2019. Partitioned coalescence support reveals biases in species-tree methods and detects gene trees that determine phylogenomic conflicts. Mol. Phylogenet. Evol. 139:106539.
Glenn T.C., Nilsen R.A., Kieran T.J., Sanders J.G., Bayona-Vásquez N.J., Finger J.W., Pierson T.W., Bentley K.E., Hoffberg S.L., Louha S., Garcia-De Leon F.J., del Rio Portilla M.A., Reed K.D., Anderson J.L., Meece J.K., Aggrey S.E., Rekaya R., Alabady M., Belanger M., Winker K., Faircloth B.C. 2019. Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). PeerJ 7:e7755.
Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotech. 29:644–652.
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.
Harris R.S. 2007. Improved pairwise alignment of genomic DNA [PhD thesis]. Computer Science and Engineering Department, Pennsylvania State University.
Hedtke S.M., Patiny S., Danforth B.N. 2013. The bee tree of life: a supermatrix approach to apoid phylogeny and biogeography. BMC Evol. Biol. 13:1–13.
Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522.
Hosner P.A., Faircloth B.C., Glenn T.C., Braun E.L., Kimball R.T. 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the Landfowl (Aves: Galliformes). Mol. Biol. Evol. 33:1110–1125.
Huelsenbeck J.P., Larget B., Alfaro M.E. 2004. Bayesian phylogenetic model selection using reversible jump Markov Chain Monte Carlo. Mol. Biol. Evol. 21:1123–1133.
Ješovnik A., Sosa-Calvo J., Lloyd M.W., Branstetter M.G., Fernández F., Schultz T.R. 2017. Phylogenomic species delimitation and host-symbiont coevolution in the fungus-farming ant genus *Sericomyrmex* Mayr (Hymenoptera: Formicidae): ultraconserved elements (UCEs) resolve a recent radiation. Syst. Entomol. 42:523–542.
Jombart T., Kendall M., Almagro-Garcia J., Colijn C. 2017. treespace: statistical exploration of landscapes of phylogenetic trees 17:1385–1392.
Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587.
Kapheim K.M., Pan H., Li C., Blatti C., Harpur B.A., Ioannidis P., Jones B.M., Kent C.F., Ruzzante L., Sloofman L., Stolle E., Waterhouse R.M., Zayed A., Zhang G., Wcislo W.T. 2019. Draft genome assembly and population genetics of an agricultural pollinator, the solitary alkali bee (Halictidae: *Nomia melanderi*). G3 (Bethesda) 9:625–634.
Kapheim K.M., Pan H., Li C., Salzberg S.L., Puiu D., Magoc T., Robertson H.M., Hudson M.E., Venkat A., Fischman B.J., Hernandez A., Yandell M., Ence D., Holt C., Yocum G.D., Kemp W.P., Bosch J., Waterhouse R.M., Zdobnov E.M., Stolle E., Kraus F.B., Helbing S., Moritz R.F.A., Glastad K.M., Hunt B.G., Goodisman M.A.D., Hauser F., Grimmelikhuijzen C.J.P., Pinheiro D.G., Nunes F.M.F., Soares M.P.M., Tanaka É.D., Simões Z.L.P., Hartfelder K., Evans J.D., Barribeau S.M., Johnson R.M., Massey J.H., Southey B.R., Hasselmann M., Hamacher D., Biewer M., Kent C.F., Zayed A., Blatti C., Sinha S., Johnston J.S., Hanrahan S.J., Kocher S.D., Wang J., Robinson G.E., Zhang G. 2015. Genomic signatures of evolutionary transitions from solitary to group living. Science 348:1139–1143.
Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.
Kendall M., Colijn C. 2016. Mapping phylogenetic trees to reveal distinct patterns of evolution. Mol. Biol. Evol. 33:2735–2743.
Kendall M., Eldholm V., Colijn C. 2018. Comparing phylogenetic trees according to tip label categories. BioRxiv 251710.
Kocher S., Li C., Yang W., Tan H., Yi S., Yang X., Hoekstra H., Zhang G., Pierce N., Yu D. 2013. The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. Genome Biol. 14:R142.
Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.
Kück P., Greve C., Misof B., Gimnich F. 2012. Automated masking of AFLP markers improves reliability of phylogenetic analyses. PLoS One 7:e49119.
Lartillot N. 2013. Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. Mol. Biol. Evol. 30:489–502.
Lartillot N., Brinkmann H., Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol. 7:1–14.
Lartillot N., Lepage T., Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25:2286–2288.
Lartillot N., Philippe H. 2004. A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21:1095–1109.
Lartillot N., Rodrigue N., Stubbs D., Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. 62:611–615.
Leebens-Mack J.H., Barker M.S., Carpenter E.J., Deyholos M.K., Gitzendanner M.A., Graham S.W., Grosse I., Li Z., Melkonian M., Mirarab S., Porsch M., Quint M., Rensing S.A., Soltis D.E., Soltis P.S., Stevenson D.W., Ullrich K.K., Wickett N.J., DeGironimo L., Edger P.P., Jordon-Thaden I.E., Joya S., Liu T., Melkonian B., Miles N.W., Pokorny L., Quigley C., Thomas P., Villarreal J.C., Augustin M.M., Barrett M.D., Baucom R.S., Beerling D.J., Benstein R.M., Biffin E., Brockington S.F., Burge D.O., Burris J.N., Burris K.P., Burtet-Sarramegna V., Caicedo A.L., Cannon S.B., Çebi Z., Chang Y., Chater C., Cheeseman J.M., Chen T., Clarke N.D., Clayton H.,

Covshoff S., Crandall-Stotler B.J., Cross H., dePamphilis C.W., Der J.P., Determann R., Dickson R.C., Di Stilio V.S., Ellis S., Fast E., Feja N., Field K.J., Filatov D.A., Finnegan P.M., Floyd S.K., Fogliani B., García N., Gâteblé G., Godden G.T., Goh F., Greiner S., Harkess A., Heaney J.M., Helliwell K.E., Heyduk K., Hibberd J.M., Hodel R.G.J., Hollingsworth P.M., Johnson M.T.J., Jost R., Joyce B., Kapralov M.V., Kazamia E., Kellogg E.A., Koch M.A., Von Konrat M., Könyves K., Kutchan T.M., Lam V., Larsson A., Leitch A.R., Lentz R., Li F.-W., Lowe A.J., Ludwig M., Manos P.S., Mavrodiev E., McCormick M.K., McKain M., McLellan T., McNeal J.R., Miller R.E., Nelson M.N., Peng Y., Ralph P., Real D., Riggins C.W., Ruhsam M., Sage R.F., Sakai A.K., Scascitella M., Schilling E.E., Schlösser E.-M., Sederoff H., Servick S., Sessa E.B., Shaw A.J., Shaw S.W., Sigel E.M., Skema C., Smith A.G., Smithson A., Stewart C.N., Stinchcombe J.R., Szövényi P., Tate J.A., Tiebel H., Trapnell D., Villegente M., Wang C.-N., Weller S.G., Wenzel M., Weststrand S., Westwood J.H., Whigham D.F., Wu S., Wulff A.S., Yang Y., Zhu D., Zhuang C., Zuidof J., Chase M.W., Pires J.C., Rothfels C.J., Yu J., Chen C., Chen L., Cheng S., Li J., Li R., Li X., Lu H., Ou Y., Sun X., Tan X., Tang J., Tian Z., Wang F., Wang J., Wei X., Xu X., Yan Z., Yang F., Zhong X., Zhou F., Zhu Y., Zhang Y., Ayyampalayam S., Barkman T.J., Nguyen N.-p., Matasci N., Nelson D.R., Sayyari E., Wafula E.K., Walls R.L., Warnow T., An H., Arrigo N., Baniaga A.E., Galuska S., Jorgensen S.A., Kidder T.I., Kong H., Lu-Irving P., Marx H.E., Qi X., Reardon C.R., Sutherland B.L., Tiley G.P., Welles S.R., Yu R., Zhan S., Gramzow L., Theißen G., Wong G.K.-S., One Thousand Plant Transcriptomes I. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574:679–685.

Longhorn S.J., Pohl H.W., Vogler A.P. 2010. Ribosomal protein genes of holometabolan insects reject the Halteria, instead revealing a close affinity of Strepsiptera with Coleoptera. Mol. Phylogenet. Evol. 55:846–859.

Longo S.J., Faircloth B.C., Meyer A., Westneat M.W., Alfaro M.E., Wainwright P.C. 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. Mol. Phylogenet. Evol. 113:33–48.

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

Mai U., Mirarab S. 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genom. 19:272.

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.

Meiklejohn K.A., Faircloth B.C., Glenn T.C., Kimball R.T., Braun E.L. 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. Syst. Biol. 65:612–627.

Michener C.D. 2007. The bees of the world. Baltimore: The Johns Hopkins University Press.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol.

Mirarab S. 2019. Species tree estimation using ASTRAL: practical considerations. arXiv preprint arXiv:1904.03826.

Mirarab S., Bayzid M., Boussau B., Warnow T. 2014a. Statistical binning improves species tree estimation in the presence of gene tree incongruence. Science 346:1250463.

Mirarab S., Bayzid M.S., Warnow T. 2014b. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65:366–380.

Mirarab S., Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44-i52.

Molloy E.K., Warnow T. 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. Syst. Biol. 67:285–303.

Murtagh F., Legendre P. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J. Class. 31:274–295.

Patel S., Kimball R.T., Braun E.L. 2013. Error in phylogenetic estimation for bushes in the tree of life. J. Phylogen. Evol. Biol. 1:1–10.

Pauly A. 1990. Classification des Nomiinae Africains (Hymenoptera Apoidea Halictidae). Musée Royal de l'Afrique Centrale Tervuren, Belgique 261:1–206.

Pauly A. 2009. Classification des Nomiinae de la Région Orientale, de Nouvelle-Guinée et des îles de l'Océan Pacifique (Hymenoptera: Apoidea: Halictidae). Bull. Inst. Roy. Sci. Nat. Belgique 79:151–229.

Philippe H., Laurent J. 1998. How good are deep phylogenetic trees? Curr. Opin. Genet. Dev. 8:616–623.

Portik D.M., Wiens J.J. 2020. Do alignment and trimming methods matter for phylogenomic (UCE) Analyses? Syst. Biol. 70:440–462. doi: 10.1093/sysbio/syaa064.

Price M.N., Dehal P.S., Arkin A.P. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

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

Roch S., Steel M. 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100:56–62.

Roch S., Warnow T. 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Syst. Biol. 64:663–676.

Rohlf F.J., Chang W., Sokal R., Kim J. 1990. Accuracy of estimated phylogenies: effects of tree topology and evolutionary model. Evolution 44:1671–1684.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Saghai-Maroof M.A., Soliman K.M., Jorgensen R.A., Allard R.W. 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA 81:8014–8018.

Sayyari E., Whitfield J.B., Mirarab S. 2017. Fragmentary gene sequences negatively impact gene tree and species tree reconstruction. Mol. Biol. Evol. 34:3279–3291.

Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C., Brumfield R.T. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63:83–95.

Smith M.R. 2019. Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets. Biol. Lett. 15:20180632.

Smith M.R. 2020. Quartet: comparison of phylogenetic trees using quartet and bipartition measures (Version v1.1.0). Zenodo Available from: http://doi.org/10.5281/zenodo.3630138.

Smith S.A., Brown J.W., Walker J.F. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PLoS One 13:e0197433.

Springer M.S., Gatesy J. 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94:1–33.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Streicher J.W., Miller E.C., Guerrero P.C., Correa C., Ortiz J.C., Crawford A.J., Pie M.R., Wiens J.J. 2018. Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2214 loci. Mol. Phylogenet. Evol. 119:128–143.

Tagliacollo V.A., Lanfear R. 2018. Estimating improved partitioning schemes for ultraconserved elements. Mol. Biol. Evol. 35:1798–1811.

Talavera G., Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Tong K.J., Duchêne D.A., Duchêne S., Geoghegan J.L., Ho S.Y.W. 2018. A comparison of methods for estimating substitution rates from ancient DNA sequence data. BMC Evol. Biol. 18:70.

Van Dam M.H., Henderson J.B., Esposito L., Trautwein M. 2020. Genomic characterization and curation of UCEs improves species tree reconstruction. Syst. Biol. 70:307–321. doi: 10.1093/sysbio/syaa063.

Van Dam M.H., Lam A.W., Sagata K., Gewa B., Laufa R., Balke M., Faircloth B.C., Riedel A. 2017. Ultraconserved elements (UCEs)

resolve the phylogeny of Australasian smurf-weevils. PLoS One 12:e0188044.

Wickett N.J., Mirarab S., Nguyen N., Warnow T., Carpenter E., Matasci N., Ayyampalayam S., Barker M.S., Burleigh J.G., Gitzendanner M.A. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA 111:E4859-E4868.

Xi Z., Liu L., Davis C.C. 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol. Phylogenet. Evol. 92:63–71.

Xi Z., Liu L., Davis C.C. 2016. The impact of missing data on species tree estimation. Mol. Biol. Evol. 33:838–860.

Xu B., Yang Z. 2016. Challenges in species tree estimation under the multispecies coalescent model. Genetics 204:1353–1368.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.

Zhou X., Shen, X.-X., Hittinger C.T., Rokas A. Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. Mol. Biol. Evol. 35:486–503.