So many genes, so little time: comments on divergence-time estimation in the genomic era

Stephen A. Smith¹, and Joseph W. Brown¹†, Joseph F. Walker¹†
¹Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, 48109, USA
†Equal contribution

Corresponding author: Stephen A. Smith, Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, 48109, USA; E-mail: eebsmith@umich.edu.

Abstract

1. Phylogenomic datasets have emerged as an important tool and have been used for addressing questions involving evolutionary relationships, patterns of genome structure, signatures of selection, and gene and genome duplications. Here, we examine these data sources for their utility in divergence-time analyses. Divergence-time estimation can be complicated by the heterogeneity of rates among lineages and through time. Despite the recent explosion of phylogenomic data, it is still unclear what the distribution of gene- and lineage-specific rate heterogeneity is over these genomic and transcriptomic datasets.

2. Here, we examine rate heterogeneity across genes and determine whether clock-like or nearly clock-like genes are present in phylogenomic datasets that could be used to reduce error in divergence-time estimation. We address these questions with six
published phylogenomic datasets including Birds, carnivorous Caryophyllales, broad
Caryophyllales, Millipedes, Hymenoptera, and Vitales. We introduce a simple and
fast method for identifying useful genes for constructing divergence-time estimates
and conduct exemplar Bayesian analyses under both clock and uncorrelated
log-normal (UCLN) models.

3. We used a “gene shopping” method (implemented in SortaDate) to identify genes
with minimal conflict, lower root-to-tip variance, and discernible amounts of
molecular evolution. We find that every empirical dataset examined includes genes
with clock-like, or nearly clock-like, behavior. Many datasets have genes that are not
only clock-like, but also have reasonable evolutionary rates and are mostly
compatible with the species tree. We used these data to conduct basic
divergence-time analyses under strict clock and UCLN models. These exemplar
divergence-time analyses show overlap in age estimates when using either clock or
UCLN models, but with much larger credibility intervals for UCLN models.

4. We find that “gene shopping” can be productive and successful in finding gene regions
that minimize lineage-specific heterogeneity. By doing relatively simple assessments
of root-to-tip variance and bipartition conflict, we not only explore datasets more
thoroughly but also may estimate ages on phylogenies with lower error. We also
suggest the need to explore more detailed and informative approaches to determine fit
and deviation from a molecular clock, as existing approaches are exceedingly strict.

**Introduction**

Datasets based on thousands of genes from genomes and transcriptomes have emerged as a
major tool in addressing broad evolutionary questions including, but not limited to,
phylogenetic reconstruction, gene and genome duplication, and inference of molecular
evolutionary patterns and processes. And while these datasets have been used for
divergence-time estimation (e.g., [Jarvis et al. 2014b] [Prum et al. 2015]), their overall utility
for divergence-time analyses has not been fully examined. In particular, it is unclear
whether within these enormous datasets there exist nearly clock-like gene regions that may
aid in producing lower error divergence-time estimates. While some authors, such as Jarvis
et al. (2014b), have suggested choosing clock-like genes, a repeatable and fast procedure to
identify these genes has not been developed for phylogenomics and an examination of the
frequency of these genes in empirical datasets has not been conducted.

Divergence-time estimation is a complicated, but often essential, step for many
phylogenetic analyses. The sources of error include the ambiguous nature of fossil
placement, significant variation in the branchwise rates of evolution, significant variation in
the sitewise rates of evolution, uncertainty in the phylogenetic tree, topological dissonance
amongst gene trees due to incomplete lineage sorting, and complexity of the model for the
molecular clock (e.g., Smith et al. 2010, Dornburg et al. 2012, Parham et al. 2012, Beaulieu
et al. 2015). While fossils give the only available information for absolute age, their
placement and age carry uncertainty. Multiple fossil calibrations and complicated tree
shape priors can interact to further complicate molecular dating (Zhu et al. 2015, Heled
and Drummond 2015). Rate variation is common among individual branches and across
sites and can contribute to extensive deviations from the molecular clock. As a result,
complex models have been developed to accommodate for these deviations (Sanderson
2002, Drummond et al. 2006, Drummond and Suchard 2010). However, these more
parameter-rich models also carry with them significant uncertainty and can, when the data
deviate significantly from the model, positively bias results (e.g., Worobey et al. 2014).
Despite these difficulties, researchers continue to use divergence-time estimates extensively
as they remain essential for many downstream evolutionary and comparative analyses.

Researchers can take steps to ease sources of errors for divergence-time analyses.
For example, better use of fossils in temporal calibrations can dramatically improve
estimations (e.g., Parham et al. 2012, Ksepka et al. 2015), as does better accounting for
rate variation in the molecular models by improving model fit. Several relaxed clock
models have been introduced over the last few decades to accommodate rate heterogeneity
because most data do not conform to a strict clock. The most commonly used relaxed clock models include penalized likelihood (PL, Sanderson 2002) as implemented in r8s (Sanderson 2003) and treePL (Smith and O’Meara 2012), and the uncorrelated lognormal model (UCLN) as implemented in BEAST (Drummond and Rambaut 2007). This is not an exhaustive list as other methods have been developed and new ones are continually released (e.g., Lepage et al. 2007; Tamura et al. 2012; Heath et al. 2014). The diversity of techniques is matched with a variety of different inputs. For example, PL implementations minimally require an estimated phylogram, calibration, smoothing penalty value, and alignment size, while Bayesian estimation in BEAST minimally requires an alignment and priors to be set for each parameter, including any fossil calibrations.

Bayesian methods that use the UCLN model of rate variation (Drummond et al. 2006), such as that implemented in BEAST, simultaneously estimate phylogenetic relationships and divergence times, and so may be preferred over PL approaches as BEAST incorporates uncertainty more easily and explicitly. However, the computational burden of these simultaneous reconstruction methods limit their use to smaller datasets (i.e., not the entire genome or transcriptome). By “gene shopping” for genes that best conform to a molecular clock or relaxed molecular clock we can reduce larger datasets to datasets that are capable of being analyzed with BEAST or other programs. Before the recent development of next-generation sequencing techniques, this was not possible because of the relatively small number of genes available for any single clade. However, as genomic and transcriptomic datasets have become more readily available, “gene shopping” has become a potentially fruitful approach for divergence-time estimation (e.g., Jarvis et al. 2014b).

Nevertheless, the utility of these large genomic datasets for divergence-time estimation and the distribution of lineage-specific rate heterogeneity has yet to be fully explored.

Next generation sequencing techniques have dramatically increased the number of gene regions available for phylogenetic analysis. This has stimulated research into questions that are specifically pertinent to datasets with hundreds or thousands of genes. What is the best method for reconstructing the species tree (e.g., Gatesy and Springer 2014).
How many genes support the dominant species tree signal (e.g., Salichos et al. 2014; Smith et al. 2015)? Genomic datasets also allow us to examine the extent of molecular rate variation in genes, genomes, and lineages (Jarvis et al. 2014b; Yang et al. 2015). For example, Yang et al. (2015) explored the distribution of lineage-specific rate heterogeneity throughout transcriptomes of the plant clade Caryophyllales as it relates to life history, and Jarvis et al. (2014b) explored rate heterogeneity and selection as it relates to errors in phylogeny reconstruction in a genomic dataset of birds. Jarvis et al. (2014b) also filtered gene regions to identify “clock-like” genes for divergence-time estimation. While Jarvis et al. (2014b) conducted a filtering analysis on their genomic data, a thorough examination of lineage-specific rate heterogeneity for divergence-time estimation has not been conducted. Nevertheless, the availability of full genomes and transcriptomes makes identifying genes with lower rate variation possible and so are more suitable for divergence-time estimation.

Here, we examine six genomic and transcriptomic datasets across animals and plants and with different temporal and taxonomic scopes to examine the extent of lineage-specific rate heterogeneity. We investigate the distribution of variation in the branchwise rates of evolution across thousands of genes to understand whether these new genomic resources may improve divergence-time estimation by allowing for simpler models of molecular evolution. Finally, we introduce a simple sorting procedure to identify informative and nearly clock-like genes. This procedure can be used to examine the overall distribution of evolution, rate heterogeneity, bipartition concordance, and potential utility of genes for divergence-time analysis.

Materials and Methods

Dataset processing.— We used six published datasets to examine rate heterogeneity: Birds (BIR, Jarvis et al. 2014b), carnivorous Caryophyllales (CAR, Walker et al. 2017), the broader Caryophyllales (CARY, Yang et al. 2015), Vitales (VIT, Wen et al. 2013), Hymenoptera (HYM, Johnson et al. 2013), and Millipedes (MIL, Brewer and Bond 2013).
The range in datasets spans different taxonomic groups, datasets sizes (e.g., CAR vs CARY), and age (e.g., from hundreds of millions of years to within the last hundred million years). Where possible, we used orthologs that were identified using the Maximum Inclusion method of Yang and Smith (2015). This was the case with every dataset but BIR for which we used the exon alignments available online (Jarvis et al. 2014a; http://gigadb.org/dataset/101041). For each ortholog we have an estimated gene tree, based on maximum likelihood (ML) analyses, and alignments, from the original studies. Gene trees, regardless of the source of orthologs, were then rooted and SH-like tests were performed to assess confidence in edges (Anisimova and Gascuel 2006).

Gene tree analyses.— Because deviation from the clock is empirically manifest in a phylogram as variation in root-to-tip length among tips within a tree, we measured the variance of root-to-tip lengths for each tree. This was performed on each rooted ortholog, for which outgroups were removed, with the pxlstr program of Phyx package (Brown et al. 2017). We performed the standard clock test for each ortholog, with outgroup removed, using PAUP* (Swofford 2001) by calculating the ML score for a gene both with and without assuming a clock, and then performing a likelihood ratio test. In addition to assessing the clock-likeness of genes, we also compared gene tree topologies to the corresponding published species tree topology. Branch lengths were not available for some species trees. To compare the individual gene trees to their corresponding species trees, we conducted bipartition comparison analyses on each gene tree using pxbp from the Phyx package (procedure described in Smith et al. 2015).

Simulations.— We conducted simulations to examine expectations of rate variation given clock-like, noisy clock-like, and uncorrelated lognormal data. We first generated simulated clock-like data using Indelible v1.03 (Fletcher and Yang 2009) using the WAG model with 500 characters for amino acid datasets, and JC with 1500 characters for nucleotide datasets, on each of the empirical species tree topologies. For these data simulations, because the species tree often had no branch lengths available, node heights were first simulated randomly using Indelible and then the tree was rescaled to a height of 0.25,
0.5, or 0.75. We used the trees generated by \textbf{Indelible} to further simulate 100 noisy clock
trees (rate=1.0, noise=0.25, and rate=1.0, noise=0.75) and uncorrelated lognormal (UCLN)
trees (mean.log=-0.5, sd.log=0.5, and mean.log=-0.5, sd.log=1.0) using \textbf{NELSI} v0.21 (Ho
\textit{et al.}, 2015). We note the ‘noise’ in \textbf{NELSI} corresponds to the standard deviation of a
normal distribution with mean = 0. For the noisy clock, branch-specific rates are a sum of
the global rate (here, 1.0) and a draw from this normal distribution. The simulations with
noise=0.75 thus are only loosely clock-like, and serve as a comparison between the more
clock-like (noise=0.25) and UCLN analyses. We used \textbf{RAxML} v8.2.3 (Stamatakis 2014) to
reconstruct each of these datasets. For each simulation, we examined the rate variation and
the root-to-tip length variation on the reconstructed phylograms.

While the focus of this study is not the performance of divergence-time estimation
methods, we still wanted to examined an exemplar from the simulations to ascertain the
variation in the results given different clock models. We used one random realization of
node heights as simulated from the \textbf{Indelible} analyses as mentioned above to generate
two datasets with \textbf{NELSI}. One dataset had three genes generated from a clock rate of 1 and
noise at 0.25, and the other dataset had three genes generated from a UCLN model and
mean.log at -0.5 and sd.log=1. As above, each amino acid gene consisted of 500 residues,
while DNA genes consisted of 1500 nucleotides. For both the noisy clock-like and UCLN
datasets, we conducted \textbf{BEAST} analyses with both a clock model and a UCLN model. A
birth-death tree prior was used as the prior for all node heights, and runs were conducted
for 10,000,000 generations with the first 10% discarded as burnin. Results were
summarized using \textbf{treeannotator} from the \textbf{BEAST} package. Median node heights as well as
95% HPD node heights were compared between the simulated datasets and the tree used
to generate these datasets.

\textit{Sorting and dating analyses on real data.}— In addition to these analyses on simulated
datasets, we conducted divergence-time analyses on a subset of the empirical datasets.
Because these datasets consist of hundreds to thousands of genes, we developed a sorting
procedure intended to mimic that which would be performed as a “gene shopping”
analysis. The sorting procedure relies on the root-to-tip variance statistic, bipartition
calculation to determine the similarity to the species tree, and total treelength. We sorted
first by the similarity to the species tree, then root-to-tip variance, and finally treelength.
We limited the results to the top three genes reported from the sorting procedure. Because
we filtered for genes that were consistent with the species tree, these genes were then
concatenated and the topology was fixed to be consistent with the species tree. For each of
these datasets, we conducted two BEAST analyses, one assuming a strict clock and the other
assuming a UCLN model. Because specific dates were not the focus of this examination,
the birth-death tree prior was used instead of fossil priors for nodes. The analyses were run
for 10,000,000 generations with the first 1,000,000 discarded as burnin. As above, results
were summarized using treannotator from the BEAST package. Median node heights and
95% HPD node heights were compared between the clock and UCLN runs as the node
heights on the true phylogeny are unknown.

Availability of procedures.— The analyses that were performed above can be conducted
using the SortaDate package (with source code and instructions available at
https://github.com/FePhyFoFum/sortadate). This package is written in Python and
available as an Open Source set of procedures. In some cases, external programs are used
(e.g., those found in the PhyX package) that are also Open Source and freely available.

Results and Discussion

A fundamental question for each of the empirical datasets is: are there clock-like gene
regions present within the genome? Results were broad, from 0.4% of genes passing the
clock test for the VIT dataset to 17% for the MIL dataset (see Table 1). The variation in
the percentage of deviation from the clock may reflect dataset size and the age of the clade
involved. As for size, the CAR dataset has 7 taxa that are not included in the CARY
dataset but otherwise overlaps partially and has far fewer taxa in total. The CARY
dataset, in addition to being much larger, also contains known shifts in life history (Yang
et al. 2015). These differences may account for the variation between these two datasets.
As for clade age, HYM and MIL are significantly older than the other datasets, which may account for their rate variation. Nevertheless, each dataset indeed had at least a few orthologs that passed a strict clock test even if these orthologs were in the small minority. Because passing a clock test does not necessarily indicate that the gene would be good for phylogenetic reconstruction, we also measured treelength and root-to-tip variance for each ortholog (see Figures 2-3). Clock tests are stringent in their need to conform to the clock (see below) and so by examining the root-to-tip variation and lineage-specific variation, we are more directly examining the deviation from ultrametricity. Although this is primarily descriptive and does not include a formal test, this provides an easily interpretable characterization of rate variation. We found that the datasets vary dramatically with no discernible general pattern for both root-to-tip variance and treelength. For example, the BIR dataset demonstrates very little molecular evolution as demonstrated by the short treelengths. For this dataset, we analyzed nucleotides (rather than amino acids) to maximize treelengths as [Jarvis et al. (2014b)] demonstrated low rates of evolution, especially deep in the phylogeny. However, the inferred rates of evolution (as determined by overall tree length) were still low. Given the difficulty in resolving the avian phylogeny, this pattern is perhaps to be expected ([Jarvis et al. 2014b]). This same pattern is present in the VIT dataset, though this was not explored as thoroughly in the original publication. Both the CAR and CARY datasets show a pattern of increasing variance with greater treelength (Figures 2). This contrasts with the HYM and MIL datasets that are clock-like even with longer treelengths (Figure 3). Lineage-specific rate variation in each dataset was idiosyncratic with most extreme variation in the outgroups. While outgroups were excluded for clock tests and in determining root-to-tip variance for “gene shopping”, we allowed outgroups to remain for lineage-specific rate variation analyses as in the right handed plots of Figure 3. The VIT dataset was an exception with several lineages other than the outgroup having high rates. In each dataset, there were genes that fell within the distribution of simulated trees that are clock-like or clock-like with low noise. One potential benefit of identifying orthologs with lower lineage-specific rate
variation within phylogenomic datasets is to use these, or a subset of these, orthologs to conduct divergence-time analyses. The hope is that by using clock-like genes, we may overcome or lessen the impact of lineage-specific rate variation on the error of divergence time analyses. The non-identifiability of rates and dates (e.g., longer branch lengths may be the result of a long time or fast evolution) is exacerbated by lineage-specific rate heterogeneity. We used a subset of orthologs to conduct divergence time analyses and we implemented a sorting procedure (packed in SortaDate) to (i) filter the genes that best reflect the species tree (i.e., higher bipartition concordance with the species tree), (ii) have lower root-tot-ip variance (i.e., most clock-like), and discernible amounts of molecular evolution (i.e., greater tree length; Figure 1). For each empirical dataset, we generated such an alignment (see Table 2). The genes that were filtered and used for divergence-time analyses for the BIR, CARY, VIT, and HYM datasets rejected the clock. The genes for the CAR and MIL datasets either didn’t or weakly rejected the clock. Resulting HPD trees were rescaled so that the root heights were equivalent to allow for easier comparisons between datasets. Typically, fossil placements would be used for scaling but because these are not intended to be runs for future use, we eliminated fossil placements as one source of variation. We found rough correspondence of node heights between the clock and UCLN analyses, especially for the four smallest datasets (see Figure 4). The UCLN analyses, as expected, had far greater variance in the 95% HPDs for node ages. We found the greatest differences in the larger BIR and CARY datasets (see Table 3). For both datasets, we saw major differences in tree heights, especially for CARY. This may reflect the size of the dataset or the underlying rate variation in the datasets. In general, strict clock estimates resulted in younger median node ages than analogous UCLN estimates, as well as younger maximum and older minimum 95% HPD values (see Table 3). The covariance statistics for UCLN runs ranged from the lowest mean values of -0.002 (stdev=0.04) in MIL to the highest of 0.246 (stdev=0.12) in BIR.

As is always the problem with real datasets, the true divergence-times are unknown. So we conducted exemplar analyses. For each empirical dataset, we simulated data for
three genes under both noisy clock and UCLN models to examine the variation in the
resulting divergence-time analyses where the true dates were known. For these simulated
datasets, a strict clock was rejected in each case, including those datasets that were
simulated under a clock with noise. We compared the resulting node heights from the
divergence time analyses under clock and UCLN models with the tree used for simulation
(see Tables 4-5 and Figure 5). For the datasets generated under a noisy clock model, more
of the true node heights were found in the 95% HPD interval when using the UCLN model
for inference than the strict clock model for inference. However, the precision as measured
by the total width of the 95% HPD interval for the UCLN runs were much lower than the
clock runs (see Tables 4-5). Those nodes that were not within the interval of the 95% HPD
when using the strict clock model for reconstruction, were close to the true value. So, while
fewer true node ages were contained in the strict clock HPDs, the overall error rate was
lower. For example, in the CARY dataset, while fewer nodes in the clock estimate were
found to be within the interval (52 vs 67 for the UCLN), the distance of the interval from
the estimate was lower for the clock dataset for both the high and low value for the 95%
HPD. Stated another way, the UCLN intervals were large enough that the true age was
often included, but this was at the cost of far lower precision. Because of this error relative
to the strict clock, the UCLN perhaps should not be the preferred model, especially if the
researcher is going to use a single summary tree for future analyses.

Several gene trees from the examples discussed fail a standard strict clock test but
have low tip-to-root variance. To explore this further, we simulated strict clock amino acid
and nucleotide data on orthologs from each empirical dataset and examined the frequency
of incorrectly rejecting a strict clock. The false rejection rate for clock tests using amino
acid data and a strict clock were between 5% and 8%. For the two nucleotide datasets, the
rejection rate was much higher at 23% and 46%. This suggests that for amino acid data,
the false rejection rate was near the nominal value, while for the nucleotide datasets the
false rejection rate was unreliable. Both nucleotide datasets (BIR and CARY) also had the
largest number of species and so the rejection rate may be a function of the number of
taxa. Sensitivity of the clock-test to nucleotide data is not the focus of this study, but should be examined in more detail. Also, it would be more informative to examine the deviation from the clock instead of a boolean test of significant fit. In regard to divergence time estimation, if a strict or stricter clock can be used, molecular phylogenies may be dated with significantly lower error. As an added benefit, fewer fossils would be necessary to calibrate nodes (and indirectly, rates). We suggest that the community explore model fit to relaxed clock models as well as potential alternatives to the prevailing strict clock test that may be more beneficial for divergence time estimates and more informative in regard to rate heterogeneity in phylogenomic datasets.

Acknowledgements

We would like to thank Oscar Vargas, Greg Stull, Jianjun Jin, Drew Larson, Ning Wang, Caroline Parins-Fukuchi, Edwige Moyroud, and Lijun Zhao for discussion of the manuscript and method. JFW and SAS were supported by NSF 1354048. JWB and SAS were supported by NSF 1207915.

Data availability

All unpublished analyses and datasets are available through Data Dryad (#XXXXXX). Associated scripts related to the method are available on GitHub.

Author contributions

SAS, JWB, and JFW conceived of the project. SAS and JFW analyzed the data. SAS, JWB, and JFW wrote the manuscript.
References

Anisimova, M. and Gascuel, O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, 55(4): 539–552.

Beaulieu, J. M., O’Meara, B. C., Crane, P., and Donoghue, M. J. 2015. Heterogeneous rates of molecular evolution and diversification could explain the triassic age estimate for angiosperms. *Systematic Biology*, 64(5): 869–878.

Brewer, M. S. and Bond, J. E. 2013. Ordinal-level phylogenomics of the arthropod class diplopoda (millipedes) based on an analysis of 221 nuclear protein-coding loci generated using next-generation sequence analyses. *PLOS ONE*, 8(11).

Brown, J. W., Walker, J. F., and Smith, S. A. 2017. phyx: Phylogenetic tools for unix. *Bioinformatics*. accepted.

Dornburg, A., Brandley, M. C., McGowen, M. R., and Near, T. J. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (mammalia: Cetacea). *Molecular Biology and Evolution*, 29(2): 721–736.

Drummond, A. J. and Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1): 214.

Drummond, A. J. and Suchard, M. A. 2010. Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, 8(1): 114.

Drummond, A. J., Ho, S. Y., Phillips, M. J., Rambaut, A., et al. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5): 699.

Fletcher, W. and Yang, Z. 2009. Indelible: a flexible simulator of biological sequence evolution. *Molecular Biology and Evolution*, 26(8): 1879–1888.
Gatesy, J. and Springer, M. S. 2014. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution*, 80: 231 – 266.

Heath, T. A., Huelsenbeck, J. P., and Stadler, T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, 111(29): E2957–E2966.

Heled, J. and Drummond, A. J. 2015. Calibrated birth–death phylogenetic time-tree priors for bayesian inference. *Systematic Biology*, 64(3): 369–383.

Ho, S. Y. W., Duchêne, S., and Duchêne, D. 2015. Simulating and detecting autocorrelation of molecular evolutionary rates among lineages. *Molecular Ecology Resources*, 15(4): 688–696.

Jarvis, E., Mirarab, S., Aberer, A., Houde, P., Li, C., Ho, S., Faircloth, B., Nabholz, B., Howard, J., Suh, A., Weber, C., Fonseca, R., Alfaro-Nunez, A., Narula, N., Liu, L., Burt, D., Ellegren, H., Edwards, S., Stamatakis, A., Mindell, D., Cracraft, J., Braun, E., Warnow, T., Jun, W., Gilbert, M., and Zhang, G. 2014a. Phylogenomic analyses data of the avian phylogenomics project.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y. W., Faircloth, B. C., Nabholz, B., Howard, J. T., Suh, A., Weber, C. C., da Fonseca, R. R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M. S., Zavidovych, V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W. C., Ray, D., Green, R. E., Bruford, M. W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E. P., Bertelsen, M. F., Sheldon, F. H., Brumfield, R. T., Mello, C. V., Lovell, P. V., Wirthlin, M., Schneider, M. P. C., Prosdocimi, F., Samaniego, J. A., Velazquez, A. M. V., Alfaro-Núñez, A., Campos, P. F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D. M., Zhou, Q., Perelman, P.,
Driskell, A. C., Shapiro, B., Xiong, Z., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J.,
Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F. E., Braun,
M., Fjeldsa, J., Orlando, L., Barker, F. K., Jønsson, K. A., Johnson, W., Koepfli, K.-P.,
O’Brien, S., Haussler, D., Ryder, O. A., Rahbek, C., Willerslev, E., Graves, G. R.,
Glenn, T. C., McCormack, J., Burt, D., Ellegren, H., Alström, P., Edwards, S. V.,
Stamatakis, A., Mindell, D. P., Cracraft, J., Braun, E. L., Warnow, T., Jun, W., Gilbert,
M. T. P., and Zhang, G. 2014b. Whole-genome analyses resolve early branches in the
tree of life of modern birds. *Science*, 346(6215): 1320–1331.

Johnson, B. R., Borowiec, M. L., Chiu, J. C., Lee, E. K., Atallah, J., and Ward, P. S. 2013.
Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. *Current
Biology*, 23(20): 2058 – 2062.

Ksepka, D. T., Parham, J. F., Allman, J. F., Benton, M. J., Carrano, M. T., Cranston,
K. A., Donoghue, P. C. J., Head, J. J., Hermen, E. J., Irmis, R. B., Joyce, W. G., Kohli,
M., Lamm, K. D., Lee, D., Patané, J. L., Polly, P. D., Phillips, M. J., Smith, N. A.,
Smith, N. D., van Tuinen, M., Ware, J. L., and Warnock, R. C. M. 2015. The fossil
calibration database—a new resource for divergence dating. *Systematic Biology*, 64(5):
853–859.

Lepage, T., Bryant, D., Philippe, H., and Lartillot, N. 2007. A general comparison of
relaxed molecular clock models. *Molecular Biology and Evolution*, 24(12): 2669–2680.

Mirarab, S., Bayzid, M. S., Boussau, B., and Warnow, T. 2014. Statistical binning enables
an accurate coalescent-based estimation of the avian tree. *Science*, 346(6215).

Parham, J. F., Donoghue, P. C. J., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A.,
Inoue, J. G., Irmis, R. B., Joyce, W. G., Ksepka, D. T., Patané, J. S. L., Smith, N. D.,
Tarver, J. E., van Tuinen, M., Yang, Z., Angielczyk, K. D., Greenwood, J. M., Hipsley,
C. A., Jacobs, L., Makovicky, P. J., Müller, J., Smith, K. T., Theodor, J. M., Warnock,
R. C. M., and Benton, M. J. 2012. Best practices for justifying fossil calibrations.
*Systematic Biology*, 61(2): 346–359.
Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., and
Lemmon, A. R. 2015. A comprehensive phylogeny of birds (aves) using targeted
next-generation dna sequencing. *Nature*, 526(7574): 569–573.

Roch, S. and Warnow, T. 2015. On the robustness to gene tree estimation error (or lack
thereof) of coalescent-based species tree methods. *Systematic Biology*, 64(4): 663–676.

Salichos, L., Stamatakis, A., and Rokas, A. 2014. Novel information theory-based measures
for quantifying incongruence among phylogenetic trees. *Molecular Biology and
Evolution*, 31(5): 1261–1271.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence
times: a penalized likelihood approach. *Molecular Biology and Evolution*, 19(1): 101–109.

Sanderson, M. J. 2003. r8s: inferring absolute rates of molecular evolution and divergence
times in the absence of a molecular clock. *Bioinformatics*, 19(2): 301–302.

Smith, S. A. and O’Meara, B. C. 2012. treepl: divergence time estimation using penalized
likelihood for large phylogenies. *Bioinformatics*, 28(20): 2689–2690.

Smith, S. A., Beaulieu, J. M., and Donoghue, M. J. 2010. An uncorrelated relaxed-clock
analysis suggests an earlier origin for flowering plants. *Proceedings of the National
Academy of Sciences*, 107(13): 5897–5902.

Smith, S. A., Moore, M. J., Brown, J. W., and Yang, Y. 2015. Analysis of phylogenomic
datasets reveals conflict, concordance, and gene duplications with examples from animals
and plants. *BMC Evolutionary Biology*, 15(1): 150.

Stamatakis, A. 2014. Raxml version 8: a tool for phylogenetic analysis and post-analysis of
large phylogenies. *Bioinformatics*, 30(9): 1312–1313.

Swofford, D. L. 2001. PAUP*: Phylogenetic Analysis Using Parsimony (* and other
methods) 4.0.b5.
Tamura, K., Battistuzzi, F. U., Billing-Ross, P., Murillo, O., Filipski, A., and Kumar, S. 2012. Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences*, 109(47): 19333–19338.

Walker, J. F., Yang, Y., Moore, M. J., Mikenas, J., Brockington, S. F., Timoneda, A., and Smith, S. A. 2017. Widespread paleopolyploidy, gene tree conflict and recalcitrant relationships among the carnivorous caryophyllales. *New Phytologist*. in review.

Wen, J., Xiong, Z., Nie, Z.-L., Mao, L., Zhu, Y., Kan, X.-Z., Ickert-Bond, S. M., Gerrath, J., Zimmer, E. A., and Fang, X.-D. 2013. Transcriptome sequences resolve deep relationships of the grape family. *PLOS ONE*, 8(9).

Worobey, M., Han, G.-Z., and Rambaut, A. 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature*, 508: 254–257.

Yang, Y., Moore, M. J., Brockington, S. F., Soltis, D. E., Wong, G. K.-S., Carpenter, E. J., Zhang, Y., Chen, L., Yan, Z., Xie, Y., Sage, R. F., Covshoff, S., Hibberd, J. M., Nelson, M. N., and Smith, S. A. 2015. Dissecting molecular evolution in the highly diverse plant clade caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution*.

Zhu, T., Dos Reis, M., and Yang, Z. 2015. Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. *Systematic Biology*, 64(2): 267–280.
Table 1: Dataset size and results of likelihood ratio tests for strict clock-like gene behavior.

| Dataset | Orthologs | Clocklike (%) |
|---------|-----------|---------------|
| BIR     | 7116      | 440 (6.18)    |
| CAR     | 3767      | 274 (7.27)    |
| CARY    | 583       | 3 (0.51)      |
| HYM     | 1161      | 22 (1.89)     |
| MIL     | 152       | 26 (17.10)    |
| VIT     | 2267      | 8 (0.35)      |

Table: Dataset size and results of likelihood ratio tests for strict clock-like gene behavior.
| Gene name | Variance   | Tree length | Bipartition proportion |
|-----------|------------|-------------|------------------------|
| **BIR**   |            |             |                        |
| 12969     | 0.000791644| 2.73068     | 0.5                    |
| 1173      | 0.00205589 | 3.01712     | 0.457                  |
| 12123     | 8.32228e-05| 0.825943    | 0.413                  |
| **CAR**   |            |             |                        |
| cluster259MIortho7 | 9.07832e-05 | 0.346618        | 1.0                    |
| cluster3790MIortho1 | 0.000245644 | 0.739886    | 1.0                    |
| cluster234MIortho1 | 0.0004849  | 1.19575     | 1.0                    |
| **CARY** |            |             |                        |
| cc7674-1-tolortho | 0.0183029  | 10.9821     | 0.701                  |
| cc4427-1MIortho1 | 0.0093838  | 8.7827      | 0.657                  |
| cc7873-1MIortho1 | 0.0206222  | 10.4773     | 0.657                  |
| **HYM**  |            |             |                        |
| cluster3024-1-1ortho1 | 0.00159156 | 2.64137     | 0.706                  |
| cluster5160-1-1ortho1 | 0.00294197 | 2.0815      | 0.706                  |
| cluster1251-1-1ortho1 | 0.00621115 | 4.99913    | 0.706                  |
| **MIL**  |            |             |                        |
| cluster89-1-lortho1 | 0.00200945 | 0.909593    | 0.875                  |
| cluster1437-1-lortho1 | 0.00872612 | 2.96511     | 0.875                  |
| cluster1615-1-lortho1 | 0.010942  | 3.56434     | 0.875                  |
| **VIT**  |            |             |                        |
| cluster9579-1MIortho1 | 0.000978163 | 0.519373   | 1.0                    |
| cluster1236-1MIortho1 | 0.00106778 | 0.547562    | 1.0                    |
| cluster461-1MIortho1 | 0.001227  | 0.607536    | 1.0                    |

Table 2: Properties of the genes used in the empirical dating analyses. Variance regards the root-to-tip paths. Tree length is measured in units of expected substitutions per site across all branches. Bipartition proportion measures agreement to the species tree topology (1.0 indicates complete concordance).

| Dataset | Height | Lower | Higher |
|---------|--------|-------|--------|
| **BIR** | -0.26  | 0.27  | -1.49  |
| **CAR** | -0.004 | 0.04  | -0.2   |
| **CARY**| -3.93  | 0.52  | -8.56  |
| **HYM** | -0.12  | 0.1   | -0.63  |
| **MIL** | -0.09  | 0.04  | -1.12  |
| **VIT** | -0.02  | 0.08  | -0.56  |

Table 3: The cumulative difference in the height, lower 95% HPD, and higher 95% HPD of each node comparing the UCLN estimates to the clock estimates from the individual empirical dating analyses. A value lower than 0 results when the cumulative difference in the clock values of height or HPD are younger than the associated UCLN values.
| Dataset | Height | Lower | Higher | Nodes | Error |
|---------|--------|-------|--------|-------|-------|
|         | CL UC  | CL UC | CL UC  | CL UC | CL UC |
| BIR     | 0.63 0.5 | 0.48 0.69 | 0.96 1.33 | 16 39 | 1.44 2.02 |
| CAR     | 0.26 0.2 | 0.37 0.85 | 0.25 0.63 | 5 12 | 0.62 1.49 |
| CARY    | 0.54 0.63 | 1.23 2.12 | 0.64 1.11 | 52 67 | 1.88 3.24 |
| HYM     | 0.16 0.76 | 0.21 0.65 | 0.38 0.94 | 15 3 | 0.58 1.59 |
| MIL     | 0.17 0.42 | 0.12 0.45 | 0.33 0.46 | 5 3 | 0.44 0.91 |
| VIT     | 0.27 0.26 | 0.42 0.32 | 0.2 0.27 | 5 8 | 0.62 0.59 |

Table 4: Assessment of dating error for the clock (CL) and UCLN (UC) analyses of the simulated clock data. All measures involve distance from the true node age, and are cumulative sums across all nodes. Height is the inferred node age. Lower and Higher regard the 95% HPD node age bounds. Nodes indicates the number of true node ages contained within the HPD interval. Error is the total error involved, equivalent to Low + High.

| Dataset | Height | Lower | Higher | Nodes | Error |
|---------|--------|-------|--------|-------|-------|
|         | CL UC  | CL UC | CL UC  | CL UC | CL UC |
| BIR     | 1.26 3.21 | 1.26 6.43 | 1.37 1.52 | 12 24 | 2.64 7.95 |
| CAR     | 0.76 0.69 | 0.89 1.59 | 0.68 0.58 | 2 9 | 1.57 2.17 |
| CARY    | 2.29 3.51 | 2.37 8.98 | 2.38 4.97 | 15 55 | 4.75 13.95 |
| HYM     | 0.14 0.91 | 0.61 3.01 | 0.61 1.58 | 18 16 | 1.22 4.65 |
| MIL     | 0.14 0.61 | 0.32 1.6 | 0.57 1.17 | 11 10 | 0.89 2.77 |
| VIT     | 0.29 1.12 | 0.82 2.43 | 0.29 0.73 | 14 9 | 1.11 3.16 |

Table 5: Assessment of dating error for the clock (CL) and UCLN (UC) analyses of the simulated ucln data. All measures involve distance from the true node age, and are cumulative sums across all nodes. Height is the inferred node age. Lower and Higher regard the 95% HPD node age bounds. Nodes indicates the number of true node ages contained within the HPD interval. Error is the total error involved, equivalent to Low + High.
Figure 1: Measures used for sorting genes for use in dating analyses. The order presented here is arbitrary.
Figure 2: Plots of gene tree properties (left, including root-to-tip variance and treelength for simulated and empirical datasets) and tip-specific root-to-tip variance for empirical datasets (right). When the outgroup is present, the taxa are labeled with a red dot.
Figure 3: Plots of gene tree properties (left, including root-to-tip variance and treelength for simulated and empirical datasets) and tip-specific root-to-tip variance for empirical datasets (right). When the outgroup is present, the taxa are labeled with a red dot.
Figure 4: A comparison of strict clock and UCLN estimates of node ages for the six curated empirical datasets.
Figure 5: A comparison of strict clock and UCLN estimates of node ages for the simulated clock and ucln datasets. Red and pink are scenarios where the generating and inference are identical, while green and blue are where the models are mismatched.