RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Statistical binning enables an accurate coalescent-based estimation of the avian tree

Siavash Mirarab, Md. Shamsuzzoha Bayzid, Bastien Boussau, Tandy Warnow*

INTRODUCTION: Reconstructing species trees for rapid radiations, as in the early diversification of birds, is complicated by biological processes such as incomplete lineage sorting (ILS) that can cause different parts of the genome to have different evolutionary histories. Statistical methods, based on the multispecies coalescent model and that combine gene trees, can be highly accurate even in the presence of massive ILS; however, these methods can produce species trees that are topologically far from the species tree when estimated gene trees have error. We have developed a statistical binning technique to address gene tree estimation error and have explored its use in genome-scale species tree estimation with MP-EST, a popular coalescent-based species tree estimation method.

ON OUR WEB SITE

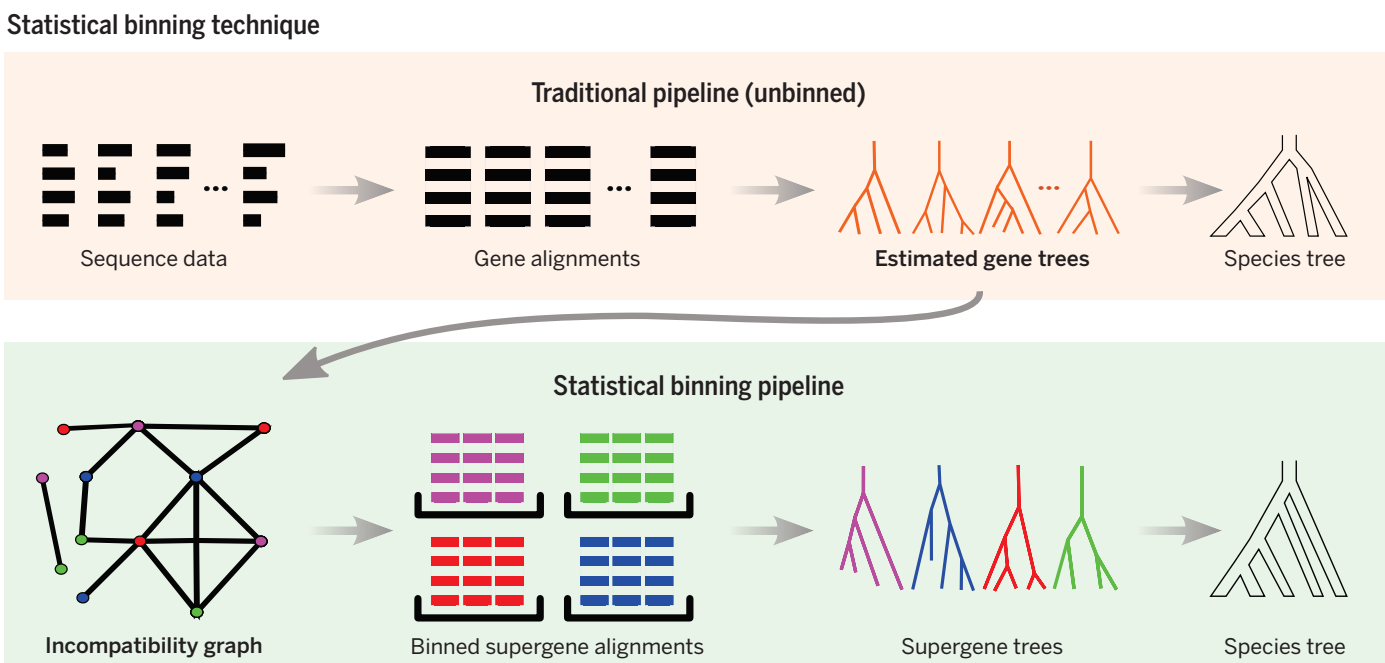
Read the full article at http://dx.doi.org/10.1126/science.1250463

RATIONALE: In statistical binning, phylogenetic trees on different genes are estimated and then placed into bins, so that the differences between trees in the same bin can be explained by estimation error (see the figure). A new tree is then estimated for each bin by applying maximum likelihood to a concatenated alignment of the multiple sequence alignments of its genes, and a species tree is estimated using a coalescent-based species tree method from these supergene trees.

RESULTS: Under realistic conditions in our simulation study, statistical binning reduced the topological error of species trees estimated using MP-EST and enabled a coalescent-based analysis that was more accurate than concatenation even when gene tree estimation error was relatively high. Statistical binning also reduced the error in gene tree topology and species tree branch length estimation, especially when the phylogenetic signal in gene sequence alignments was low. Species trees estimated using MP-EST with statistical binning on four biological data sets showed increased concordance with the biological literature. When MP-EST was used to analyze 14,446 gene trees in the avian phylogenomics project, it produced a species tree that was discordant with the concatenation analysis and conflicted with prior literature. However, the statistical binning analysis produced a tree that was highly congruent with the concatenation analysis and was consistent with the prior scientific literature.

CONCLUSIONS: Statistical binning reduces the error in species tree topology and branch length estimation because it reduces gene tree estimation error. These improvements are greatest when gene trees have reduced bootstrap support, which was the case for the avian phylogenomics project. Because using unbinned gene trees can result in overestimation of ILS, statistical binning may be helpful in providing more accurate estimations of ILS levels in biological data sets. Thus, statistical binning enables highly accurate species tree estimations, even on genome-scale data sets. ■

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: warnow@illinois.edu
Cite this article as S. Mirarab *et al.*, *Science* 346, 1250463 (2014). DOI: 10.1126/science.1250463



The statistical binning pipeline for estimating species trees from gene trees. Loci are grouped into bins based on a statistical test for combinabilty, before estimating gene trees.

Published by AAAS

RESEARCH ARTICLE

Statistical binning enables an accurate coalescent-based estimation of the avian tree

Siavash Mirarab,[1] Md. Shamsuzzoha Bayzid,[1] Bastien Boussau,[2] Tandy Warnow[1,3]*

Gene tree incongruence arising from incomplete lineage sorting (ILS) can reduce the accuracy of concatenation-based estimations of species trees. Although coalescent-based species tree estimation methods can have good accuracy in the presence of ILS, they are sensitive to gene tree estimation error. We propose a pipeline that uses bootstrapping to evaluate whether two genes are likely to have the same tree, then it groups genes into sets using a graph-theoretic optimization and estimates a tree on each subset using concatenation, and finally produces an estimated species tree from these trees using the preferred coalescent-based method. Statistical binning improves the accuracy of MP-EST, a popular coalescent-based method, and we use it to produce the first genome-scale coalescent-based avian tree of life.

Species trees provide a basis for understanding how life evolved on earth, as well as having applications to comparative genomics, orthology detection, protein function inference, and biodiversity analysis. Estimations of species trees are typically built with multiple loci (sometimes complete genes—but not always), in some cases, from throughout the genome. One advantage of such a phylogenomic approach is that it enables more data to be used in tree estimation (*1*). However, there is increasing evidence that loci can have conflicting evolutionary histories (so that their phylogenetic trees are topologically different) because of many biological causes, including incomplete lineage sorting (ILS) (*2*, *3*), a process that is especially common in rapid radiations, characterized by a succession of short branches in the phylogenetic tree, such as is believed to have occurred in the avian and mammalian evolutionary lineages (*4*–*9*). However, the standard phylogenetic estimation technique of concatenation, which concatenates alignments for individual loci into a combined data set called a supermatrix and then estimates the species tree from the supermatrix, can return incorrect species trees with high confidence in the presence of substantial ILS (*10*–*13*).

For this reason, many methods have been developed to estimate species trees that can be accurate even with high levels of ILS (*4*, *8*, *12*–*19*). For example, coestimation methods produce estimated gene trees and species trees directly from sequence alignments (*14*, *15*, *18*), and summary methods operate by combining estimated gene trees into a species tree. (Not all the loci in a phylogenomic analysis may be complete genes, e.g., some may contain only the exons or only the introns of some gene, and some may not be based on genes at all. However, to be consistent with other literature on the subject (*2*, *10*), we refer to phylogenetic trees on genomic loci as gene trees.) Some of these summary methods are created on the basis of the multispecies coalescent model (*20*) and are statistically consistent under that model (*12*, *13*), which means they will reconstruct the true species tree with high probability given a sufficiently large number of estimated gene trees that are error-free (*12*, *13*, *17*, *18*). A new type of coalescent-based method estimates the species tree directly from unlinked markers without also estimating gene trees (*21*–*23*).

However, the performance of coalescent-based methods has been mixed. Coestimation methods can have excellent accuracy but are too computationally intensive to use on data sets with hundreds of genes (*24*). The methods that estimate species trees directly from the sequence data without also estimating gene trees are not as computationally intensive as the coestimation methods but are much less well understood because they have only recently been developed. Furthermore, some [like SNAPP (*21*)] can only be used with biallelic markers and so are not suitable for estimating species trees for large data sets with deep divergences where biallelic markers are rare. Summary methods are by far the most frequently used method for species tree estimation and have produced good results on some biological data sets (*14*, *15*); however, for other data sets, the summary methods have not been able to produce highly supported trees (*25*), even with a large quantity of data (*26*). Simulation studies show that species trees estimated with summary methods can be less accurate than species trees estimated with concatenation, even in the presence of substantial ILS (*11*, *27*, *28*). A main reason for this disparity in performance is poor phylogenetic signal (e.g., because of short sequence lengths) in individual genes, which is a potential problem for coalescent-based summary methods (*28*, *29*). Moreover, many realistic biological conditions (including short branches in gene trees) make completely accurate gene tree estimation from limited sequence data highly unlikely (*30*).

Phylogenomic analyses can utilize very large numbers of genomic loci to estimate the species tree, but genome-scale data sets can contain loci that have reduced phylogenetic signal so that their estimated gene trees have reduced bootstrap support (BS) (*31*). Although it is not known how summary methods are affected when only some of the loci have low signal, coalescent-based summary methods have reduced accuracy on data sets where all the gene sequence alignments are short (*28*).

This challenge confronted the avian phylogenomics project (*31*), where a species tree estimated with a concatenated maximum-likelihood analysis on 14,446 loci had a succession of short branches suggestive of a radiation, and the tree also conflicted with estimated gene trees. Furthermore, most loci had low phylogenetic signal, which resulted in average BS of only 32% for the bifurcating maximum-likelihood trees estimated on these loci (fig. S1). Although much of the distance between estimated gene trees and the estimated species tree was related to the low support branches, even after collapsing low support branches, there was still substantial conflict among the gene trees (fig. S15 and supplementary text), suggestive of ILS. Thus, not only is there gene tree conflict (reducing the accuracy of concatenation) but the gene trees were generally poorly estimated (reducing the accuracy of summary methods).

Constructing phylogenies from genes with low phylogenetic signal is challenging, even if ILS is not an issue, and several approaches for selecting loci for use within a concatenated analysis have been suggested (*32*). However, restricting loci is problematic for statistically consistent coalescent-based summary methods, because the conditions under which they are guaranteed to be accurate (with high probability) require a large enough random sample of true gene trees; removing loci can violate this condition and potentially bias the analysis.

Statistical binning technique

A phylogenomic pipeline that uses a coalescent-based summary method begins with sequence alignments on different loci, estimates gene trees on each locus, and then combines the estimated gene trees into an estimated species tree using the summary method. The statistical binning step modifies the pipeline by using a binning technique to produce a different set of estimated gene trees that can be used with the summary method. We call the use of binning with a given summary method the binned version of the summary method.

How the statistical binning pipeline operates, given an input set of loci with their estimated

[1]Department of Computer Science, University of Texas at Austin, Austin, TX 78712, USA. [2]Laboratoire de Biométrie et Biologie Evolutive, CNRS, UMR5558, Université Lyon 1, 69622, Villeurbanne, France. [3]Department of Bioengineering and Computer Science, University of Illinois Urbana-Champaign, Champaign, IL 61820, USA.
*Corresponding author. E-mail: warnow@illinois.edu

sequence alignments and trees is shown in Fig. 1. We use BS values on branches of the estimated gene trees to partition the set of loci into bins of roughly equal sizes, so that each bin consists of a set of loci where differences in the estimated gene trees can be explained by gene tree estimation error. We concatenate the alignments of loci in each bin into a large alignment (called a supergene alignment) and compute trees on each supergene alignment using maximum likelihood (*33*); this produces a set of trees (called supergene trees), with one supergene tree for each bin. We then construct a species tree from the set of supergene trees using the desired summary method. Thus, the difference between the unbinned and binned versions of a summary method is the set of trees it uses to compute the species tree: The unbinned summary method uses the original set of gene trees, and the binned summary method uses the set of supergene trees.

Evaluation

We used biological and simulated data sets (*34*) to evaluate species trees estimated by using binned and unbinned summary methods, as well as concatenation using maximum likelihood under the GTR+Γ model, computed by randomized accelerated maximum likelihood (RAxML) (*33*). Three summary methods—the greedy consensus, matrix representation with parsimony (MRP) (*35*), and maximum pseudo-likelihood estimation of species trees (MP-EST) (*13*)—were applied to simulated avian and mammalian data sets by using the site-only multilocus bootstrapping procedure (*34*) [see (*36*) for a discussion of more elaborate approaches for bootstrapping multilocus data sets]. We chose MP-EST because it is statistically consistent under the multispecies coalescent model, has been used in several studies (*6*, *37*–*39*), and had better accuracy than other summary methods in some studies (*13*). However, the greedy consensus is inconsistent under the multispecies coalescent (*40*), and MRP and concatenation may also be inconsistent (*11*).

We generated simulated data sets (*34*) from two model species trees: one based on the avian phylogenomics project data set with 48 species and 14,446 loci (*31*), and one based on a mammalian data set with 37 species and 447 loci studied in (*6*). The default model conditions center on the average gene tree BS and ILS levels of these biological data sets, and we varied the model parameters to produce lower and higher ILS levels and estimated gene trees with varying BS to understand the impact of binning under a wide range of conditions. Each model species tree was computed by running MP-EST on the biological data, and we modified the branch lengths on the model species trees to produce other model conditions with different amounts of ILS (the 2× condition has doubled branch lengths and so reduces ILS, and the 0.5× condition has halved branch lengths and so increases ILS). We simulated gene trees within the model species trees (table S1) on the basis of the multispecies coalescent model (*20*). We evolved sequences of different lengths down the gene trees and used RAxML (*33*) with 200 bootstrap replicates to estimate gene trees with branch support on these sequence alignments.

The avian biological gene trees have very low average BS. Of the three types of genomic markers—exons, introns, and UCEs (ultraconserved markers) analyzed—the exons have the least signal (average BS 24%), the introns have the most (average BS 48%), and the UCEs are intermediate in support (average BS 39%). The longest introns (with at least 10,000 bp) have the highest average BS (59%) but represent a very small fraction of the total set of gene trees examined (only 638 of 14,446).

We modeled conditions that resembled the avian exons-only, UCEs-only, introns-only, and long introns–only data sets (fig. S1), with respect to their average BS values, and refer to these different model conditions by the partition type (below). The simulated mammalian data sets exhibit support levels of 63 and 79%, bracketing the 71% average BS values in the biological data. We varied the number of genes from 200 to 2000 for the avian data set and from 200 to 800 for the mammalian data set. Finally, we created mixed-model conditions, one with 14,350 genes for the avian simulation experiment and the other with 400 genes for the mammalian simulation experiment, to closely approximate the biological data sets in terms of the number of loci and average BS. Overall, the avian simulated data sets have higher levels of ILS and lower BS values than the mammalian data sets and so present a more challenging condition.

For the simulated data sets, we recorded the true species tree and true gene trees generated during the simulation process, which allows us to exactly quantify the topological error in the estimated tree (*34*). We computed the missing branch rate (also called the false-negative rate), which is the proportion of branches in the true tree that are missing from the estimated tree, as well as the false-positive rate, which is the proportion of branches in the estimated tree that

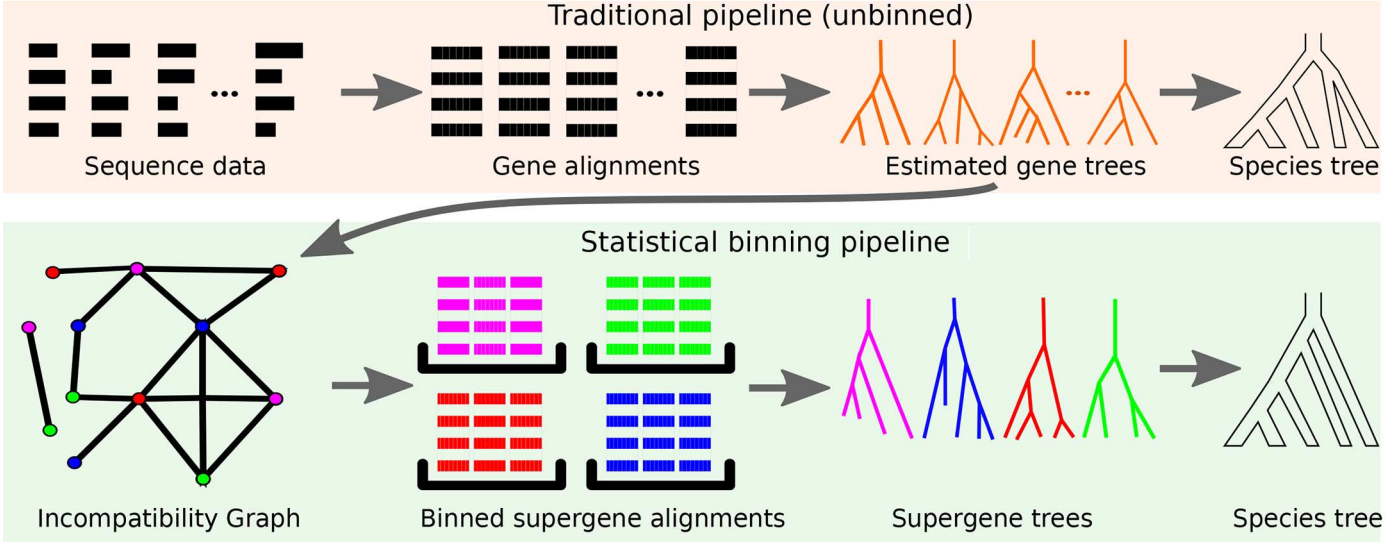


Fig. 1. Binning procedure. In traditional pipelines, gene trees are estimated from input sequence alignments and then combined into a species tree using a coalescent-based summary method. Statistical binning takes estimated gene trees and builds an incompatibility graph in which each node represents an estimated gene tree and each edge represents a detected incompatibility between two estimated gene trees at the specified statistical support threshold or higher. We use a heuristic we developed to color the nodes of the graph so that no two adjacent nodes have the same color and so that the color classes are of similar sizes. This coloring of the nodes defines a partition of genes into bins and ensures that no two genes with strongly supported conflict are put in the same bin. For each bin, individual gene alignments are concatenated to get a supergene alignment, from which a supergene tree is estimated using maximum-likelihood analysis. The supergene trees are then used as input to the summary method of choice to produce an estimated species tree.

are not present in the true tree. We measured how well the distribution of rooted gene trees is estimated by comparing triplet frequency distributions calculated from true gene trees and estimated gene trees; this is important because MP-EST uses estimated triplet distributions to construct the species tree. We measured estimation error in the species tree branch lengths as follows: Given a branch in an estimated species tree that is also present in the true species tree, we record the ratio of the length estimated for that branch by MP-EST to the true length of the branch (both in coalescent units) in the true (model) species tree (*34*).

Results

Unbinned MP-EST, binned MP-EST, and concatenation on 1000 avian genes with varying BS support in the gene trees are shown in Fig. 2A (see also table S2). Binned MP-EST was consistently and significantly more accurate than concatenation [$P < 10^{-5}$; all the statistical significance results reported henceforth are based on the two-way analysis of variance (ANOVA) test with Benjamini-Hochberg (BH) correction, and all the P values are reported in tables S4 and S5] and was also significantly more accurate than unbinned MP-EST ($P = 0.0001$). For gene trees with the highest BS values (i.e., long intron–like genes), both binned and unbinned MP-EST

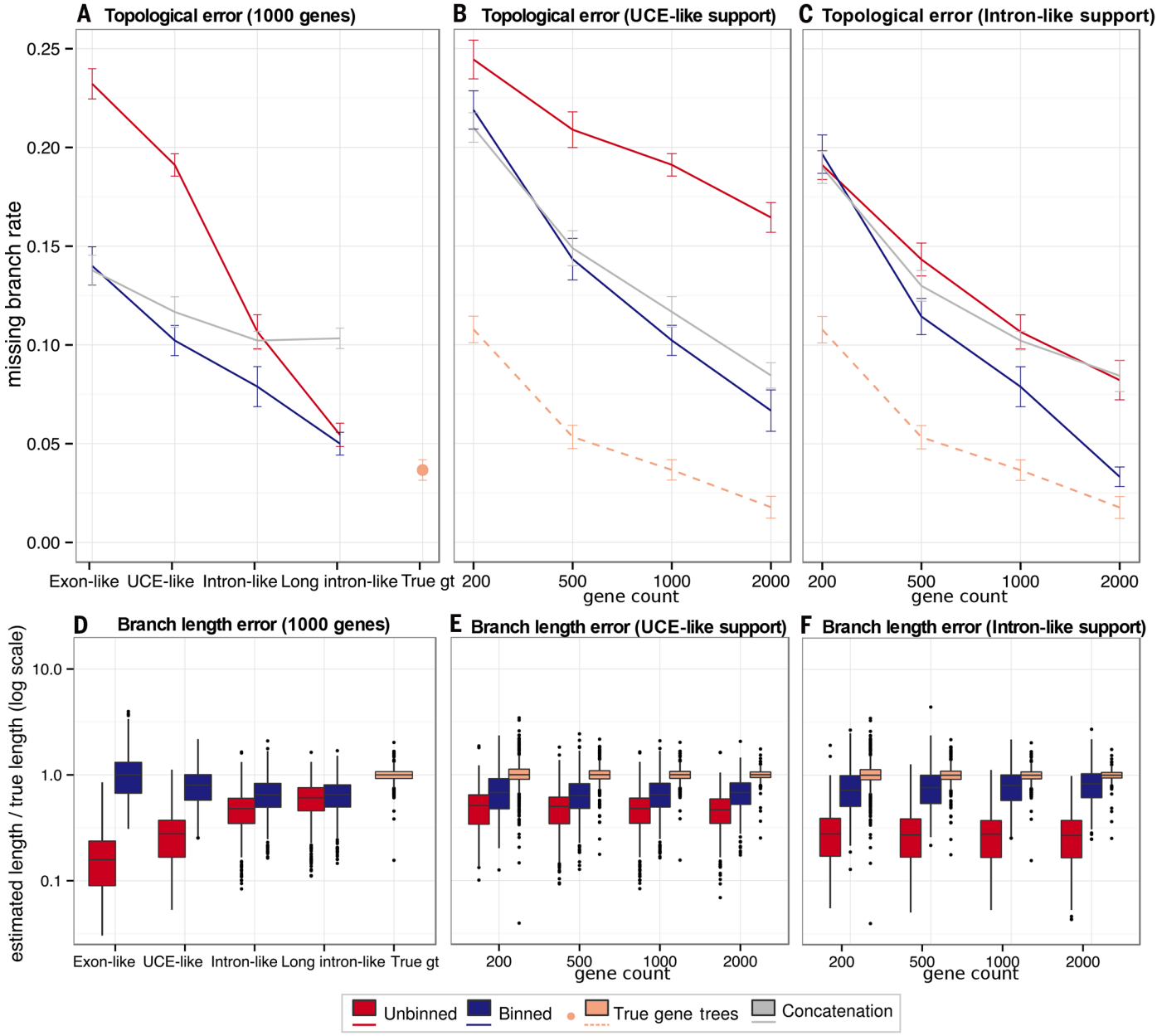


Fig. 2. Effect of binning on MP-EST on the simulated avian data sets with 1× ILS. (**A** to **C**) Species tree-topology error and (**D** to **F**) species tree branch–length error (boxplots with the ratio of estimated branch length to true branch length for branches of the true tree that appear in the estimated tree, 1 indicates correct estimation). MP-EST assigned an arbitrarily small length in the model tree to one branch, which we exclude from branch length calculations. (A) and (D) fix the number of genes to 1000 genes and vary gene tree support. The other panels fix gene tree support to UCE-like (B) and (E) and intronlike (C) and (F) and vary the number of genes. Results are over 10 replicates for the condition with 2000 genes, and 20 replicates for all other conditions. MP-EST has significantly lower topological error compared with MP-EST for (A), (B), and (C) ($P < 10^{-5}$, $P < 10^{-5}$, and $P = 0.002$, respectively), and concatenation for (A) ($P = 0.0001$) and (C) ($P = 0.01$). See also tables S2, S4, S5, and S6, and see fig. S7 for false-positive rates.

species trees had approximately the same error. However, as gene tree BS values decreased, the improvements obtained by binned MP-EST compared with unbinned MP-EST increased (P = 0.003 for the interaction effect; table S5). Concatenation was generally more accurate than unbinned MP-EST, except for gene trees with the highest BS. Results for MRP and Greedy showed similar trends (figs. S6 and S8; supplementary online text).

When loci exhibit BS corresponding to gene trees calculated on the UCE-like (moderate BS) or intronlike (high BS) markers (Fig. 2, B and C), binned MP-EST was more accurate than unbinned MP-EST ($P < 10^{-5}$ for UCE-like and P = 0.002 for intronlike markers). Furthermore, the advantage provided by binning increased with the number of UCE-like loci (the impact is significant with P = 0.003). Binned MP-EST tended to be more accurate than concatenation on both UCE-like and intronlike loci, but the differences are significant only for intronlike genes (P = 0.011). The improvement of binned MP-EST over concatenation appeared to increase with the number of intronlike loci, but the interaction effect is not significant (P = 0.087). Finally, on the mixed-model condition, concatenation and binned MP-EST each had 7% error, whereas all the other methods had at least 11% error (fig. S6). On the simulated mammalian data sets, binned MP-EST generally either matched or improved upon both unbinned MP-EST and concatenation (Fig. 3A). On the moderate (63%) BS trees, binned MP-EST and concatenation had close accuracy (with no significant differences), but unbinned MP-EST was significantly less accurate than binned MP-EST ($P < 10^{-5}$), and some conditions showed substantial differences (e.g., 800 loci). On higher BS (79%) loci, binned MP-EST was significantly more accurate than concatenation (P = 0.003), but there were no statistically significant differences between binned MP-EST and unbinned MP-EST. On the mixed-model condition, which most closely resembles the real mammalian data set in terms of the number of genes and gene tree support, binned MP-EST had only 1.8% error, concatenation had 3.7% error, and unbinned MP-EST had 4.6% error (Fig. 3C).

MP-EST always underestimated species tree branch lengths in coalescent units when analyzing estimated gene trees (in some cases by close to an order of magnitude), whereas the binned MP-EST trees had more accurate branch lengths (Figs. 2, D to F, and 3B). Because branch lengths are model parameters that determine the amount of ILS, underestimating branch lengths directly means overestimating ILS.

In the experiments where we varied the amount of ILS, binned MP-EST had lower average tree error than both unbinned MP-EST and concatenation, regardless of the amount of ILS (Fig. 4). The differences between binned and unbinned MP-EST are significant for both avian and mammalian data sets ($P < 10^{-5}$ and P = 0.0001, respectively), and differences between binned MP-EST and concatenation are significant on the avian data sets (P = 0.004). Furthermore, for

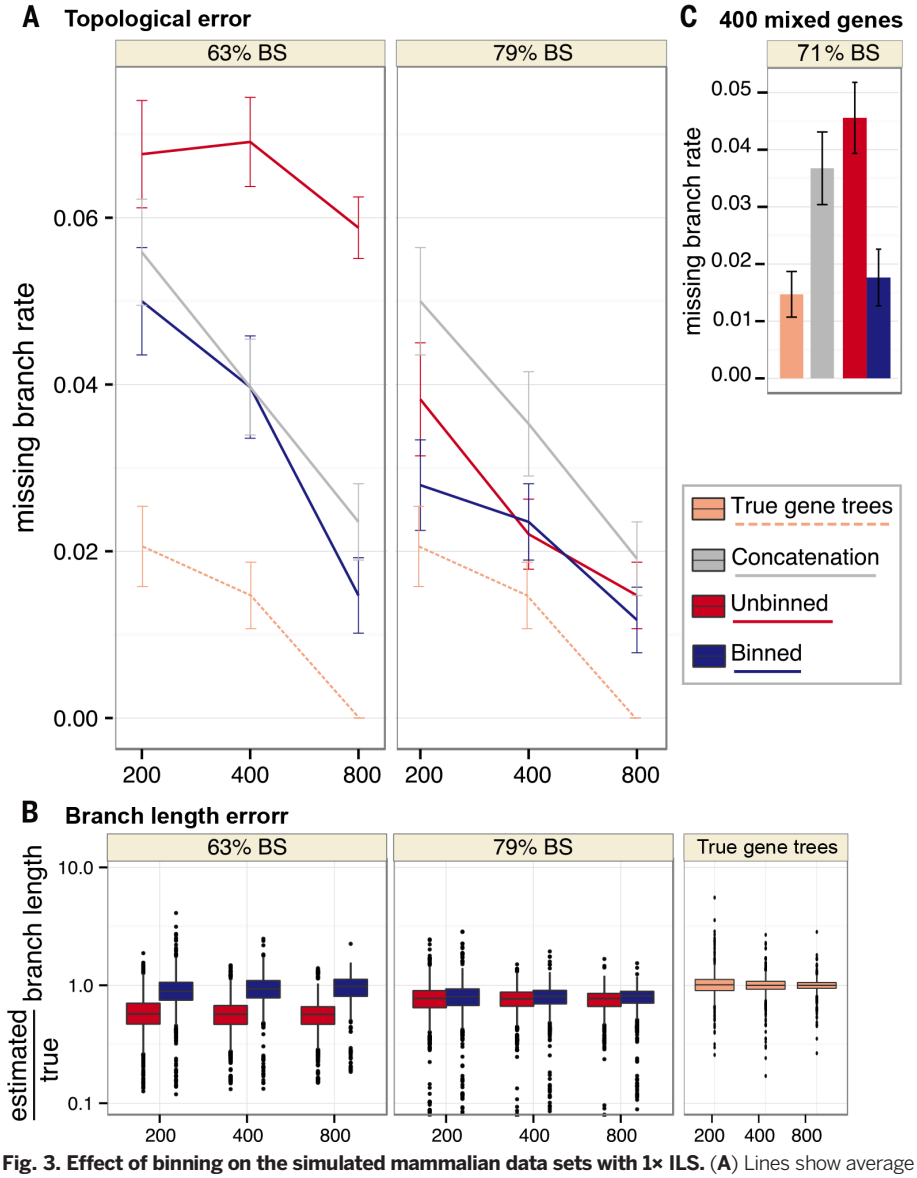


Fig. 3. Effect of binning on the simulated mammalian data sets with 1× ILS. (**A**) Lines show average topological tree error (missing branch rate) over 20 replicates for 200, 400, and 800 genes. Results are shown separately for gene trees with 63% and 79% BS. Binned MP-EST has significantly lower error compared to unbinned MP-EST for the 63% BS gene trees ($P < 10^{-5}$) and concatenation for the 79% BS gene trees (P = 0.003). (**B**) Error in branch lengths estimated by MP-EST in coalescent units is shown. (**C**) Topological error is shown for a mixed data set with 200 genes of 63% BS level and 200 genes of 79% BS level. See also table S3 and fig. S8.

the avian data sets, reducing the ILS level (2× condition) increased the impact of binning, and increasing the ILS level (0.5× condition) decreased the impact ($P < 10^{-5}$ for the interaction effect). The impact of ILS level on the mammalian data sets was similar but less pronounced and not statistically significant. For the reduced ILS (2×) models on both the avian and mammalian data sets, binned MP-EST was more accurate than unbinned MP-EST at estimating species tree topologies and branch lengths. For example, with 1000 UCE-like avian loci, unbinned MP-EST had 17.2% tree error, whereas binned MP-EST had only 5.9%. Performance on true gene trees provides an upper bound on what a summary method can achieve on estimated trees, and, as expected, MP-EST had its highest accuracy when run on true gene trees (Figs. 2 to 4).

Statistical binning improved the estimation of gene tree topologies (Table 1, but see also figs. S2 and S3), with the largest reductions in gene tree estimation error for the exonlike genes, and decreasing impact as the gene trees increased in BS. The reductions in triplet gene tree distribution estimation error were even larger (Table 1 and figs. S4 and S5), especially for loci with the lowest BS.

Biological data sets

We studied the avian data set (*31*) with 14,446 genes and 48 species, a mammalian data set with

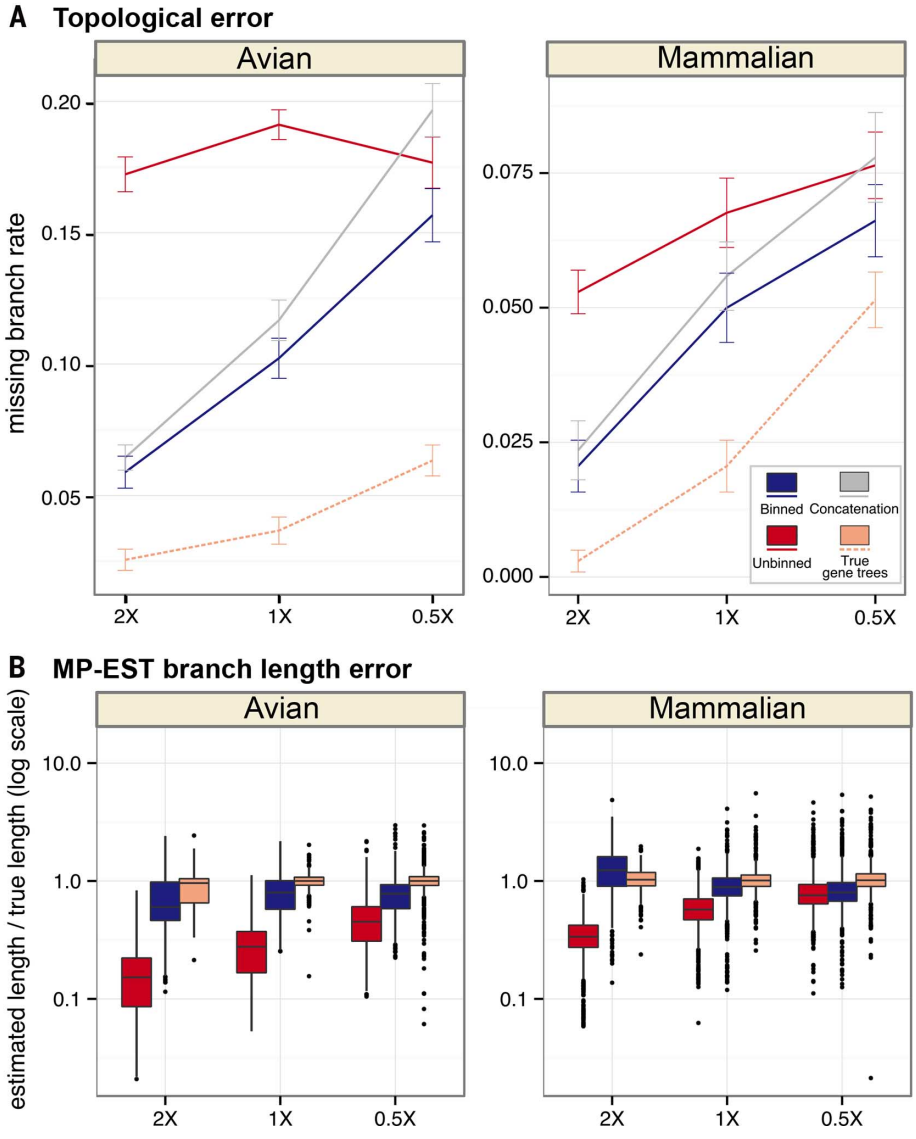


Fig. 4. Effects of ILS levels for the simulated avian and mammalian data sets. Levels of ILS are changed by multiplying all branch lengths in the model species tree by a factor of 0.5 (to increase ILS) or 2 (to reduce ILS). (**A**) Lines show average topological tree error (missing branch rate) over 20 replicates of 1000 UCE-like gene trees for avian and 200 gene trees with 63% BS for mammalian data sets. Binned MP-EST has significantly lower error compared with unbinned MP-EST for the avian ($P < 10^{-5}$) and the mammalian ($P = 0.0001$) data sets and concatenation for the avian data set ($P = 0.004$). (**B**) Error in branch lengths estimated by MP-EST in coalescent units.

447 genes and 37 species (*6*), yeast with 23 species and 1070 genes, vertebrates with 15 species and 1087 genes, and metazoa with 21 species and 225 genes (*32*). Each of these data sets shows evidence of gene tree discord (fig. S13), but they vary with respect to average BS (fig. S14). Species trees estimated using concatenation and gene trees were available (*31*, *32*), except for the maximum-likelihood gene trees from the mammalian data set, which we recomputed. We partitioned on the basis of individual loci when estimating supergene trees for the avian, metazoa, vertebrates, and yeast data sets (supplementary text). We present bin sizes (fig. S12), as well as a summary of the results obtained with binned and unbinned MP-EST on these data (Table 2).

Avian

The avian data set has very low average BS for almost all loci (fig. S1) and large topological distances between estimated gene trees. An unbinned MP-EST analysis of the full 14,446 loci produced a tree (Fig. 5A) with low to moderate support for some branches and failed to recover four key clades [Columbea (flamingo, grebe, pigeon, mesite, sandgrouse), Cursores (crane, killdeer), Otidimorphae (bustard, turaco, cuckoo), and Australaves (parrot, passerine, falcon, seriema)] that are recovered consistently in other analyses on the full genome data set (31), including concatenation analyses (fig. S17). Failure to recover Australaves is particularly surprising, as it has been recovered across different studies and types of data (*25*, *26*, *41*, *42*). In contrast, binned MP-EST on all 14,446 loci (Fig. 5A) had more highly supported branches and recovered all key clades.

An unbinned MP-EST tree generated on the introns-only data set (*31*) had 31 out of 45 edges with 100% support and 34 edges with 95% or higher BS; it also recovered all the key clades missing from the unbinned MP-EST tree computed on the full set of 14,446 loci. However, the binned MP-EST analysis (fig. S16) on the introns-only data set also recovered all the key clades and had higher support (33 edges with 100% support and 35 with 95% support or more), with increased support for some key novel clades (*31*).

Metazoa

The Metazoan data set also represents a challenging analysis, because the average BS is low (only 49%). The most important difference between the unbinned and binned MP-EST trees (Fig. 5B) is among Chordates, where the unbinned MP-EST tree put Cephalochordates (represented by *Branchiostoma floridae*) as sister to vertebrates (Craniates), and the binned MP-EST tree [as in the concatenation analysis, (fig. S20)], put Urochordates (represented by *Ciona intestinalis*) as sister to vertebrates. Although Cephalochordates were traditionally thought to be the sister to all the extant vertebrates (*43*), recent evidence supports Urochordates as the sister to all vertebrates (*44*–*46*), and hence, the binned MP-EST tree is likely correct. There are also some differences between the two trees within Protostomia, but both MP-EST trees had low support for those relations, and neither was congruent with the literature (supplementary text).

Mammalian

The mammalian data set has gene trees with substantially higher average BS (71%) but also demonstrates substantial gene tree incongruence. Differences between MP-EST and concatenation (using maximum likelihood) were observed for tree shrews and bats: The concatenated analysis put Scandentia (tree shrews) as sister to Glires (Rodentia/Lagomorpha), whereas the MP-EST analysis put Scandentia as sister to primates (*6*). We reanalyzed this data set and identified 21 loci with mislabeled sequences (subsequently confirmed by the authors) plus two outlier loci (fig. S18 and supplementary text). We removed these 23 loci and reanalyzed the data using concatenation and both binned and unbinned MP-EST (fig. S19). We recovered a concatenation tree topologically identical to the concatenation tree in (*6*). The unbinned MP-EST tree on this reduced gene set was similar to the unbinned MP-EST tree reported in (*6*) but had lower support for tree shrews as sister to primates [99% in (*6*), 64% with our analysis], and there was

Table 1. Gene tree estimation error, with and without binning for simulated data sets. Results are shown for fixed number of genes (1000 for avian and 200 for mammalian) and levels of ILS (1×, i.e., observed), but also see figs. S2 to S5. Individual gene tree (GT) error is mean topological distance, measured using the missing branch rate between the true gene tree and all 200 bootstrap replicates of each estimated gene tree. For the supergene trees, each bootstrap replicate of each supergene tree is compared separately against each true gene tree for the genes put in that bin. We also characterize gene tree distributions by calculating the triplet frequencies for all possible triplets, and we do this both for true and estimated gene trees (using all 200 bootstrap replicates of all genes and supergenes in the case of estimated trees). Thus, we obtain a true and an estimated triplet frequency distribution for each of the triplets. We report the mean Kullback-Leibler (KL) divergence of the estimated distribution from the true distribution.

Tree	Genomic markers	Length (bp)	Individual GT error (%)		GT distribution error (KL)	
			Unbinned	Binned	Unbinned	Binned
Avian (1000 genes)	Exonlike	250	79	57	0.234	0.025
	UCE-like	500	69	57	0.120	0.008
	Intronlike	1000	55	51	0.033	0.008
	Long intron–like	1500	46	45	0.011	0.007
Mammalian (200 genes)	63% BS	500	43	35	0.119	0.019
	79% BS	1000	27	26	0.038	0.027

Table 2. Results on the biological data sets. We compare MP-EST trees in terms of BS (the number of edges with BS equal to 100%, edges with BS at least 95%, and average BS), the distance between concatenation (DC) and MP-EST trees (number of missing branches), and with respect to biologically interesting differences between the two trees. The total number of branches in each tree is given parenthetically in the first column.

Tree (branches)	Gene trees	Bootstrap Support			DC	Interesting clades
		100%	>95%	Mean		
Avian (45)	Unbinned	29	34	0.95	12	Did not recover Australaves, Columbea, Cursores, and Otidimorphae, all recovered by other phylogenomic analysis
	Binned-50%	36	39	0.96	5	
Mammals (34)	Unbinned	30	30	0.98	2	Recovers Scandentia/Primates
	Binned-75%	30	31	0.98	1	Recovers Scandentia/Glires
Metazoa (18)	Unbinned	10	10	0.83	5	Rejects Olfactores (urochordates/vertebrates) and Eumetazoa
	Binned-75%	10	12	0.89	2	Rejects Eumetazoa
Vertebrates (15)	Unbinned	14	15	1	0	
	Binned-50%	14	14	0.99	0	
Yeast (20)	Unbinned	19	20	1	1	
	Binned-50%	19	19	0.98	1	

one topological difference among low support edges; the exact cause of these differences is not clear to us.

The binned MP-EST and unbinned MP-EST trees on the reduced gene set were very similar, but tree shrews were sister to Glires with 80% support in the binned MP-EST tree, just as their position was recovered in the concatenation tree. Thus, the placement of Scandentia, and whether it is sister to primates or to Glires, depends on the mode of analysis. This agreement between the binned MP-EST analysis and concatenated analysis of the reduced data set may be an important finding, but contradicts Janecka *et al.* (*47*) (which specifically addressed this question) and (*48*). However, these two studies did not use coalescent-based methods to estimate species trees. The unbinned and binned MP-EST trees placed bats identically as sister to all other Laurasitheria (except for the basal Eulipotyphyla) and so differed from the concatenation tree with respect to bats.

Vertebrates

The vertebrate data set had the highest average BS (76%) of all data sets we examined. Binned and unbinned MP-EST trees had the same topology, and both were topologically identical to the concatenation tree on the same data (fig. S21 and supplementary text).

Yeast

The yeast data set has relatively high average BS (72%). The binned and unbinned MP-EST topologies were identical, and both had 100% support for all but one branch (fig. S22).

Discussion

Our simulation results demonstrate that binning reduces error in estimated species tree topologies and branch lengths, gene tree topologies, and gene tree distributions under the conditions we studied. These reductions in error result in estimations of ILS that are closer to correct ILS levels than unbinned MP-EST, which tends to overestimate ILS levels. In our analyses, although unbinned methods are rarely more accurate than concatenation, binned MP-EST is almost always at least as accurate as concatenation, and there are many model conditions in which binned MP-EST is more accurate than concatenation, whereas unbinned MP-EST is less accurate than concatenation.

The biological data sets examined here show that binning affects analyses of data sets with less well supported gene trees (avian and metazoa) and has little impact on the yeast and vertebrate data sets, both of which have very well resolved gene trees. Notably, binning impacts the MP-EST analysis of the mammalian data set, which also has fairly well resolved gene trees. Where binning has an impact, binned MP-EST typically produces trees in closer agreement with accepted reconstructions than unbinned MP-EST. Binning reduces gene tree incongruence on biological data sets (fig. S13), which suggests that binned MP-EST may not overestimate ILS on biological data sets as much as unbinned MP-EST does. This trend is consistent with performance on simulated data and suggests that more-accurate estimates of ILS in biological data may be obtained through the use of statistical binning.

As binning can group genes together with different true topologies, it can result in misspecified models in the supergene tree estimation step and can reduce the accuracy of the estimated gene tree distributions. However, our simulations suggest that estimated gene tree distributions are more accurate after binning.

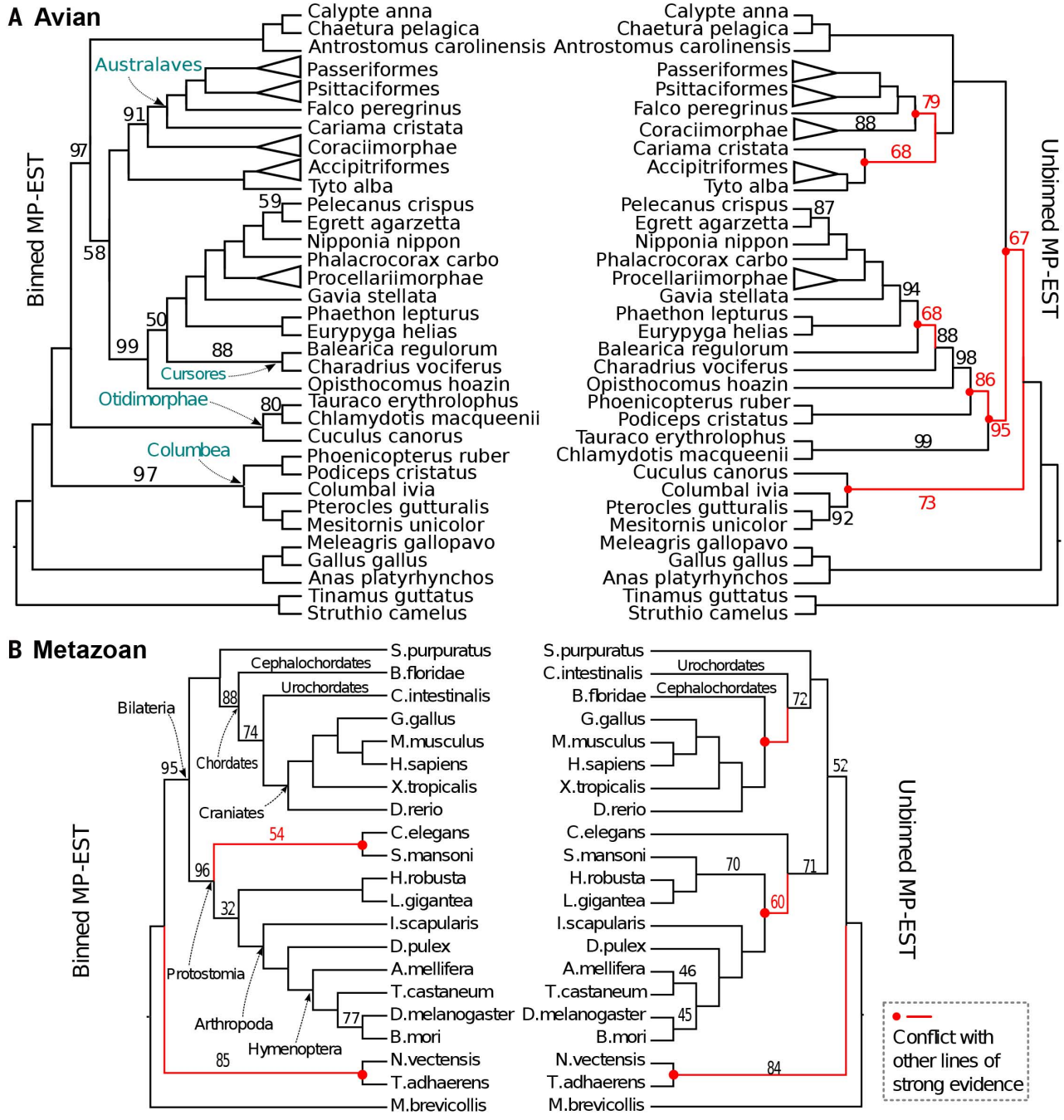


Fig. 5. Results on the (A) avian and (B) metazoan biological data sets using binned and unbinned MP-EST. Branches without designation represent 100% support.

We suggest that this is because binning will never group genes with different topologies together unless the conflicting branches had low support, which likely results from insufficient phylogenetic signal. As we have shown, the inclusion of poorly estimated gene trees distorts the estimated triplet gene tree distribution, and binning reduces this noise, which suggests that the overall impact of binning is beneficial. These results are also consistent with the observation that coalescent-based summary methods can be robust to recombination (*49*).

Our study explored gene tree estimation error arising from insufficient phylogenetic signal in the gene sequences; however, gene tree estimation error can also come from poorly estimated alignments (*50*) or errors introduced during the tree inference (*51*, *52*). Because our studies focused on insufficient phylogenetic signal, we have no evidence that binning could reduce phylogenetic error due to alignment error or misspecification for the sequence evolution model. Consequently, appropriate care should be devoted to obtaining good alignments and choosing an adequate model of sequence evolution to reconstruct both gene and supergene trees.

In our simulation, we only allowed ILS as a source of discord between true gene trees and true species trees; hence, these model conditions favor MP-EST (which is based on the same model used for simulations) over concatenation (which assumes no ILS is present). Given this, the fact that unbinned MP-EST is less accurate than concatenation in many conditions is noteworthy. Future studies based on model conditions in which other sources of gene tree discord

(e.g., duplication and loss, incorrect orthology assessments, recombination, introgression, horizontal gene transfer, and hybridization) are included would enable a better understanding of the relative accuracy of concatenation and coalescent-based species tree estimation and the impact of using binning under those conditions.

Statistical binning is just one step in a pipeline that begins with sequence alignments on different loci and ends with an estimated species tree, and variations of this pipeline might lead to improved accuracy. For example, gene trees could be estimated using Bayesian methods instead of maximum likelihood (*53*), and rigorous statistical tests for combinability (*54*) could be used instead of bootstrap branch support values; both variants might improve species tree estimations but would result in substantially increased running time. Bins could also be created without attempting to produce balanced sizes; in which case, bins could be weighted by their size. However, binning without attempting to evaluate whether genes have a common tree [as used in the "naïve binning" technique (*28*)] may not provide the improvements in accuracy seen here. Statistical binning is not likely to be useful for methods that estimate species trees directly from sequence data [e.g., (*21*–*23*)], because binning reduces the amount of data given to the method and can only be beneficial if it also improves the quality of the input data.

Our study demonstrates that binning is useful when the input is a set of estimated gene trees, because the method typically reduces the estimation error in the gene trees; however, binning sites together cannot improve the quality of sequence data. Furthermore, binning cannot be used with methods (like SNAPP) that are restricted to biallelic markers, because binning biallelic markers would create markers exhibiting at least three states and thus would violate that property.

The algorithmic techniques used in statistical binning are fast enough to use on data sets with many thousands of loci (such as the avian phylogenomics data set). Thus, statistical binning enables coalescent-based methods to be used on genome-scale data and can help to resolve challenging phylogenetic questions, including the avian Tree of Life.

Methods

The statistical binning technique includes a combinability test that evaluates whether a given pair of genes is likely to have significant topological incongruence, so that a concatenated analysis of those two genes is likely to be problematic. Because supergene trees can be estimated using partitioned concatenated analyses (which would allow the branch lengths and other model parameters to be reestimated for each gene within a partition), we only need to consider topological incongruence. Thus, we can place genes with the same true gene tree topology in the same bin, even if the trees differ in other respects (e.g., branch lengths).

We use maximum likelihood with bootstrapping to estimate gene trees with branch support values, and we say that a given pair of trees exhibits conflict at threshold t if there is a pair of incompatible branches (meaning they cannot coexist in any tree), one in each of the two gene trees, each with BS of at least t. Two trees that do not exhibit conflict at threshold t are combinable, and a set of trees for which all pairs are combinable is a combinable set.

Saying that two branches are incompatible means that no tree can be constructed that has both of these branches (*55*) (more specifically, no tree exists with branches that induce the bipartitions defined by these two branches). Thus, to test two trees for incompatibility at threshold t or higher, we collapse all branches in each tree with support below t, and then ask whether a tree exists that is a common refinement of these two collapsed trees. Testing for compatibility of two trees can be performed in linear time (*55*); hence, this calculation is fast.

The partitioning step uses a graph-based optimization, in which we build a graph in which each gene is represented by a node and an edge is present between two nodes (i.e., genes) if the estimated trees on that pair of genes exhibit conflict at threshold t. By definition, the graph depends on the parameter t; thus, smaller values for t will generally consider trees less likely to be combinable than larger values.

The graph created is called an incompatibility graph. To create bins from this graph, we color the nodes of the graph so that no two nodes with the same color are adjacent, and put all nodes with the same color into a common bin. Each bin thus contains a set of genes where no pairwise incompatibility has support of t or greater; hence, adjusting the support threshold t enables more aggressive or conservative binning. Once bins are formed, alignments of genes in the same bin are concatenated into a supergene alignment, and supergene trees are estimated on these supergene alignments using maximum likelihood. These supergene trees are used as input to the summary method of choice.

Because the statistically consistent methods use the distribution of gene trees to estimate the species tree, it is important for the supergene tree distribution to be close to the gene tree distribution; for this reason, we seek a node coloring in which the different color classes have approximately the same size (i.e., are balanced). We also seek a node coloring with a small number of colors, so that we have the largest bins we can, given the constraints imposed by combinability. However, finding a minimum node coloring (regardless of whether bins are balanced) is NP-hard and, consequently, is believed to not be solvable in polynomial time (*56*). Therefore, we developed a heuristic algorithm [based on the Brélaz heuristic (*56*) (see fig. S10)] for finding the smallest number of balanced bins (*34*). Our greedy heuristic algorithm processes genes one by one in a particular order and adds each gene to the smallest bin that has no incompatibility with it (*34*). When two or more bins have the same smallest size, the algorithm breaks the ties arbitrarily. Our experiments indicate that variations causes by these arbitrary choices mostly affect low support branches (see fig. S11), but also that a consensus tree from multiple runs of the binning approach can have higher average accuracy than individual runs (fig. S11). Therefore, if computational resources permit, we recommend that several runs of statistical binning be applied (each breaking ties differently) to produce several different estimated species trees. These trees can be compared with each other, to explore sensitivity and reliability, or a consensus tree of these trees can be used as a point estimate of the species tree.

We set the statistical support threshold t as follows. We note that using 75% for the BS has been a standard threshold for branch reliability (*57*), and so 75% represents a reasonable setting for t; however, when the data sets are large, we can afford to be more conservative and pick a smaller threshold. We also explored the effect of the support threshold (fig. S9) and saw that setting t to either 50 or 75% gave good results. Therefore, we set two thresholds: a conservative threshold of t = 50% that we use for data sets with at least 1000 genes and a moderate threshold of t = 75% that we use for the other data sets.

REFERENCES AND NOTES

1. A. Rokas, B. L. Williams, N. King, S. B. Carroll, Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804 (2003). doi: 10.1038/nature02053; pmid: 14574403
2. W. P. Maddison, Gene trees in species trees. *Syst. Biol.* **46**, 523–536 (1997). doi: 10.1093/sysbio/46.3.523
3. S. V. Edwards, Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1–19 (2009). doi: 10.1111/j.1558-5646.2008.00549.x; pmid: 19146594
4. L. L. Knowles, Estimating species trees: Methods of phylogenetic analysis when there is incongruence across genes. *Syst. Biol.* **58**, 463–467 (2009). doi: 10.1093/sysbio/syp061; pmid: 20525600
5. S. J. Hackett *et al.*, A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008). doi: 10.1126/science.1157704; pmid: 18583609
6. S. Song, L. Liu, S. V. Edwards, S. Wu, Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14942–14947 (2012). doi: 10.1073/pnas.1211733109; pmid: 22930817
7. R. W. Meredith *et al.*, Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011). doi: 10.1126/science.1211028; pmid: 21940861
8. J. H. Degnan, N. A. Rosenberg, Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**, 332–340 (2009). doi: 10.1016/j.tree.2009.01.009; pmid: 19307040
9. N. A. Rosenberg, Discordance of species trees with their most likely gene trees: A unifying principle. *Mol. Biol. Evol.* **30**, 2709–2713 (2013). doi: 10.1093/molbev/mst160; pmid: 24030555
10. S. V. Edwards, L. Liu, D. K. Pearl, High-resolution species trees without concatenation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5936–5941 (2007). doi: 10.1073/pnas.0607004104; pmid: 17392434
11. L. S. Kubatko, J. H. Degnan, Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* **56**, 17–24 (2007). doi: 10.1080/10635150601146041; pmid: 17366134
12. B. R. Larget, S. K. Kotha, C. N. Dewey, C. Ané, BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* **26**, 2910–2911 (2010). doi: 10.1093/bioinformatics/btq539; pmid: 20861028

13. L. Liu, L. Yu, S. V. Edwards, A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302 (2010). doi: 10.1186/1471-2148-10-302; pmid: 20937096
14. L. Liu, D. K. Pearl, Species trees from gene trees: Reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst. Biol.* **56**, 504–514 (2007). doi: 10.1080/10635150701429982; pmid: 17562474
15. L. Liu, BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* **24**, 2542–2543 (2008). doi: 10.1093/bioinformatics/btn484; pmid: 18799483
16. L. Liu, L. Yu, D. K. Pearl, S. V. Edwards, Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* **58**, 468–477 (2009). doi: 10.1093/sysbio/syp031; pmid: 20525601
17. E. Mossel, S. Roch, Incomplete lineage sorting: Consistent phylogeny estimation from multiple loci. *IEEE/ACM Trans. Comp. Biol. Bioinform.* **7**, 166–171 (2010). doi: 10.1109/TCBB.2008.66; pmid: 20150678
18. J. Heled, A. J. Drummond, Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* **27**, 570–580 (2010). doi: 10.1093/molbev/msp274; pmid: 19906793
19. Y. Yu, T. Warnow, L. Nakhleh, Algorithms for MDC-based multi-locus phylogeny inference: Beyond rooted binary gene trees on single alleles. *J. Comput. Biol.* **18**, 1543–1559 (2011). doi: 10.1089/cmb.2011.0174; pmid: 22035329
20. B. Rannala, Z. Yang, Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656 (2003). pmid: 12930768
21. D. Bryant, R. Bouckaert, J. Felsenstein, N. A. Rosenberg, A. RoyChoudhury, Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* **29**, 1917–1932 (2012). doi: 10.1093/molbev/mss086; pmid: 22422763
22. N. De Maio, C. Schlötterer, C. Kosiol, Linking great apes genome evolution across time scales using polymorphism-aware phylogenetic models. *Mol. Biol. Evol.* **30**, 2249–2262 (2013). doi: 10.1093/molbev/mst131; pmid: 23906727
23. J. Chifman, L. Kubatko, Quartet inference from SNP data under the coalescent model. *Bioinformatics* (2014). doi: 10.1093/bioinformatics/btu530
24. B. T. Smith, M. G. Harvey, B. C. Faircloth, T. C. Glenn, R. T. Brumfield, Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst. Biol.* **63**, 83–95 (2014). doi: 10.1093/sysbio/syt061; pmid: 24021724
25. R. T. Kimball, N. Wang, V. Heimer-McGinn, C. Ferguson, E. L. Braun, Identifying localized biases in large datasets: A case study using the avian tree of life. *Mol. Phylogenet. Evol.* **69**, 1021–1032 (2013). doi: 10.1016/j.ympev.2013.05.029; pmid: 23791948
26. J. E. McCormack *et al.*, A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLOS ONE* **8**, e54848 (2013). doi: 10.1371/journal.pone.0054848; pmid: 23382987
27. M. DeGiorgio, J. H. Degnan, Fast and consistent estimation of species trees using supermatrix rooted triples. *Mol. Biol. Evol.* **27**, 552–569 (2010). doi: 10.1093/molbev/msp250; pmid: 19833741
28. M. S. Bayzid, T. Warnow, Naive binning improves phylogenomic analyses. *Bioinformatics* **29**, 2277–2284 (2013). doi: 10.1093/bioinformatics/btt394; pmid: 23842808
29. S. Patel, R. Kimball, E. Braun, Error in phylogenetic estimation for bushes in the Tree of Life. *J. Phylogenet. Evol. Biol.* **1**, 2 (2013). doi: 10.4172/2329-9002.1000110
30. L. Nakhleh, U. Roshan, K. St. John, J. Sun, T. Warnow, Designing fast converging phylogenetic methods. *Bioinformatics* **17** (suppl. 1), S190–S198 (2001). doi: 10.1093/bioinformatics/17.suppl_1.S190; pmid: 11473009
31. E. D. Jarvis *et al.*, Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
32. L. Salichos, A. Rokas, Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**, 327–331 (2013). doi: 10.1038/nature12130; pmid: 23657258
33. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006). doi: 10.1093/bioinformatics/btl446; pmid: 16928733
34. Materials and methods are available as supplementary material on *Science* Online.
35. M. A. Ragan, Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* **1**, 53–58 (1992). doi: 10.1016/1055-7903(92)90035-F; pmid: 1342924
36. T.-K. Seo, Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol. Biol. Evol.* **25**, 960–971 (2008). doi: 10.1093/molbev/msn043; pmid: 18281270
37. A. D. Leaché, R. B. Harris, B. Rannala, Z. Yang, The influence of gene flow on species tree estimation: A simulation study. *Syst. Biol.* **63**, 17–30 (2013). doi: 10.1093/sysbio/syt049; pmid: 23939193
38. L. Zhao *et al.*, Phylogenomic analyses of nuclear genes reveal the evolutionary relationships within the BEP clade and the evidence of positive selection in Poaceae. *PLOS ONE* **8**, e64642 (2013). doi: 10.1371/journal.pone.0064642; pmid: 23734211
39. B. Zhong, L. Liu, Z. Yan, D. Penny, Origin of land plants using the multispecies coalescent model. *Trends Plant Sci.* **18**, 492–495 (2013). doi: 10.1016/j.tplants.2013.04.009;pmid: 23707196
40. J. H. Degnan, M. DeGiorgio, D. Bryant, N. A. Rosenberg, Properties of consensus methods for inferring species trees from gene trees. *Syst. Biol.* **58**, 35–54 (2009). doi: 10.1093/sysbio/syp008; pmid: 20525567
41. A. Suh *et al.*, Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Commun.* **2**, 443 (2011). doi: 10.1038/ncomms1448; pmid: 21863010
42. N. Wang, E. L. Braun, R. T. Kimball, Testing hypotheses about the sister group of the passeriformes using an independent 30-locus data set. *Mol. Biol. Evol.* **29**, 737–750 (2012). doi: 10.1093/molbev/msr230; pmid: 21940640
43. C. Nielsen, *Animal Evolution: Interrelationships of the Living Phyla* (Oxford Univer Press, Oxford, 2012).
44. F. Delsuc, H. Brinkmann, D. Chourrout, H. Philippe, Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006). doi: 10.1038/nature04336; pmid: 16495997
45. S. J. Bourlat *et al.*, Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006). doi: 10.1038/nature05241; pmid: 17051155
46. T. R. Singh *et al.*, Tunicate mitogenomics and phylogenetics: Peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. *BMC Genomics* **10**, 534 (2009). doi: 10.1186/1471-2164-10-534; pmid: 19922605
47. J. E. Janecka *et al.*, Molecular and genomic data identify the closest living relative of primates. *Science* **318**, 792–794 (2007). doi: 10.1126/science.1147555; pmid: 17975064
48. B. Boussau *et al.*, Genome-scale coestimation of species and gene trees. *Genome Res.* **23**, 323–330 (2013). doi: 10.1101/gr.141978.112; pmid: 23132911
49. H. C. Lanier, L. L. Knowles, Is recombination a problem for species-tree analyses? *Syst. Biol.* **61**, 691–701 (2012). doi: 10.1093/sysbio/syr128; pmid: 22215721
50. K. Liu, S. Raghavan, S. Nelesen, C. R. Linder, T. Warnow, Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science* **324**, 1561–1564 (2009). doi: 10.1126/science.1171243; pmid: 19541996
51. J. Felsenstein, Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401 (1978). doi: 10.2307/2412923
52. W. G. Weisburg, S. J. Giovannoni, C. R. Woese, The Deinococcus-Thermus phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst. Appl. Microbiol.* **11**, 128–134 (1989). doi: 10.1016/S0723-2020(89)80051-7; pmid: 11542160
53. M. DeGiorgio, J. H. Degnan, Robustness to divergence time underestimation when inferring species trees from estimated gene trees. *Syst. Biol.* **63**, 66–82 (2014). doi: 10.1093/sysbio/syt059; pmid: 23988674
54. J. W. Leigh, E. Susko, M. Baumgartner, A. J. Roger, Testing congruence in phylogenomic analysis. *Syst. Biol.* **57**, 104–115 (2008).doi: 10.1080/10635150801910436; pmid: 18288620
55. T. Warnow, Tree compatibility and inferring evolutionary history. *J. Algorithms* **16**, 388–407 (1994). doi: 10.1006/jagm.1994.1018
56. D. Brélaz, New methods to color the vertices of a graph. *Commun. ACM* **22**, 251–256 (1979). doi: 10.1145/359094.359101
57. T. P. Wilcox, D. J. Zwickl, T. A. Heath, D. M. Hillis, Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* **25**, 361–371 (2002). doi: 10.1016/S1055-7903(02)00244-0; pmid: 12414316

ACKNOWLEDGMENTS

SM was supported by a graduate fellowship from Howard Hughes Medical Institute (HHMI) M.S.B. was supported by a graduate fellowship from the Fulbright Foundation, by NSF under grant DBI-10735191 to iPLANT, and by the University of Alberta, made possible through a donation from Musea Ventures, which is held by G. K.-S. Wong; T.W. was supported by NSF under grant DBI-1062335; and B.B. was supported by a postdoctoral fellowship from the Human Frontier Science Program, the CNRS, and Agence Nationale de la Recherche (ANR) through grant ANR-10-BINF-01-01 Ancestrome. The authors thank E. Braun, G. Chacko, S. Edwards, T. Gilbert, E. Jarvis, L. Liu, D. Posada, and G. Zhang for helpful comments. The computational analyses in this study were performed using generous allocations on TACC (both Lonestar and Stampede) and on the Condor cluster of the University of Texas Computer Science department. All the data sets used in the paper and the software for performing statistical binning can be found at http://www.ideals.illinois.edu/handle/2142/55319.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1250463/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S6
References (*58*–*70*)

6 January 2014; accepted 3 September 2014
10.1126/science.1250463

Statistical binning enables an accurate coalescent-based estimation of the avian tree

Siavash Mirarab *et al.*

Science **346**, (2014);
DOI: 10.1126/science.1250463

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.