

---

# EMPIRICAL ANALYSIS OF PHYLOGENETIC QUASI-TERRACES

---

A PREPRINT

**Paula Breitling**

Institute for Theoretical Informatics  
Karlsruhe Institute of Technology  
76128 Karlsruhe, Germany,  
paula.breitling@kit.edu

**Alexandros Stamatakis \***

Institute for Theoretical Informatics  
Karlsruhe Institute of Technology  
76128 Karlsruhe, Germany,  
and  
Computational Molecular Evolution Group  
Heidelberg Institute for Theoretical Studies  
69118 Heidelberg, Germany  
Alexandros.Stamatakis@h-its.org  
*\* corresponding author*

**Olga Chernomor**

Center for Integrative Bioinformatics Vienna  
Max Perutz Labs  
University of Vienna  
Medical University of Vienna  
Dr. Bohr Gasse 9  
1030 Vienna, Austria  
olga.chernomor@univie.ac.at

**Ben Bettisworth**

Computational Molecular Evolution Group  
Heidelberg Institute for Theoretical Studies  
69118 Heidelberg, Germany  
Ben.Bettisworth@h-its.org

**Lukasz Reszczyński**

Center for Integrative Bioinformatics Vienna  
Max Perutz Labs  
University of Vienna  
Medical University of Vienna  
Dr. Bohr Gasse 9  
1030 Vienna, Austria  
lukasz.reszczyński@univie.ac.at

October 18, 2019

## ABSTRACT

Terraces in phylogenetic tree space are, among other things, important for the design of tree space search strategies. While the phenomenon of phylogenetic terraces is already known for unlinked partition models on partitioned phylogenomic data sets, it has not yet been studied if an analogous structure is present under linked and scaled partition models. To this end, we analyze aspects such as the log-likelihood distributions, likelihood-based significance tests, and nearest neighborhood interchanges on the trees residing on a terrace and compare their distributions among unlinked, linked, and scaled partition models. Our study shows that there exists a terrace-like structure under linked and scaled partition models as well. We denote this phenomenon as quasi-terrace. Therefore quasi-terraces should be taken into account in the design of tree search algorithms as well as when reporting results on 'the' final tree topology in empirical phylogenetic studies.

**Keywords** Phylogenomics · Phylogenetic terrace · Maximum likelihood · Partition models · Quasi-terraces

## 1 Introduction

Phylogenetic trees are widely used to explain and visualize the evolutionary relationships of species. One approach to reconstructing phylogenies consists in using a supermatrix, which is a collection of multiple partitions, typically representing genes that are concatenated into one multiple sequence alignment. The alignment (matrix) size is determined by the number of homologous sites (columns) and the number of taxa (rows), often also denoted as species or sequences. Hence, a partition is a subset of alignment sites. A common problem with these large phylogenomic alignments are missing data. That is, a taxon can have no data present in a specific partition. This can either be due to errors in the sequencing process or because some species simply do not have data in this partition (do not have the specific gene) or if the gene has not been sampled yet (Kearney, 2002; Nyakatura & Bininda-Emonds, 2012; Wiens & Morrill, 2011). This type of missing data complicates phylogenetic tree inference under likelihood-based criteria (maximum likelihood (ML) or Bayesian inference) and under parsimony.

Two recent studies (Chernomor, von Haeseler, & Minh, 2016; Dobrin, Zwickl, & Sanderson, 2018) have analyzed a wide range of empirical data sets highlighting the importance and prevalence of terraces in tree in large contemporary phylogenomic studies. Dobrin et al. (2018) have primarily focused on the presence of terraces for published alignments and discussed their properties. Among others, for each data set, they provide the theoretical minimum number of genes that would be necessary to reduce the terrace sizes to just one tree. Moreover, the authors also discussed implications of terraces on bootstrap values.

The second study (Chernomor et al., 2016) introduced a production level implementation of techniques to account for terraces during tree searches in ML inference under the most general partition model, that is, the unlinked branch (UB) partition model (Z. Yang, 1996). The authors also introduced techniques for saving computational time in the presence of missing sequences under more restrictive linked (LB) and scaled branch (SB) partition models (Z. Yang, 1996). They obtained substantial speedups for all data sets tested under all three partition models. Awareness about quasi-terraces could help to further improve these results under the LB and SB models. During the search one can simply use one representative tree from a quasi-terrace to calculate its log-likelihood (LnL) score and treat it as an approximation of the log-likelihood scores for all other trees on the quasi-terrace. This will increase the speedup under LB and SB models by the same fraction as currently achieved under the UB partition model using terrace-aware tree inference procedures (in some cases up to 4.5 times faster, than in terrace-unaware computations).

## 2 Definitions

First we introduce some formal definitions and concepts. Let  $T$  and  $Y$  denote a tree and a partition, respectively. The induced partition tree, denoted by  $T|Y$ , is obtained from  $T$  by removing species, which have no data present in partition  $Y$  (Sanderson, McMahon, & Steel, 2011). A set of all trees that have an identical set of induced partition trees for all  $k$  partitions is denoted by  $P = \{T_1, T_2, \dots, T_N\}$  (also called a *stand* in Sanderson, McMahon, Stamatakis, Zwickl, and Steel (2015)). This set is fully determined by the presence/absence matrix (presence/absence of genes per species/partition) and one representative tree  $T$ , which can be any of  $T_1, \dots, T_N$ .

For a given alignment (supermatrix) the tree space is partitioned into multiple disjoint  $P$  sets with (often) different sizes. The size (the number of trees in  $P$ ) can vary between 1 tree (a trivial set) and more than a billion of trees (Sanderson et al., 2011). Note that, for a given set  $P$ , currently available algorithms can compute its size and enumerate all trees, only if the alignment contains a so-called comprehensive taxon (Biczok et al., 2018). That is, there is at least one taxon that has data for all partitions.

The three partition models (LB, SB, and UB; (Z. Yang, 1996)) differ in the way they optimize branch lengths. Under UB, a separate independent set of branch lengths is estimated for each partition in the phylogenomic data set. Under LB, a single branch length over all partitions is being estimated. Under SB, the same underlying branch lengths are scaled via a single parameter for each partition. Under all models, the LnL of a tree is computed as the sum of the LnLs over all induced partition trees. When the branch lengths are estimated independently for each induced partition tree, their LnLs, and as a result, also their sum are identical for all trees in  $P$ . That is, under the UB partition model, all trees in  $P$  have an identical LnL (Sanderson et al., 2015, 2011). In this case  $P$  is called a terrace. From theory we know, that under LB and SB, trees in  $P$  have distinct LnL (Sanderson et al., 2015). However, we expect that the LnL scores will not be significantly different. We call the set  $P$  under LB and SB a quasi-terrace.

As many phylogenetic studies of large empirical alignments use LB and SB partition models (Duchêne et al., 2018; Rosa, Melo, & Barbeitos, 2019; Wang, Susko, & Roger, 2019), the phenomenon of quasi-terraces is highly relevant for contemporary analyses. LB and SB partition models induce a substantially smaller number of free model parameters compared to UB. Consequently, they are less computationally expensive. Moreover, for the same reason, LB and SB are less prone to overfitting. For instance, it appears more reasonable to deploy the SB model to account for different

evolutionary rates for partitions derived from the 1st, 2nd, and 3rd codon position, rather than using a generic UB partition model.

In this paper we conduct a study of 7 published empirical phylogenomic data sets to assess if quasi-terraces exist.

### 3 Data sets

Two recently published papers (Chernomor et al., 2016; Dobrin et al., 2018) addressed different aspects of terraces in tree space and collected a wide range of empirical alignments with missing data.

Dobrin et al. (2018) investigated how frequently terraces occur in empirical phylogenomic studies and analyzed data sets from 12 different published studies (Burleigh, Kimball, & Braun, 2015; Meredith et al., 2011; Miadlikowska et al., 2014; Misof et al., 2014; Rabosky, Donnellan, Grudler, & Lovette, 2014; Shi & Rabosky, 2015; Soltis et al., 2013; Springer et al., 2012; Tolley, Townsend, & Vences, 2013; Wickett et al., 2014; Y. Yang et al., 2015; Zanne et al., 2014). In total, they analyzed 23 DNA and 3 protein alignments. The terrace sizes among the data sets varied between a single tree and  $1.30 \times 10^{388}$  trees.

Chernomor et al. (2016) analyzed 12 empirical data sets (9 DNA and 3 protein alignments) published in (Bouchenak-Khelladi et al., 2008; Dell’Ampio et al., 2013; Fabre, Rodrigues, & Douzery, 2009; Hinchliff & Roalson, 2012; Nyakatura & Bininda-Emonds, 2012; Pyron et al., 2011; Springer et al., 2012; Stamatakis & Alachiotis, 2010; Van Der Linde, Houle, Spicer, & Steppan, 2010). The terraces corresponding to ML trees for the data sets from Chernomor et al. (2016) are either trivial (comprising just one tree) or extremely large (in some cases, exceeding 850,000 trees).

In our study of quasi-terraces we use 7 phylogenomic DNA data sets from (Dobrin et al., 2018). Table 1 provides an overview of the analyzed data sets. The number of taxa varies from 106 to 213 and the number of partitions varies from 4 to 6. Other data sets from (Chernomor et al., 2016; Dobrin et al., 2018) had to be excluded, either because the terraces of interest were trivial (comprising just one tree), or excessively large ( $> 588,735$  trees) to be analyzed within the 24 hour job run time limit on the cluster.

Table 1: Overview of data sets

Data set	Datatype	#Taxa	#Partitions	#Sites	Reference
Eucalyptus*	DNA	136	4	6,205	(Zanne et al., 2014)
Euphorbia*	DNA	131	6	9,154	(Zanne et al., 2014)
Primula*	DNA	185	5	7,321	(Zanne et al., 2014)
Rabosky.scincids	DNA	213	6	5,373	(Rabosky et al., 2014)
Ranunculus*	DNA	170	6	10,799	(Zanne et al., 2014)
Rhododendron*	DNA	117	5	7,321	(Zanne et al., 2014)
Szygium*	DNA	106	4	5,815	(Zanne et al., 2014)

\*: *Subsampled data sets. Data sets were subsampled until they contained a comprehensive taxon. See Section 4.1 for details*

### 4 Experimental setup

For our quantitative analysis, we first designed a data preparation and analysis pipeline, so that every data set is prepared and analyzed in exactly the same way to obtain comparable results. We first outline the basic steps of our pipeline and subsequently discuss each of them in more detail.

For each data set we conducted the following steps (see also Table 2):

1. Data preprocessing: Create a PHYLIP formatted alignment, a binary presence/absence matrix (indicating which species has data for which partition), and a partition file (defining which columns/sites of the multiple sequence alignment belong to which partition) from the original NEXUS input files.
2. Conduct ML tree searches under UB, LB, and SB (inferred ML trees are denoted as  $T_{UB}$ ,  $T_{LB}$ , and  $T_{SB}$ , respectively).
3. For each ML tree from the previous step determine if the corresponding  $P$  set has size greater than 1 and enumerate the trees in  $P$  (for notations see Table 2).

4. Calculate the LnL scores of all trees in the respective  $P$  set under the corresponding partition model ( $P_1$  under UB,  $P_2$  under LB,  $P_3$  under SB).
5. Additionally, for  $P_1$  calculate the LnL score under the LB and the SB partition models.
6. Further analyses:
  - (a) Calculate RF-distance (Robinson & Foulds, 1981) between the best ML trees:  $T_{LB}$ ,  $T_{SB}$ , and  $T_{UB}$ .
  - (b) Compute significance tests based on the LnL scores for the trees in  $P_2$  and  $P_3$  under LB and SB.

The additional analysis of the tree neighborhood of the Rhododendron data set comprised the following steps:

1. For  $P_1$  generate its tree neighborhood by nearest-neighbor interchange (NNI<sup>1</sup>): for each tree from  $P_1$  collect all trees one NNI away from it.
2. Compute LnL for all trees in the NNI-neighborhood of  $P_1$  under the UB, LB, and SB partition models.
3. Compute significance tests of the trees in  $P_1$  and their NNI neighborhood for LnL scores under UB, LB, and SB.
4. The same NNI neighborhood analysis (generation of one NNI neighboring trees, LnL calculation, and significance tests) was also performed for the set  $P_4$  generated by a random Yule–Harding (YH) tree (Harding, 1971)).

Table 2: Analysis steps and notations

Analyses performed for each data set					
Step	Software used	Corresponding notations			
		Object	UB	LB	SB
Tree inference and RF distance computation	RAxML-NG	ML tree	$T_{UB}$	$T_{LB}$	$T_{SB}$
Size and enumeration of trees in corresponding $P$ sets	Terraphast I	$P$ set	$P_1$	$P_2$	$P_3$
LnL calculation for all trees in $P$ under corresponding partition model	RAxML-NG		✓	✓	✓
LnL calculation for all trees in $P$ under all three partition models	RAxML-NG		✓		
Significance tests of LnL scores of trees in $P$	IQ-Tree			✓	✓
Additional analysis for Rhododendron data set					
The tree used to generate a corresponding set $P$			$T_{UB}$	$T_{YH}$ *	
Notation for set $P$			$P_1$	$P_4$	
Calculation of LnL for all trees in the NNI-neighborhood under UB, LB, and SB			✓	✓	
Significance tests of LnL scores of trees in $P$ and its NNI-neighborhood			✓	✓	

\*: *Random Yule-Harding tree*

## 4.1 Data preprocessing

As we had a collection of NEXUS data files, but required other formats for our analyses, we initially transformed our data accordingly. To analyze (quasi-)terraces we required a binary presence/absence matrix, which describes for which species we have data in which partitions.

An important aspect for the downstream analysis of (quasi-)terraces is that all data sets need to comprise a comprehensive taxon. However, not all of our data sets comprised such a comprehensive taxon *a priori*. Therefore, we reduced (subsampling) some data sets by removing partitions, such that the reduced data set comprised at least one comprehensive taxon. We conduct this reduction as follows: for each taxon we count in how many partitions it has data and chose the taxon that has the largest number of partitions *with* data. Once such a taxon has been selected, we remove all partitions from the data set for which this taxon does not have data. When such a reduction is applied, we need to propagate it to all subsequent files used within the analysis pipeline.

## 4.2 Tree inference

We inferred the best-known ML trees under the UB, LB, and SB models with RAxML-NG (Kozlov, Darriba, Flouri, Morel, & Stamatakis, 2019) (denoted as  $T_{UB}$ ,  $T_{LB}$  and  $T_{SB}$ ). For each data set and each partition model, we performed 20 independent tree searches, and recorded the tree with the best LnL score among these 20 independent runs. We used 10 random trees and 10 randomized stepwise addition order parsimony trees as starting trees.

<sup>1</sup>Nearest-neighbor interchange is the simplest topological rearrangement operation, when one tree is obtained from another by changing subtrees across an internal branch of the tree.

### 4.3 Enumeration of trees in $P$ sets

We used Terraphast I (Biczok et al., 2018) to enumerate all trees and computed the sizes of  $P_1$ ,  $P_2$ , and  $P_3$  (collection of trees with identical induced partition trees), generated by the corresponding ML trees  $T_{UB}$ ,  $T_{LB}$ , and  $T_{SB}$ . Recall that, under UB, any set  $P$  is a terrace, while under LB and SB, any set  $P$  is a quasi-terrace. Apart from the ML tree file, Terraphast I requires the aforementioned binary presence/absence matrix as input. Terraphast I then outputs the number of trees in  $P$  (terrace/quasi-terrace) and enumerates all trees in  $P$  in Newick format.

### 4.4 Calculation of log-likelihood scores

We used RAxML-NG to compute the LnL scores of all tree topologies in the respective  $P$  sets. The input is a PHYLIP alignment file, the corresponding partition file, and the terrace tree file (trees in the corresponding  $P$  set). We calculated the LnLs of the trees in  $P_1$ ,  $P_2$ , and  $P_3$  under the UB, LB, and SB partition models. For the sake of comparison, we also calculated the LnLs of the trees in  $P_1$  (generated by using  $T_{UB}$ ) under the LB and SB models. Recall that, any set  $P$  is defined by the topology of the representative tree. Therefore, the terrace  $P_1$  generated using  $T_{UB}$  is also a quasi-terrace under LB and SB.

### 4.5 Further analyses

To assess the similarity of the best ML trees inferred under each partition model and if partitioning 'does matter' (i.e., yields a distinct tree topology), we computed the Robinson-Foulds (RF) distances (Robinson & Foulds, 1981) between  $T_{UB}$ ,  $T_{LB}$ , and  $T_{SB}$ .

To confirm our expectation that the trees on a quasi-terrace are indeed not significantly different from each other with respect to the LnL score, we conducted standard significance tests with IQ-Tree (Chernomor et al., 2016; Nguyen, Schmidt, von Haeseler, & Minh, 2014) for the trees in  $P_2$  and  $P_3$  and the LnL scores computed under the LB and SB partition models, respectively. As input we used the partition file, the trees in the respective  $P$  set, and the PHYLIP alignment file.

IQ-Tree implements the following common LnL-based significance tests:

- Bootstrap proportions using the REL method (Kishino, Miyata, & Hasegawa, 1990)
- Kishino-Hasegawa test (one sided and weighted) (Kishino & Hasegawa, 1989)
- Shimodaira-Hasegawa test (weighted and unweighted) (Shimodaira & Hasegawa, 1999)
- Expected likelihood weight (Strimmer & Rambaut, 2002)
- Approximately unbiased (AU) test by Shimodaira (Shimodaira, 2002)

To explore the trees surrounding a (quasi-)terrace, on one of our data sets (Rhododendron) we also performed a NNI neighborhood analysis. For each tree in  $P_1$  we applied all possible NNIs to obtain a collection of all one-NNI neighboring trees surrounding the (quasi-)terrace. We then compared the trees in  $P_1$  with the generated NNI tree set. We performed this comparison by evaluating LnL scores under all three partition models.

We identified the range between the maximum and minimum LnL of the trees in  $P_1$  (i.e., trees on the (quasi-)terrace). Then, we computed the fraction of NNI trees falling within this range and the fraction of NNI trees with a LnL higher than the maximum LnL of the trees in  $P_1$ . We also conducted LnL score significance tests on an extended tree set, containing, the trees in  $P_1$  as well as the NNI trees.

With this analysis, we intended to assess how trees on a quasi-terrace differ from those in its NNI neighborhood. In other words, we test if all trees in  $P$  behave similarly, that is, if the majority of the trees on a quasi-terrace are significantly better or significantly worse than the surrounding trees. If all trees in  $P$  do behave similarly, it suffices to only use one representative tree from the quasi-terrace to save time and not evaluate other trees in  $P$ , that essentially behave similarly to that representative.

Since  $P_1$  was obtained using the best-known ML tree, the difference in LnLs between trees on the quasi-terrace and their NNI-neighbors might be small, as the tree search has converged on a tree with a 'good' ML score. Therefore, to assess, if the LnL differences are more pronounced for a quasi-terrace that was encountered during the early stages of the tree search, we also performed an analysis of the NNI neighborhood for a set  $P_4$  generated by a random Yule-Harding tree.

## 5 Results

### 5.1 Comparison of LnL scores for the trees from $P_1$ under UB, LB and SB models

Table 3 summarizes the results for the corresponding  $P_1$  sets for all tested alignments. The size of  $P_1$  varies from 3 to 1,863 trees. The average LnL (Avg. LnL) of the trees in  $P_1$  computed under UB is better (higher) than under SB and in turn under LB for all data sets. This is expected as the number of free model parameters decreases. The standard deviation (Std. Dev) of LnL scores is near zero for all data sets under UB and also for three data sets (Ranunculus, Rhododendron, and Szygium) under LB and SB. For Eucalyptus and Euphorbia, the Std. Dev for LB and SB is considerably larger (between 2.14 and 4.58). For Primula and Rabosky.scincids, it ranges between 1.21 and 1.78. The relative RF distance (Rel. RF) between the best ML trees  $T_{UB}$ ,  $T_{LB}$  and  $T_{SB}$  is smaller than 0.25 for 6 out of 7 data sets. For Eucalyptus we observe the highest value of 0.5.

Table 3: Overview of results for  $P_1$  sets

Data set	Size of $P_1$	Avg. LnL			Std. Dev			Rel. RF
		UB	LB	SB	UB	LB	SB	
Euphorbia	1,863	-39,580.3	-41,201.1	-40,173.6	0.0009	4.5787	3.7331	0.15
Primula	1,125	-37,034.1	-38,491.0	-38,084.1	0.0006	1.2102	1.2684	0.15
Eucalyptus	267	-10,902.7	-11,263.9	-11,098.0	0.0120	3.2669	2.1400	0.50
Rhododendron	27	-17,830.8	-18,216.8	-18,181.8	0.0005	0.1619	0.1644	0.25
Ranunculus	9	-28,709.8	-29,674.1	-29,447.7	0.0009	0.0922	0.1452	0.18
Szygium	5	-11,931.6	-12,214.3	-12,125.0	0.0003	0.1604	0.0155	0.18
Rabosky.scincids	3	-125,550.6	-128,649.7	-127,973.0	0.0006	1.7823	1.6070	0.13

Figure 1 provides a comparison between the LnL scores of the trees in set  $P_1$  for the Rhododendron data set (plots for other data sets are available in the appendix, Figures A.1-A.6). The LnL was computed under all partition models (LB, SB, and UB). Here, the  $x$  axis represents the trees in  $P_1$  in the order they are enumerated by Terraphast I. Figure 1a shows the LnLs as computed under UB model, which should, in principle, all be exactly identical. The slight deviations are due to numerical round of error propagation. The maximal difference amounts to 0.0012 LnL units. Note that, for the LnL scores under the SB and LB models shown in Figure 1b and Figure 1c, the differences between the maximum and the minimum LnLs are also comparatively small (0.5558 and 0.4683 LnL units, respectively). The same holds for Ranunculus and Szygium, where the LnL differences under all partition models are near zero. For the remaining four data sets (Eucalyptus, Euphorbia, Primula, and Rabosky.scincids) the differences in LnL under SB and LB models are larger and vary between 3.4 and 22.4 LnL units. To verify the apparently structured LnL variations that are visible in the plots (presumably connected with the order by which the recursive function in Terraphast I enumerates trees on a terrace), we repeated the calculation of the LnL scores with IQ-Tree. The results resemble those of RAXML-NG (data not shown).

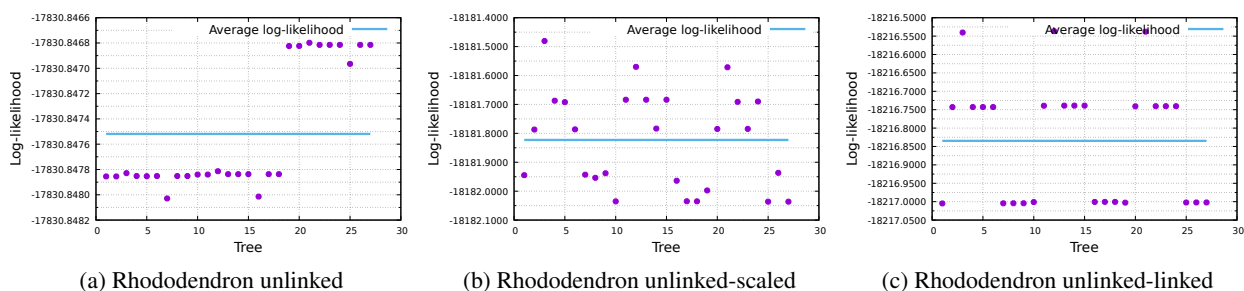


Figure 1: Log-likelihoods for Rhododendron data set under 1a UB, 1b SB, and 1c LB model

### 5.2 Significance tests for the trees from $P_2$ and $P_3$ with LnL under LB and SB partition models, respectively

Table 4 shows the results of the significance tests for the trees on quasi-terraces  $P_2$  and  $P_3$  and the LnL scores under LB and SB, respectively. For each data set, we report the fraction of trees on the quasi-terrace ( $P_2$  or  $P_3$ ), which are not significantly different with respect to their LnL from the best tree in the set (i.e., the tree with the highest LnL).

For the  $P_2$  sets and under the LB model, for 3 out of 7 data sets (Rabosky.scincids, Ranunculus, and Szygium) we find that all tests report that the trees are not significantly different from the best tree. For Primula and Rhododendron 6 tests suggest that the majority of the trees (79.5%-100%) are not significantly different from the best tree. For Eucalyptus and Euphorbia the results are inconclusive and the fractions of trees, which are not significantly different from the best tree, range from 8% to 91%.

Table 4: Significance tests for the trees on quasi-terraces

Data set	Size	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
<i>Results for quasi-terraces <math>P_2</math> with LnL computed under LB partition model</i>								
Primula	2,835	21.2%	96.6%	100%	91.6%	100%	79.5%	97.2%
Euphorbia	897	13.7%	23.1%	52.3%	22.4%	84.4%	27.1%	40.2%
Eucalyptus	267	11.6%	27.3%	91.0%	8.0%	33.3%	39.0%	18.7%
Rhododendron	45	57.8%	100%	100%	93.3%	97.8%	95.6%	100%
Rabosky.scincids	3	100%	100%	100%	100%	100%	100%	100%
Ranunculus	3	100%	100%	100%	100%	100%	100%	100%
Szygium	3	100%	100%	100%	100%	100%	100%	100%
<i>Results for quasi-terraces <math>P_3</math> with LnL computed under SB partition model</i>								
Primula	2,835	21.1%	97.7%	100%	96.1%	100%	78.2%	98.0%
Euphorbia	1,863	6.7%	11.1%	56.5%	11.1%	84.8%	29.5%	41.5%
Eucalyptus	267	20.2%	35.2%	87.3%	20.6%	55.8%	69.7%	25.5%
Rhododendron	45	57.8%	100%	100%	100%	100%	95.6%	100%
Szygium	5	100%	100%	100%	100%	100%	100%	100%
Rabosky.scincids	3	100%	100%	100%	100%	100%	100%	100%

*The values are in percent of trees, which are not significantly worse than the best tree in the set*

The results for  $P_3$  quasi-terraces and LnL SCORES computed under SB resemble those of  $P_2$  and LB. Similarly, for Rabosky.scincids and Szygium all tests report that the trees in  $P_3$  are not significantly different from the best tree. For Primula and Rhododendron, 6 tests report that the majority of trees (78.2%-100%) are not significantly different from the best tree in the set. For Eucalyptus and Euphorbia the results are again inconclusive (the fractions of not significantly different trees vary from 6.7% to 84.8%). Note, that for the Ranunculus data set,  $P_3$  only contains one tree (i.e., the quasi-terrace is trivial). Therefore, significance tests are not available for Ranunculus.

### 5.3 NNI analysis for Rhododendron and a random Yule-Harding tree

For the NNI analysis we choose the Rhododendron data set. It contains sufficient number of trees in the  $P_1$  set (27), but is, at the same time, still small enough in terms of taxa (117) such that the resulting NNI trees (in total 6141, with duplicates already removed) could be analyzed within the cluster runtime limit. For each partition model, LnL computations were completed within the runtime limits for a different number of NNI neighbors. Namely, 2389 NNI neighbors (38.9% out of total number of NNI trees) were evaluated under UB, 3720 NNI neighbors (60.58%) under LB, and 3705 (60.33%) under SB.

For all partition models, we calculated the fraction of NNI trees, whose LnL fall within the range between the maximum and minimum LnL of the trees *on* the (quasi-)terrace  $P_1$ . These fractions amount to 15.8%, 30.2% and 27.9% under UB, LB, and SB, respectively. In addition, we also computed the fraction of NNI trees with a LnL greater than the maximum LnL for the trees in  $P_1$ . The respective fractions of NNI trees are very small with only 3.9%, 1.4%, and 2.2% under UB, LB, and SB, respectively.

We performed the same analysis for a (quasi-)terrace generated using a random Yule-Harding tree. This random tree yields a set  $P_4$  of size 9 and the corresponding NNI neighborhood comprises 2048 trees (duplicates already removed). Here, we found that 18.9% NNI trees fall within the maximum-minimum LnL range of the trees *on* the (quasi-)terrace for UB, 30.3% for LB, and 51.5% for SB. These fractions for UB and LB models are similar to those obtained for  $P_1$ . However, the difference is more pronounced under the SB model, where the fraction of NNI trees falling within

the minimum-maximum LnL range of trees on the quasi-terrace increases from 27.9% to 51.5%. We also calculated NNI trees with LnL scores that are greater than those of the trees *on* the (quasi-)terrace  $P_4$ . The results are: 38.2%, 36.5%, and 18.2% under UB, LB, and SB, respectively. These fractions are substantially greater than those for the (quasi-)terrace  $P_1$  generated using the ML tree. This is expected, as at toward the end of the tree search, the majority of trees surrounding the ML tree (and a (quasi-)terrace) are expected to have a LnL score that is similar to that of the ML tree.

Table 5: Significance tests for the trees on (quasi-)terraces and their NNI neighborhoods (Rhododendron data set)

Model	Trees	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW	p-AU
<i>Results for <math>P_1</math> and its NNI neighborhood</i>								
<b>UB</b>	All trees	6.89%	76.46%	97.37%	59.34%	81.71%	59.57%	80.89%
	Terrace trees	0.00%	100.00%	100.00%	0.00%	0.00%	100.00%	100.00%
	NNI trees	6.92%	76.36%	97.36%	59.60%	82.07%	59.39%	80.80%
<b>LB</b>	All trees	6.29%	78.75%	96.50%	48.98%	87.01%	60.30%	83.04%
	Terrace trees	0.00%	100.00%	100.00%	0.00%	55.56%	100.00%	100.00%
	NNI trees	6.32%	78.65%	96.48%	49.19%	87.15%	61.12%	82.97%
<b>SB</b>	All trees	6.91%	79.12%	96.50%	40.24%	84.14%	61.69%	83.80%
	Terrace trees	0.00%	100.00%	100.00%	0.00%	74.07%	100.00%	100.00%
	NNI trees	6.94%	79.03%	96.48%	40.42%	84.19%	61.52%	83.73%
<i>Results for <math>P_4</math> and its NNI neighborhood</i>								
<b>UB</b>	All trees	1.94%	4.23%	19.54%	4.38%	40.25%	2.43%	7.00%
	Terrace trees	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	NNI trees	1.95%	4.25%	19.63%	4.39%	40.43%	2.44%	7.03%
<b>UB-LB</b>	All trees	1.56%	3.89%	13.08%	2.04%	37.92%	1.99%	2.92%
	Terrace trees	0.00%	0.00%	0.00%	0.00%	22.22%	0.00%	0.00%
	NNI trees	1.56%	3.91%	13.13%	2.05%	37.99%	2.00%	2.93%
<b>UB-SB</b>	All trees	1.51%	3.60%	14.05%	2.14%	47.59%	1.94%	2.92%
	Terrace trees	0.00%	0.00%	0.00%	0.00%	55.56%	0.00%	0.00%
	NNI trees	1.51%	3.61%	14.11%	2.15%	47.56%	1.95%	2.93%

*The values are in percent of trees, which are not significantly worse than the best tree in the set*

For all three models we performed significance tests with IQ-Tree on the expanded set of trees including both, trees on the (quasi-)terrace, and trees in its neighborhood. As in the previous tests, we recorded the fraction of trees that are not significantly different from the best tree in the set. Table 5 summarizes the results: 27 trees in  $P_1$  are denoted in the Table 5 as "Terrace trees" and 6141 NNI trees as "NNI trees". Under the UB model, 3 tests (bp-RELL, p-WKH, and p-WSH) report that all trees in  $P_1$  are significantly worse than the best tree in the set. The remaining 4 tests suggest the opposite (all trees in  $P_1$  are not significantly different). Similarly, under the LB and SB models, bp-RELL and p-WKH report that all trees in  $P_1$  are significantly worse than the best tree, while p-WSH reports that half of trees on quasi-terrace  $P_1$  is significantly worse than the best tree. Again, the remaining 4 tests suggest that all trees in  $P_1$  are not significantly different from the best tree. Overall, 6 tests suggested that the trees on the quasi-terrace behave alike. However, the trend (either all significantly or not significantly worse than the best tree) is not consistent among different tests. For the NNI trees, the results are inconclusive and the fraction of NNI trees that are not significantly worse than the best tree in the set varies substantially for all three branch length models.

We also performed significance tests for the (quasi-)terrace  $P_4$  generated by a random tree and its NNI neighborhood (Table 5). There are 9 terrace trees, 2048 NNI trees, and hence a total of 2059 analyzed trees. Here, the behavior of the trees on the (quasi-) terraces is more consistent across different models and different tests. In contrast to the results for  $P_1$ , under all three models 6 tests report that all trees in  $P_4$  are not significantly worse than the tree with the best LnL.



The p-WSH test reported the same for the UB model, but also suggests that approximately half the trees from  $P_4$  are significantly worse than the best tree under the LB and SB models.

## 6 Discussion

In this study we analyzed 7 empirical data sets and their (quasi-)terraces as generated on the best-found ML trees. We compared the LnL scores of the trees on the (quasi-)terraces under all three partition models. The significance tests showed that trees on quasi-terraces are not significantly different from the best tree in the set with respect to their LnL scores. Therefore, during the tree search we can approximate the LnL scores of all trees on a quasi-terrace by evaluating the LnL of only one representative tree on that quasi-terrace.

For one of our data sets (Rhododendron), we also performed an analysis of the NNI neighborhood of the quasi-terraces  $P_1$  and  $P_4$ , generated by the ML tree under the UB model ( $T_{UB}$ ) and by a random Yule-Harding tree, respectively. Significance tests on these extended tree sets, including trees from the quasi-terrace and its NNI neighborhood, suggest that trees on a quasi-terrace behave similarly. This again implies that we can potentially use only one representative tree from the quasi-terrace during the tree search to accelerate the exploration of the tree space region containing that quasi-terrace. Such an approximation (using only one representative tree) can be beneficial, in particular during the initial phases of the tree search, to rapidly navigate out of regions with low LnL scores.

Our study constitutes a preliminary inspection of quasi-terraces and a more thorough study is required to confirm our results at a broader scale. Nevertheless, our findings suggest that quasi-terraces indeed exist and can be helpful for increasing the efficiency of tree search algorithms.

## References

- Biczok, R., Bozsoky, P., Eisenmann, P., Ernst, J., Ribizel, T., Scholz, F., ... Stamatakis, A. (2018, 05). Two C++ libraries for counting trees on a phylogenetic terrace. *Bioinformatics*, 34(19), 3399–3401. Retrieved from <https://doi.org/10.1093/bioinformatics/bty384> doi: 10.1093/bioinformatics/bty384
- Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, F., van der Bank, M., Chase, M. W., & Hodkinson, T. R. (2008). Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular phylogenetics and evolution*, 47(2), 488–505. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1055790308000584> doi: <https://doi.org/10.1016/j.ympev.2008.01.035>
- Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, 84, 53–63. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1055790314004217> doi: <https://doi.org/10.1016/j.ympev.2014.12.003>
- Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016, 04). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology*, 65(6), 997–1008. Retrieved from <http://dx.doi.org/10.1093/sysbio/syw037> doi: 10.1093/sysbio/syw037
- Dell’Ampio, E., Meusemann, K., Szucsich, N. U., Peters, R. S., Meyer, B., Borner, J., ... Misof, B. (2013, 10). Decisive Data Sets in Phylogenomics: Lessons from Studies on the Phylogenetic Relationships of Primarily Wingless Insects. *Molecular Biology and Evolution*, 31(1), 239–249. Retrieved from <https://doi.org/10.1093/molbev/mst196> doi: 10.1093/molbev/mst196
- Dobrin, B. H., Zwickl, D. J., & Sanderson, M. J. (2018, April). The prevalence of terraced treescapes in analyses of phylogenetic data sets. *BMC Evolutionary Biology*, 18(1), 46. Retrieved from <https://doi.org/10.1186/s12862-018-1162-9> doi: 10.1186/s12862-018-1162-9
- Duchêne, D. A., Tong, K. J., Foster, C. S. P., Duchêne, S., Lanfear, R., & Ho, S. Y. W. (2018). Linking Branch Lengths Across Loci Provides the Best Fit for Phylogenetic Inference. *bioRxiv*. Retrieved from <https://www.biorxiv.org/content/early/2018/11/09/467449> doi: 10.1101/467449
- Fabre, P. H., Rodrigues, A., & Douzery, E. J. P. (2009). Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear DNA. *Molecular Phylogenetics and Evolution*, 53(3), 808–825. Retrieved from <http://www.sciencedirect.com/science/article/pii/S10557903090003169> doi: <https://doi.org/10.1016/j.ympev.2009.08.004>
- Harding, E. F. (1971). The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability*, 3(1), 44–77. Retrieved from <https://doi.org/10.2307/1426329> doi: 10.2307/1426329
- Hinchliff, C. E., & Roalson, E. H. (2012, 12). Using Supermatrices for Phylogenetic Inquiry: An Example Using the Sedges. *Systematic Biology*, 62(2), 205–219. Retrieved from <https://doi.org/10.1093/sysbio/sys088> doi: 10.1093/sysbio/sys088

- Kearney, M. (2002). Fragmentary Taxa, Missing Data, and Ambiguity: Mistaken Assumptions and Conclusions. *Systematic Biology*, 51(2), 369–381. Retrieved from <http://www.jstor.org/stable/3070797>
- Kishino, H., & Hasegawa, M. (1989, Aug 01). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution*, 29(2), 170–179. Retrieved from <https://doi.org/10.1007/BF02100115> doi: 10.1007/BF02100115
- Kishino, H., Miyata, T., & Hasegawa, M. (1990, Aug 01). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution*, 31(2), 151–160. Retrieved from <https://doi.org/10.1007/BF02109483> doi: 10.1007/BF02109483
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *bioRxiv*. Retrieved from <https://www.biorxiv.org/content/early/2019/03/05/447110> doi: 10.1101/447110
- Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., ... Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, 334(6055), 521–524. Retrieved from <https://science.sciencemag.org/content/334/6055/521> doi: 10.1126/science.1211028
- Miadlikowska, J., Kauff, F., Högnabba, F., Oliver, J. C., Molnár, K., Fraker, E., ... Stenroos, S. (2014). A multigene phylogenetic synthesis for the class Lecanoromycetes (Ascomycota): 1307 fungi representing 1139 infrageneric taxa, 317 genera and 66 families. *Molecular Phylogenetics and Evolution*, 79, 132–168. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1055790314001298> doi: <https://doi.org/10.1016/j.ympev.2014.04.003>
- Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., ... Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346(6210), 763–767. Retrieved from <https://science.sciencemag.org/content/346/6210/763> doi: 10.1126/science.1257570
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014, 11). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. Retrieved from <https://doi.org/10.1093/molbev/msu300> doi: 10.1093/molbev/msu300
- Nyakatura, K., & Bininda-Emonds, O. R. P. (2012, February). Updating the evolutionary history of Carnivora (Mammalia): a new species-level supertree complete with divergence time estimates. *BMC biology*, 10(1), 12. Retrieved from <https://doi.org/10.1186/1741-7007-10-12> doi: 10.1186/1741-7007-10-12
- Pyron, R. A., Burbrink, F. T., Colli, G. R., de Oca, A. N. M., Vitt, L. J., Kuczynski, C. A., & Wiens, J. J. (2011). The phylogeny of advanced snakes (Colubroidea), with discovery of a new subfamily and comparison of support methods for likelihood trees. *Molecular Phylogenetics and Evolution*, 58(2), 329–342. Retrieved from <http://www.sciencedirect.com/science/article/pii/S105579031000429X> doi: <https://doi.org/10.1016/j.ympev.2010.11.006>
- Rabosky, D. L., Donnellan, S. C., Grundler, M., & Lovette, I. J. (2014, 3). Analysis and Visualization of Complex Macroevolutionary Dynamics: An Example from Australian Scincid Lizards. *Systematic Biology*, 63(4), 610–627. Retrieved from <https://doi.org/10.1093/sysbio/syu025> doi: 10.1093/sysbio/syu025
- Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, 53(1-2), 131–147. Retrieved from <http://www.sciencedirect.com/science/article/pii/0025556481900432> doi: [https://doi.org/10.1016/0025-5564\(81\)90043-2](https://doi.org/10.1016/0025-5564(81)90043-2)
- Rosa, B. B., Melo, G. A. R., & Barbeitos, M. S. (2019, 01). Homoplasy-Based Partitioning Outperforms Alternatives in Bayesian Analysis of Discrete Morphological Data. *Systematic Biology*, 68(4), 657–671. Retrieved from <https://doi.org/10.1093/sysbio/syz001> doi: 10.1093/sysbio/syz001
- Sanderson, M. J., McMahon, M. M., Stamatakis, A., Zwickl, D. J., & Steel, M. (2015, 05). Impacts of Terraces on Phylogenetic Inference. *Systematic Biology*, 64(5), 709–726. Retrieved from <https://doi.org/10.1093/sysbio/syv024> doi: 10.1093/sysbio/syv024
- Sanderson, M. J., McMahon, M. M., & Steel, M. (2011). Terraces in Phylogenetic Tree Space. *Science*, 333(6041), 448–450. Retrieved from <http://science.sciencemag.org/content/333/6041/448> doi: 10.1126/science.1206357
- Shi, J. J., & Rabosky, D. L. (2015). Speciation dynamics during the global radiation of extant bats. *Evolution*, 69(6), 1528–1545. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.12681> doi: 10.1111/evo.12681
- Shimodaira, H. (2002, 05). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, 51(3), 492–508. Retrieved from <https://doi.org/10.1080/10635150290069913> doi: 10.1080/10635150290069913
- Shimodaira, H., & Hasegawa, M. (1999, 08). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution*, 16(8), 1114–1114. Retrieved from <https://doi.org/10.1093/oxfordjournals.molbev.a026201> doi: 10.1093/oxfordjournals.molbev.a026201
- Soltis, D. E., Mort, M. E., Latvis, M., Mavrodiev, E. V., O'Meara, B. C., Soltis, P. S., ... Rubio de Casas, R.

- (2013). Phylogenetic relationships and character evolution analysis of Saxifragales using a supermatrix approach. *American Journal of Botany*, 100(5), 916–929. Retrieved from <https://bsapubs.onlinelibrary.wiley.com/doi/abs/10.3732/ajb.1300044> doi: 10.3732/ajb.1300044
- Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., ... Murphy, W. J. (2012, November). Macroevolutionary Dynamics and Historical Biogeography of Primate Diversification Inferred from a Species Supermatrix. *PLOS ONE*, 7(11), 1-23. Retrieved from <https://doi.org/10.1371/journal.pone.0049521> doi: 10.1371/journal.pone.0049521
- Stamatakis, A., & Alachiotis, N. (2010, 06). Time and memory efficient likelihood-based tree searches on phylogenomic alignments with missing data. *Bioinformatics*, 26(12), i132–i139. Retrieved from <http://dx.doi.org/10.1093/bioinformatics/btq205> doi: 10.1093/bioinformatics/btq205
- Strimmer, K., & Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1487), 137–142. Retrieved from <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2001.1862> doi: 10.1098/rspb.2001.1862
- Tolley, K. A., Townsend, T. M., & Vences, M. (2013). Large-scale phylogeny of chameleons suggests African origins and Eocene diversification. *Proceedings of the Royal Society B: Biological Sciences*, 280(1759), 20130184. Retrieved from <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2013.0184> doi: 10.1098/rspb.2013.0184
- Van Der Linde, K., Houle, D., Spicer, G. S., & Stepan, S. J. (2010). A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genetics Research*, 92(1), 25–38. Retrieved from <https://www.cambridge.org/core/journals/genetics-research/article/supermatrixbased-molecular-phylogeny-of-the-family-drosophilidae/8C9158886CC27DA191816203BA3462DA> doi: 10.1017/S001667231000008X
- Wang, H.-C., Susko, E., & Roger, A. J. (2019, 04). The Relative Importance of Modeling Site Pattern Heterogeneity Versus Partition-Wise Heterotachy in Phylogenomic Inference. *Systematic Biology*. Retrieved from <https://doi.org/10.1093/sysbio/syz021> doi: 10.1093/sysbio/syz021
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., ... Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, 111(45), E4859–E4868. Retrieved from <https://www.pnas.org/content/111/45/E4859> doi: 10.1073/pnas.1323926111
- Wiens, J. J., & Morrill, M. C. (2011, 03). Missing Data in Phylogenetic Analysis: Reconciling Results from Simulations and Empirical Data. *Systematic Biology*, 60(5), 719–731. Retrieved from <https://doi.org/10.1093/sysbio/syr025> doi: 10.1093/sysbio/syr025
- Yang, Y., Moore, M. J., Brockington, S. F., Soltis, D. E., Wong, G. K.-S., Carpenter, E. J., ... Smith, S. A. (2015, 04). Dissecting Molecular Evolution in the Highly Diverse Plant Clade Caryophyllales Using Transcriptome Sequencing. *Molecular Biology and Evolution*, 32(8), 2001–2014. Retrieved from <https://doi.org/10.1093/molbev/msv081> doi: 10.1093/molbev/msv081
- Yang, Z. (1996, May 01). Maximum-likelihood models for combined analyses of multiple sequence data. *Journal of Molecular Evolution*, 42(5), 587–596. Retrieved from <https://doi.org/10.1007/BF02352289> doi: 10.1007/BF02352289
- Zanne, A. E., Tank, D. C., Cornwell, W. K., Eastman, J. M., Smith, S. A., FitzJohn, R. G., ... others (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature Publishing Group*, 506(7486), 89. Retrieved from <https://doi.org/10.1038/nature12872> doi: 10.1038/nature12872

## A Additional plots

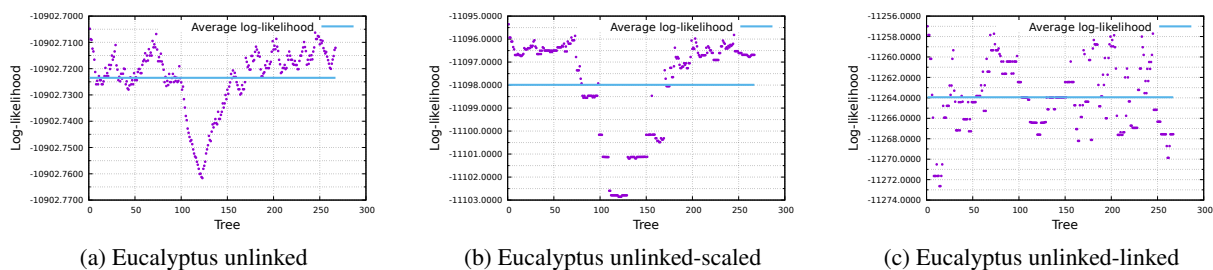


Figure A.1: Log-likelihoods for the data set Eucalyptus under A.1a UB, A.1b UB-SB, and A.1c UB-LB model

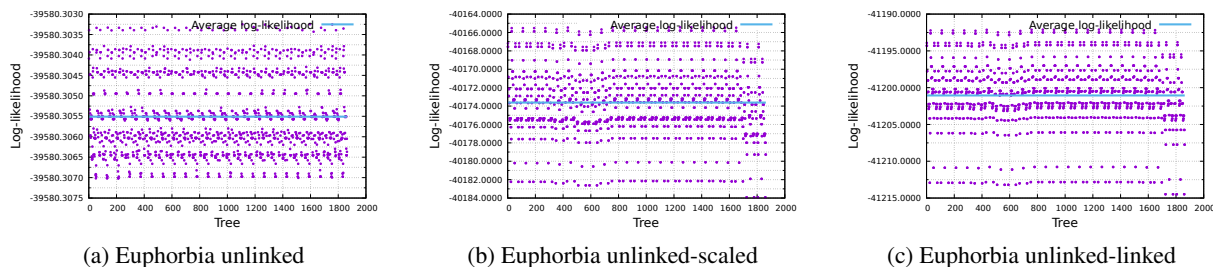


Figure A.2: Log-likelihoods for the data set Euphorbia under A.2a UB, A.2b UB-SB, and A.2c UB-LB model

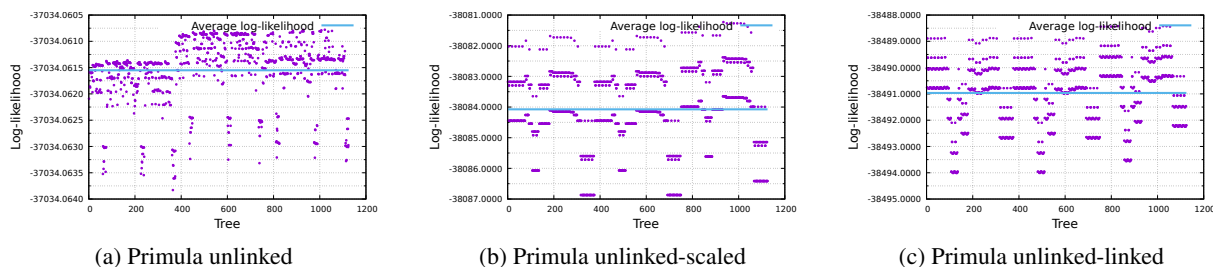


Figure A.3: Log-likelihoods for the data set Primula under A.3a UB, A.3b UB-SB, and A.3c UB-LB model

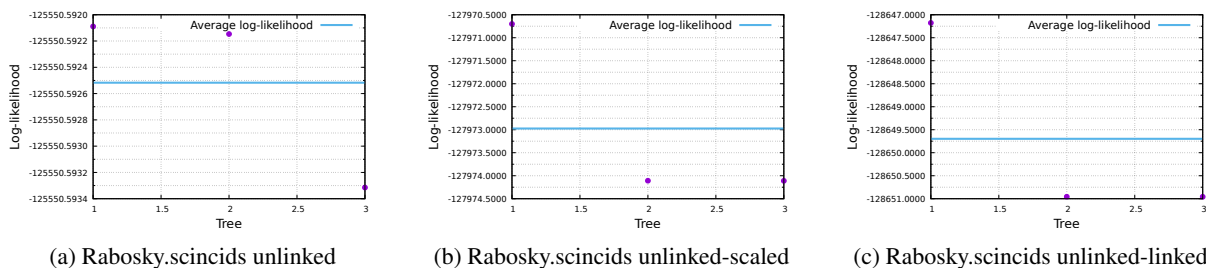


Figure A.4: Log-likelihoods for the data set Rabosky.scincids under A.4a UB, A.4b UB-SB, and A.4c UB-LB model

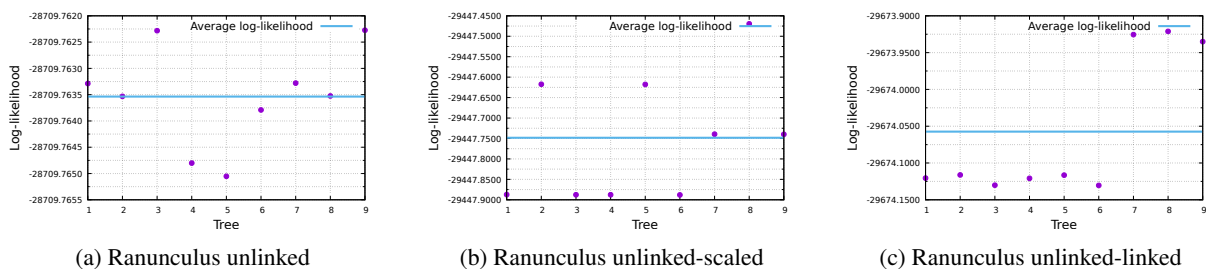


Figure A.5: Log-likelihoods for the data set Ranunculus under A.5a UB, A.5b UB-SB, and A.5c UB-LB model

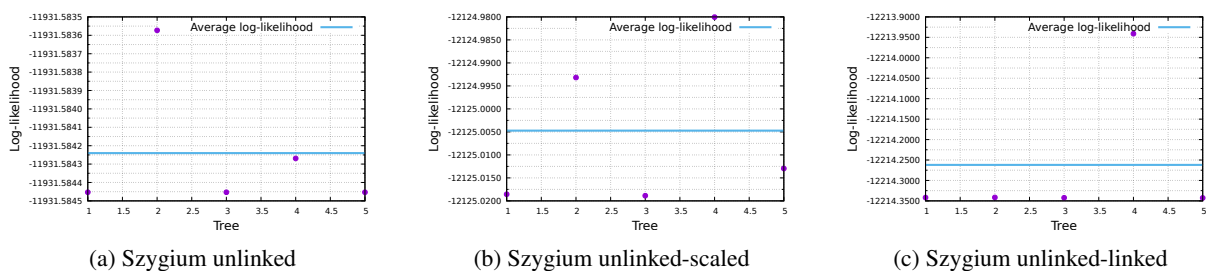


Figure A.6: Log-likelihoods for the data set Szygium under A.6a UB, A.6b UB-SB, and A.6c UB-LB model