Supplementary Information for:

Estimation of species divergence times in presence of cross-species gene flow
George P. Tiley†, Tomás Flouri², Xiyun Jiao²,³, Jelmer W. Poelstra¹, Bo Xu⁴, Tianqi Zhu⁵,⁶, Bruce Rannala⁷, Anne D. Yoder¹*, Ziheng Yang²*
¹Department of Biology, Duke University, Durham, NC, USA
²Department of Genetics, Evolution and Environment, University College London, London, UK
³Department of Statistics and Data Science, China Southern University of Science and Technology, Shenzhen, Guangdong, China
⁴Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
⁵National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
⁶Key Laboratory of Random Complex Structures and Data Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
⁷Department of Evolution and Ecology, University of California, Davis, Davis, CA, USA
†Present address: Royal Botanic Gardens Kew, Richmond, Surrey, UK
Supplementary Methods

Parameter estimates from simulated data
Parameters were estimated from simulated data with BPP v4.1.4 for both the MSC and MSci models. We used a burnin of 20,000 MCMC iterations and collected 200,000 samples sampling every 2 iterations. The prior on $\theta$ was an inverse gamma with shape parameter $\alpha = 3$ and scale parameter $\beta = 2 \theta$. This is a diffuse prior with the prior mean $\frac{\beta}{(\alpha-1)}$ at the true value.

Similarly, the prior on the root age $\tau_r$ was an inverse gamma with $\alpha = 3$ and $\beta = 6 \theta$ so that the prior mean is $3 \theta$. The other node ages follow a uniform Dirichlet distribution (Yang and Rannala 2010, Eq. 2). For the MSci model only, a prior Beta distribution with both shape parameters set at 1 was specified on $\phi$, which is equivalent to the uniform distribution for the interval from 0 to 1. Visual inspection of a few analyses from the simulated data were done to assure MCMC settings were appropriate, and that parameter estimates had converged.

Additional scrutiny was given to seemingly aberrant runs. Inspection of MCMC samples and posterior summaries revealed that some replicates converged to the true value with efficient MCMC sampling while some tended to samples with biologically implausibly large $\theta$ and wide HPD intervals despite the posterior median being very close to the true value. Were we to collect many more MCMC samples for these simulation analyses with poorly estimated $\theta$ from the MSC model, it is reasonable to expect improved estimates. However, we present these estimates as poor MCMC sampling occurs in empirical data analysis.

Calibrating divergence times
We considered two common approaches to calibrations that convert divergence times from substitutions per site to absolute time in years. First, we used a node calibration on the root, which could come from fossil evidence placed near the crown of a group (e.g. Benton and Donoghue 2007) or a well-justified geological event (e.g., Da Baets et al. 2016). Second, we used a substitution rate calibration. A de novo mutation rate could be used to calibrate divergences (Yoder and Tiley 2021), or a phylogenetic substitution rate from a previous study could be used as well (Campbell et al. 2021). It is up to the investigator to determine the best calibration, but we make the comparison here since a rate calibration will depend on the absolute node differences between models (as presented in our simulation results), while a node calibration depends only on the relative distances between nodes. The differences between calibration strategies could lead to different conclusions about absolute ages between the MSC and MSci models. For example a root node with an age of 0.01 substitutions per site under the MSC and 0.02 substitutions per site under the MSci would both have an age of 10 Ma ago if the same node calibration were applied. However, a rate calibration applied to all nodes of a tree would cause the MSci root divergence time to be twice as old as the MSC divergence time.

Here we apply simple node and rate calibrations, to demonstrate differences between the MSci and MSC model with real data rather than attempting to obtain the best possible absolute divergence times for the two systems. Since the rate calibration presented was based on substitutions per-site from the MSC analyses and we adhere to a strict molecular clock, the node-calibrated and rate-calibrated MSC analyses are equivalent.

Baobabs
We assumed that the sampled outgroup Scleronema diverged from Adansonia 18.2 Ma ago (Marinho et al. 2014). Divergence times were estimated under MSC and MSci using this node calibration by setting the mean posterior estimate of $\tau_c$ (Supplementary Figure S25) to 18.2 and rescaling all other mean estimates and HPD intervals by $\frac{\text{node age}}{\tau_c} \times 18.2$. A rate calibration in substitutions per site per million years was obtained from the MSC model by $\text{rate} = \frac{\tau_c}{18.2} = \ldots$
The rate-calibrated MSci estimates were then generated by $\frac{\text{node age}}{\text{rate}}$. All MSC and MSci estimates in substitutions per site are available in Supplementary Table S1 and calculating the calibrated values is trivial given the assumption of a strict clock.

**Jaltomata**

The procedure for deriving node-calibrated and rate-calibrated divergence times was the same as for baobabs. As node calibration, we assumed that *Solanum* and *Jaltomata* diverged 17 Ma ago ($t_\alpha$ in Supplementary Figure S32) (Sarkinen et al. 2013). All node ages under the MSC and MSci models in substitutions per site are in Supplementary Table S2.

**References**


Supplementary Figures

Figure S1 – Divergence Time Estimates under the MSC with Introgression Between Sister Lineages. Points are posterior means and bars are 95% HPD CIs. Dashed lines are true values. Node labels and divergence times correspond to Fig. 1a.
Figure S2 – Divergence Time and Introgression Probability Estimates Under the MSci with Introgression Between Sister Lineages. Points are posterior means and error bars are the 95% HPD CIs. Dashed lines are true values. Node labels and divergence times correspond to Fig. 1a.
Figure S3 – Divergence Time Estimates under the MSC with Introgression Between Non-Sister Lineages. Node labels and divergence times correspond to Fig. 1b.
Figure S4 – Divergence Time and Introgression Probability Estimates under the MSci with Introgression between Non-Sister Lineages. Node labels and divergence times correspond to Fig. 1b.
Figure S5 – Divergence Time Estimates under MSC with Introgression from a Ghost Lineage into an Ingroup Species. Node labels and divergence times correspond to Fig. 1c.
Figure S6 – Divergence Time and Introgression Probability Estimates under the MSci with Introgression from a Ghost Lineage into an Ingroup Species. Node labels and divergence times correspond to Fig. 1c.
Figure S7 – Population Size Estimates under the MSC with Introgression between Sister Lineages. Node labels and populations sizes correspond to Fig. 1a.
Figure S8 – Population Size Estimates under the MSci with Introgression between Sister Lineages. Node labels and population sizes correspond to Fig. 1a.
Figure S9 – Population Size Estimates under the MSC with Introggression between Non-Sister Lineages. Node labels and population sizes correspond to Fig. 1b.
Figure S10 – Population Size Estimates under the MSci with Introgression between Non-Sister Lineages. P Node labels and population sizes correspond to Fig. 1b.
Figure S11 – Population Size Estimates under the MSC with Introgression from a Ghost Lineage to an Ingroup Species. Node labels and population sizes correspond to Fig. 1c.
Figure S12 – Population Size Estimates under the MSci with Introgression from a Ghost Lineage to an Ingroup Species. Node labels and population sizes correspond to Fig. 1c.
Figure S13 – Divergence Time Estimates under the MSC when there is migration between Sister Lineages. Node labels and divergence times correspond to Fig. 1a. $Nm$ is the population immigration rate.
Figure S14 – Divergence Time and Introgression Probability Estimates under the MSci when there is migration between Sister Lineages. Node labels and divergence times correspond to Fig. 1a. $Nm$ is the population immigration rate, so there is no true value for $\varphi_h$. 
Figure S15 – Divergence Time Estimates under the MSC when there is migration between Non-Sister Lineages. Node labels and divergence times correspond to Fig. 1b.
Figure S16 – Divergence Time and Introgression Probability Estimates under the MSci when there is migration between Non-Sister Lineages. Node labels and divergence times correspond to Fig. 1b.
Figure S17 – Divergence Time Estimates under the MSC When there is Migration from a Ghost Lineage to an Ingroup Species. Node labels and divergence times correspond to Fig. 1c.
Figure S18 – Divergence Time and Introgression Probability Estimates under the MSci When there is Migration from a Ghost Lineage to an Ingroup Species. Node labels and divergence times correspond to Fig. 1c.
Figure S19 – Population Size Estimates under the MSC When There is Migration Between Sister Lineages. Node labels and populations sizes correspond to Fig. 1a.
Figure S20 – Population Size Estimates under the MSci When There is Migration Between Sister Lineages. Node labels and population sizes correspond to Fig. 1a.
Figure S21 – Population Size Estimates under the MSC When There is Migration Between Non-Sister Lineages. Node labels and population sizes correspond to Fig. 1b.
Figure S22 – Population Size Estimates under the MSci When There is Migration Between Non-Sister Lineages. Node labels and population sizes correspond to Fig. 1b.
Figure S23 – Population Size Estimates under the MSC When There is Migration from a Ghost Lineages to an Ingroup Species. Node labels and population sizes correspond to Fig. 1c.
Figure S24 – Population Size Estimates under the MSci When There is Migration from a Ghost Lineages to an Ingroup Species. Node labels and population sizes correspond to Fig. 1c.
Figure S25 – Species Tree with Node Labels for *Adansonia* Assumed in BPP Analyses.

Divergence time estimation used the inferred introgression events from PhyloNetworks. Node labels are used to identify parameters. Blue edges represent introgression events assumed in the MSci model.
Figure S26 – Population Size Estimates for *Adansonia* under the MSC. The four violin plots represent the four individual chains used for analysis.
Figure S27 – Divergence Time Estimates for *Adansonia* under the MSC. The four violin plots represent the four individual chains used for analysis. A scatterplot of median node heights for $\tau$ implies convergence of the MCMC analyses.
Figure S28 – Population Size Estimates for *Adansonia* under the MSci. The four violin plots represent the four individual chains used for analysis. Additional nodes for introgression events are represented here.
Figure S29 – Posterior Estimates of Divergence Times and Introgression Probability under the MSci for *Adansonia*. The four violin plots represent four replicate runs. The scatterplot of median node ages suggests consistency between runs.
Figure S30 – ASTRAL tree for Jaltomata. Numbers at nodes are local posterior probabilities. Coloring of tip labels and outlined clades indicate fruit color and follow the convention of the original study. The ASTRAL tree matches the original authors’ concatenated ML tree with the exception of the switched placement of *J. dendroidea* and *J. incahuasina*. This relationship has a low local posterior probability. Internal branch lengths are coalescent units \((\Delta t/\theta)\), while terminal branch lengths are not estimated by ASTRAL and arbitrarily set to 1. Note that ASTRAL estimates unrooted trees and the species tree is rooted with *S. lycopersicum*, so the branch subtending the *Jaltomata* MRCA is set arbitrarily to 0.5.
Figure S31 – Estimated Phylogenetic Network from PhyloNetworks for *Jaltomata*. The major species tree is identical to the ASTRAL tree. The admixture proportions are shown next to the parental branches at each hybridization node, with the black branches (corresponding to the greater admixture probability) representing the major species tree. The inferred introgression events (but not the introgression probabilities) were assumed in the MSci model to estimate divergence times with BPP.
Figure S32 – Species tree with Node Labels Assumed in BPP Analyses. Blue edges represent introgression events inferred using PhyloNetworks and assumed in the BPP analysis under the MSci model to estimate divergence times. Node labels are used to identify parameters. A ghost lineage was included to model introgression from an unsampled lineage sister to *Jaltomata* into the common ancestor of the green- and orange-fruited clades.
Figure S33 – Posterior Estimates of Population Sizes under the MSci in four Replicate runs for *Jaltomata*.
Figure S34 – Posterior Estimates of Divergence Times under the MSC in Four Replicate Runs for Jaltomata. The scatterplot of median node ages shows consistency between runs. The root node is excluded from the scatterplot to improve visibility.
Figure S35 – Posterior Estimates of Population Sizes for *Jaltomata* under the MSci model in Four Replicate Runs.
Figure S36 – Posterior Estimates of Divergence Times for *Jaltomata* under the MSci in Four Replicate Runs. The scatterplot of median node ages (τ) indicates convergence. The root node is excluded from the scatterplot to improve visibility.
Figure S37 – Posterior Estimates of Introgression Probabilities for *Jaltomata* under the MSci model in Four Replicate Runs.
Table S1 – Divergence Time Estimates for Adansonia. Node labels correspond to nodes in Figure S25. Node numbers correspond to parameter order in BPP log files available through Dryad.
<table>
<thead>
<tr>
<th>Node Label</th>
<th>Node Number</th>
<th>MSC Node</th>
<th>Mean</th>
<th>Lower 95% HPD</th>
<th>Upper 95% HPD</th>
<th>MSci Node</th>
<th>Mean</th>
<th>Lower 95% HPD</th>
<th>Upper 95% HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>16</td>
<td>0.027963</td>
<td>0.027153</td>
<td>0.028723</td>
<td></td>
<td>16</td>
<td>0.028</td>
<td>0.027214</td>
<td>0.028789</td>
</tr>
<tr>
<td>t</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>17</td>
<td>0.004586</td>
<td>0.004341</td>
<td>0.004866</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>18</td>
<td>0.001977</td>
<td>0.001898</td>
<td>0.002057</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>17</td>
<td>0.004125</td>
<td>0.004032</td>
<td>0.00422</td>
<td>19</td>
<td>0.004259</td>
<td>0.004144</td>
<td>0.004369</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>18</td>
<td>0.003147</td>
<td>0.003063</td>
<td>0.003234</td>
<td>20</td>
<td>0.003315</td>
<td>0.003217</td>
<td>0.003412</td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>21</td>
<td>0.002375</td>
<td>0.002156</td>
<td>0.002561</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>19</td>
<td>0.001995</td>
<td>0.001636</td>
<td>0.002332</td>
<td>22</td>
<td>0.002429</td>
<td>0.002233</td>
<td>0.00261</td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>23</td>
<td>0.002375</td>
<td>0.002156</td>
<td>0.002561</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>20</td>
<td>0.002361</td>
<td>0.002297</td>
<td>0.002423</td>
<td>24</td>
<td>0.002386</td>
<td>0.002321</td>
<td>0.002449</td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>25</td>
<td>0.001977</td>
<td>0.001898</td>
<td>0.002057</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>21</td>
<td>0.001817</td>
<td>0.001765</td>
<td>0.00187</td>
<td>26</td>
<td>0.001962</td>
<td>0.001885</td>
<td>0.002041</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>27</td>
<td>0.00095</td>
<td>0.000853</td>
<td>0.001032</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>22</td>
<td>0.000821</td>
<td>0.000721</td>
<td>0.000923</td>
<td>28</td>
<td>0.000517</td>
<td>0.000141</td>
<td>0.000828</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>23</td>
<td>0.001346</td>
<td>0.001276</td>
<td>0.001414</td>
<td>29</td>
<td>0.001374</td>
<td>0.001311</td>
<td>0.001442</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>24</td>
<td>0.001115</td>
<td>0.001079</td>
<td>0.001213</td>
<td>30</td>
<td>0.001165</td>
<td>0.001103</td>
<td>0.001122</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>25</td>
<td>0.000531</td>
<td>0.000438</td>
<td>0.000618</td>
<td>31</td>
<td>0.000521</td>
<td>0.000431</td>
<td>0.000608</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>26</td>
<td>0.000392</td>
<td>0.000119</td>
<td>0.000575</td>
<td>32</td>
<td>0.000387</td>
<td>0.00012</td>
<td>0.000566</td>
<td></td>
</tr>
<tr>
<td>l</td>
<td>27</td>
<td>0.001066</td>
<td>0.000981</td>
<td>0.001114</td>
<td>33</td>
<td>0.001103</td>
<td>0.001041</td>
<td>0.001164</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>28</td>
<td>0.000994</td>
<td>0.000879</td>
<td>0.001091</td>
<td>34</td>
<td>0.001057</td>
<td>0.000984</td>
<td>0.001123</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>29</td>
<td>0.000795</td>
<td>0.000563</td>
<td>0.000983</td>
<td>35</td>
<td>0.000976</td>
<td>0.000893</td>
<td>0.001053</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>36</td>
<td>0.00095</td>
<td>0.000853</td>
<td>0.001032</td>
<td></td>
</tr>
</tbody>
</table>

Table S2 – Divergence Time Estimates for *Jaltomata*. Node labels correspond to nodes in Figure S32. Node numbers correspond to parameter order in BPP log files available through Dryad.