Automation and Evaluation of the SOWH Test of Phylogenetic Topologies with SOWHAT

Samuel H. Church[1,*], Joseph F. Ryan[2,3], Casey W. Dunn[1]

1 Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island, United States of America

2 Whitney Laboratory for Marine Biosciences, St. Augustine, Florida, United States of America

3 Sars International Centre For Marine Molecular Biology, Bergen, Norway

∗ **E-mail: samuel_church@brown.edu**

Abstract

The Swofford-Olsen-Waddell-Hillis (SOWH) test is a method to evaluate incongruent phylogenetic topologies. It is used, for example, when an investigator wishes to know if the maximum likelihood tree recovered in their analysis is significantly different than an alternative phylogenetic hypothesis. The SOWH test compares the observed difference in likelihood between the topologies to a null distribution of differences in likelihood generated by parametric resampling. The SOWH test is a well-established and important phylogenetic method, but it can be difficult to implement and its sensitivity to various factors is not well understood. We wrote SOWHAT, a program that automates the SOWH test. In test analyses, we find that variation in parameter estimation as well as the use of a more complex model of parameter estimation have little impact on results, but that results can be inconsistent when an insufficient number of replicates are used to estimate the null distribution. We provide methods of analyzing the sampling as well as a simple stopping criteria for sufficient bootstrap replicates, which increase the overall reliability of the approach. Applications of the SOWH test should include explicit evaluations of sampling adequacy. SOWHAT is available for download from https://github.com/josephryan/SOWHAT.

Introduction

A phylogenetic topology test evaluates whether the difference in optimality criteria between incongruent hypotheses is significant. In some cases, the test is used to determine whether a dataset provides significantly more support for one of several previously proposed relationships. In other cases, the test is used to compare a novel or unexpected phylogenetic result to a previously proposed hypothesis. In each case, the observed difference in optimality criteria between trees is compared to an estimated null distribution of differences in optimality criteria. Phylogenetic topology tests differ largely in how this null distribution is created. The Approximately Unbiased test (AU) [1] and the Kishino-Hasegawa test (KH) [2] create a null distribution by analyzing datasets created by sampling with replacements from the original dataset [3]. However, this approach can lead to selection bias and is only appropriate for tests of hypotheses selected *a priori*, such as the comparison of two alternative hypotheses from the literature. In cases where hypotheses are not determined *a priori*, as when comparing a topology to the maximum likelihood tree produced from the same data, parametric tests like the SOWH test are more appropriate [3]. The SOWH test compares differences in likelihood to a distribution of differences in likelihood generated from datasets simulated under the null hypothesis [4].

Though the SOWH test is more appropriate than other tests for many of the questions that biologists routinely face, there are multiple technical barriers to its routine use. The SOWH test has been shown to produce erroneous results when the model is misspecified, which can lead to type I error [3] [5]. Several helpful step-by-step instructions are available [6] [7], but these manual approaches require extensive hands-on time and make it difficult to systematically examine the behavior of the test under different conditions. It can be especially difficult to manually conduct a SOWH test on a large dataset which includes additional complexities such as a partitioning scheme or undetermined sites. Performing such a test requires an investigator to make multiple decisions, and without clear evaluation of the behavior of the test these decisions may not be informed.

To address these challenges, we developed SOWHAT (as in, ”The maximum likelihood tree differs from my hypothesized phylogeny, so what?”), a program that automates the SOWH test (Fig 1) and is applicable to partitioned datasets. SOWHAT also provides convenient tools for assessing various aspects of its performance. Using SOWHAT, we examine the performance of the SOWH test on three datasets to better understand its strengths and limitations in typical applications. We specifically examine the sensitivity of the test to the number of replicate simulations, model selection, variability in parameter and likelihood estimation, and incorporation of undetermined sites.

Materials and Methods

Implementation of the SOWH test

SOWHAT, our implementation of the SOWH test, is available at https://github.com/josephryan/SOWHAT. This tool compares the maximum likelihood tree to an alternative topology specified by the user (Fig 1). The user specifies the model of evolution and the parameters under which the maximum likelihood tree will be evaluated, and provides as input an alignment file (in phylip format) as well as the topology to be tested against the maximum likelihood tree. Two phylogenetic maximum likelihood trees are then inferred with RAxML [8] - an unconstrained tree, and a tree constrained according to the topology to be tested. The difference between the likelihood scores of these trees is the test statistic that will be evaluated.

New alignments are simulated by Seq-Gen [9]. The alignments are generated using the topology, branch lengths, and model parameters (i.e., state frequencies, rates, and the alpha parameter for the gamma rate heterogeneity approximation) from the constrained analysis as inferred by RAxML. If the original dataset is partitioned, parameters are estimated separately for each partition and the alignments are generated following the partitioning scheme. Each of the simulated alignments is then evaluated using RAxML, again using both an unconstrained search and a constrained search. The difference in likelihood of these two searches is calculated for each simulated dataset, and the set of these differences make up the null distribution.

The test statistic is then compared against the null distribution in a one-tailed test. With each new value added to the null distribution, the program recalculates the statistical measures, including the cumulative mean and the relative standard error of the null distribution, as well as the z-score and p-value of the test statistic against this distribution. The p-value of the sample statistic represents the chance that such a difference would be observed under the null hypothesis. Plots of the relative standard error and cumulative mean are created after each iteration to aid in evaluating the sample size.

SOWHAT can be used to evaluate a hypothesized topology given datasets of nucleotide, amino acid, or binary characters. The model options recognized by the program are a subset of those accepted by RAxML. SOWHAT also allows for the parameters to be estimated and the sequences to be generated using the CAT-GTR model and the program PhyloBayes (Fig 2). Methods of model specification are discussed below.

Stopping Criteria

SOWHAT provides multiple methods of determining sufficient sample size, including a simple stopping criterion based on the convergence. This stopping criteria is adapted from criteria established for non-parametric bootstrap analyses [10].

For each new dataset simulated, the z-score is calculated. As the test is increased by one more simulation the ratio between this score and the previous is calculated. This ratio represents the similarity of the two one-tailed tests as the sample size is increased by one. As the ratio converges to 1, the value of 1−ratio of the ratio will approach 0. SOWHAT calculates the percentage of resampling events which return values of 1−ratio below a certain threshold (default is 0.01). A percentage in excess of 50% is strong evidence of convergence and subsequent sampling is unlikely to change the result of the test. For increased reliability, the option is available to extend the test for some number of resampling events (i.e., 50) beyond this point.

SOWHAT also allows multiple independent runs of the test to be executed simultaneously, each with multiple replicate simulations. These runs can be compared to each other to evaluate consistency. A plot is generated showing the mean value of the null distribution of each run as the sample size increases. As the sample size increases, the mean values of the distribution will converge, reflecting the similarity of the independent tests. This method was used to explore the efficacy of the stopping criteria proposed above.

Additional modifications to the SOWH test

In the standard SOWH test, the test statistic and model parameters are estimated once at the outset of the run. These same model parameters are then used to simulate all the replicates. SOWHAT can optionally recalculate the test statistic and parameters for each replicate (Fig 3). Rather than produce a single test statistic, this creates a distribution of test statistics that can then be summarized (e.g., by taking the mean) and compared to the null distribution.

For datasets which contain undetermined sites (gaps), SOWHAT can be instructed to propagate these gaps into each simulated alignment. This guarantees that the same number and pattern of gaps are present in each simulated alignment.

Examined datasets

We selected three datasets for examination. They are included with SOWHAT so that users can test their installation and easily reproduce the results presented here. The results presented here were prepared with the version of SOWHAT available at:

https://github.com/josephryan/sowhat/tree/88d8dbc403a67d8c41ef4ec97df377ea968a0e81.

ATGC dataset (amino acid): This dataset was generated by the ATGC group [11] to test of the program PHYML. Previous analyses recovered relationships that are well established and others that are more tentative. The matrix contains 11 taxa and 391 amino acid characters. The topology t1 differs from the most likely tree in the relationship of one taxon, referred to as taxon8.

Rodent 12S dataset (nucleotide): This dataset of mitochondrial ribosomal RNA was originally assembled by Sullivan and coworkers [12]. Analysis of this alignment produces a topology of sigmodontine Rodents, which differs from the topology produced by analysis of morphological, chromosomal, allozyme, and other DNA datasets. The topology t1 represents the accepted species topology [5]. This set was used in an extensive analysis of the SOWH test [5]. The dataset contains eight taxa and 791 characters.

Reptile 13 gene dataset (nucleotide): This dataset, assembled by Castoe et al [13], contains all 13 protein coding mitochondrial genes for 34 squamate reptiles and 6 tetrapod outgroup species. Analyses of this dataset strongly contradicts analyses of morphological and nuclear data [13]. The topology t1 represents the species topology inferred from the morphology and nuclear genetic information [5].

Analysis of the behavior of the SOWH test

SOWHAT was initially run multiple times over each dataset with 100 parametric bootstraps calculated for each run, which is an arbitrary yet commonly used sample size. SOWHAT was then run multiple times for each dataset using the stopping criteria described above. The distribution of p-values was examined and compared to the previous tests (Fig 4; Table 1). For the Rodent dataset, the test was again run multiple times with a sample size of 500 and the p-values were compared .

GTR+Γ was used as the model for evaluating the likelihood as well as for estimating the parameters used in simulating the datasets for the Rodent and Reptile datasets. For the ATGC dataset, the LG model of protein evolution was used for evaluating the likelihood [14]. For protein datasets, SOWHAT runs RAxML additionally using the GTR+unlinked model to optimize the free parameters.

For all datasets, a single SOWH test was performed using the CAT-GTR model and the program PhyloBayes to estimate the parameters and simulate the datasets for the null distribution (likelihood evaluation was still performed using the specified model and RAxML).

Each of these tests used the stopping criteria to determine sufficient sampling. For the Reptile dataset, using PhyloBayes to simulate the datasets required that we used a matrix with no fully undetermined columns. To account for this difference, we also ran a test using GTR+ with RAxML and Seq-Gen. For the Reptile dataset, SOWH tests were performed without the partitioning scheme under both the GTR+ and the CAT-GTR model with PhyloBayes. Finally, for the Reptile analysis, a single SOWH test was performed which propagated the number and location of the undetermined sites, or gaps, present in the real-world dataset into the simulated datasets.

Each of these analyses is included with the SOWHAT package as an executable file (analysis.sh). This file will produce the distribution of p-values as well as the plots like those in Figures 5.

Results and Discussion

The impact of replicate number

One of the greatest challenges to applying and interpreting the SOWH test is deciding how many replicate simulations to sample when building the null distribution. It is critical that a sufficient number of replicate samples are simulated in a run. Without adequate sampling, stochastic sampling error can lead to poor estimates of the null distribution and inconsistent evaluation of the p-value.

To explore the effects of sample size, we ran multiple SOWH tests and compared the results (Fig 4; Table 1). For two of the datasets analyzed here (the ATGC and the Reptile datasets) the SOWH test returned consistent p-values across multiple tests with a sample size of 100. For the ATGC dataset, the the test statistic was $-2.2 * 10^-5$ and the p-value was from 0.998 to 0.999. For the Reptile dataset, the the test statistic was 175.6375 and the p-value was very close to 0. However, for the Rodent dataset, multiple independent SOWH tests returned inconsistent p-values when the sample size was set to 100. The p-values ranged from $4.980209 * 10^-07$ to 0.707524 , and 33% of runs fell above the significance threshold of 0.05.

These results indicate that a single run of the SOWH test with an arbitrary number of replicates is not sufficient to merit confidence in the results. When reporting the results of any SOWH test, the chosen sample size should be justified. The stopping criteria outlined in the methods suggests sampling until it becomes increasingly likely that the subsequent parametric resamples will not significantly alter the null distribution, therefore the p-value is unlikely to change with greater sampling.

For the ATGC and Reptile matrices, using the stopping criteria returned very similar results to using an arbitrary sample size. For the ATGC matrix, the stopping criteria found that 100 is a sufficient number of resamples, so the test was unchanged. For the Reptile dataset, the stopping criteria suggested a sample size of 183, at which the p-values of all tests remained close to 0. For the Rodent data, the stopping criteria suggested increasing the sample size to 437. At this sample size, the p-values across runs were all below the significance level, ranging from ranging from 0.00394 to $1.281568 * 10^{-}14$.

SOWHAT also provides further methods to evaluate the sufficiency of the sample size, including reporting the relative standard error and variance of the null distribution, and plotting the inverse ratio of z-scores and cumulative mean as the sample size is increased (Fig 4). From these plots, it is clear that an arbitrary value of 100 is not always sufficient to ensure that the null distribution is well sampled.

The stopping criteria outlined above was successful in each of the subsequent tests outlined in this paper. With the exception of one, all tests suggested a sample size between 100 and 500, which is consistent with stopping criteria for bootstrap tests [10]. One test on the Reptile dataset, which tested the effects of model selection as described below, was unable to converge until above 900. The SOWH test is not universally informative for all datasets. When the stopping criteria suggests a sample size at a very high value such as this one, it is likely because the test statistic falls very close to the line of significance at even large sample sizes. In such a case, it might be concluded that the SOWH test is not capable, given the data, of producing an informative results on the validity of the hypothesis.

The impact of model specification

Model selection has previously been shown to have an impact on the outcome of the SOWH test [3] [5]. In order to minimize Type I errors associated with model misspecification, it has been suggested that the SOWH test be run with more complex substitution models.

To provide more flexibility in model specification, SOWHAT provides the option to use the more complex sequence evolution models generated by PhyloBayes for estimating the parameters used in simulation [15]. In this context, PhyloBayes is used only for estimating the parameters for simulating the datasets and not for estimating likelihood values (Fig 2).

Previous explorations of model selection in the SOWH test have focused on analyses which used the same model for parameter estimation and likelihood evaluation [5], and have also suggested that this may lead to Type I error or overconfidence in the hypothesized topology. Using a different model to optimize parameters for simulation may lead to a more robust test overall.

To test the impact of model complexity, we analyzed all datasets using both the GTR+ model as implemented in RAxML and the more complex CAT-GTR model as implemented in PhyloBayes. In no case did the SOWH tests run with more complex models return a p-value that differed in interpretation from the p-values calculated using the less complex model of substitution.

For the ATGC dataset the p-value was completely consistent with the values returned using GTR+Γ under RAxML. The mean values of the null distributions are slightly greater when evaluated with Phylobayes, which results in p-values closer to 1 than those calculated using the less complex LG model (Table 1).

For the Rodent dataset, Buckley (2002) reported that the most likely tree differs from the accepted species tree and that SOWH incorrectly refutes the true species tree, which is a case of Type I error. Using the CAT-GTR model, the mean of the null distribution was much lower and the distribution converged much faster than with a less complex model of substitution. This test also rejects the hypothesis with a p-value of $2.34 * 10^{-}256$,GTR+Γ. This is not consistent with the prediction that a more complex model of substitution would result in a lower rate of type 1 error. If the t1 constraint specified here is in fact the true evolutionary tree and not a product of convergent evolution, than a more complex model also produces the erroneous result. The null distribution had a lower mean using the more complex model, which also violates the hypothesis that using the same model for estimation and optimization will result in the likelihood engine more easily finding the hypothesized topology and returning a null distribution closer to 0 than otherwise.

For the Reptile dataset, the CAT-GTR model also returned a p-value with the same interpretation as the p-values using the less complex model. The mean value of the null distribution, however, was much larger and resulted in a p-value which approached 0 only after many resamples were calculated. The fact that the sample size must be very large, and that the test statistic is so close to the null distribution, leads us to believe that this test may not be able to return a reliable answer on the validity of the hypothesis.

The impact of partitioning

Partitioning schemes are a key part of many phylogenetic analyses. SOWHAT allows for the datasets to be simulated using simple partitioning schemes, which allows the SOWH test to be applied to datasets which have previously been virtually impossible.

To test the impact of the partitioning scheme, we compared SOWH tests which used no partitioning scheme using both the GTR+Γ and the CAT-GTR models and the Reptile dataset. Without multiple partitions, the test statistic is changed slightly, which is not unexpected. The p-values returned by these tests, however, are consistent with the partitioned tests. The mean value of the null distribution generated is somewhat different in both cases, though not enough to alter the outcome of the test. This shows that SOWHAT can be reliably used on datasets which are partitioned, though in this singular dataset there was no impact on results

The impact of parameter estimation

Another possible source of unreliable results is variability in the calculations of test statistic and estimation of the parameter values. This can be caused by the heuristic phylogenetic searches not always finding the same tree when presented with the same data.

We provide an option in SOWHAT that allows the initial trees, parameters, and test statistics to be recalculated. Recalculating these values creates a distribution of test statistics, the mean of which is tested against the null distribution to calculate the p-value (Figure 3). Any variation in the parameters and topologies, as well as the test statistic, is accounted for using this method.

Similar p-values are observed using this adjustment (Table 1). This indicates that variation in parameter estimation and the calculation of the test statistic are not a source of inconsistency in these analyses.

The impact of gaps

Many datasets include undetermined sites, yet these are not commonly a component of SOWH tests. The purpose of simulating datasets under parametric conditions is to recreate the situation under which the real-world data was generated. In order to simulate datasets which more closely reflect the nature of the real-world data, SOWHAT can propagate the gaps present in the real data into all simulated datasets.

We tested the effects of this feature using the Reptile dataset under the GTR+Γ model, and found that there was little impact. The mean distribution was somewhat smaller, resulting in a p-value which approached 0 more quickly (Table 1).

Conclusion

Performing a SOWH test on a large, multilocus dataset requires many decisions to be made, all of which have the potential to change the effectiveness of the test. We find that the most impactful choice an investigator must make is that of sample size. Though the primary concern with the application of the SOWH test has been model specification, the analyses presented here were robust to model specification. The test was also robust to the incorporation of a partitioning scheme, parameter estimation and calculation of the test statistic, and the presence or absence of gaps in simulated datasets. In future applications of the SOWH test, investigators should explicitly justify the number of replicates. Our new tool, SOWHAT, makes it much simpler for investigators to examine the adequacy of replicate number and to explore the sensitivity of the test to other factors.

Acknowledgments

SHC was funded by the LINK Award through Brown University and the Marine Biological Laboratory at Woods Hole, as well as the Sars International Centre for Marine Molecular Biology. JFR was funded by The Whitney Laboratory for Marine Bioscience and Sars International Centre for Marine Molecular Biology. We thank Nicholas Sinnott-Armstrong, Felipe Zapata, Mark Howison, Peisi Yan, Stefan Siebert, Andreas Hejnol for comments, ideas, and support. Those acknowledged do not necessarily endorse the findings in this paper.

References

1. Shimodaira H (2001) Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection. Communications in Statistics-Theory and Methods 30: 1751–1772.

2. Kishino H, Hasegawa M (1989) 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: 170–179.

3. Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Systematic Biology 49: 652–670.

4. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) {Phylogenetic inference} .

5. Buckley TR (2002) Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Systematic Biology 51: 509–523.

6. Crawford AJ (2009). horribly detailed instructions to running a parametric bootstrap test @ONLINE. URL `http://dna.ac/genetics.html`.

7. Anderson J, Goldman N, Rodrigo A. Guidelines for performing the sowh test @ONLINE. URL `http://www.ebi.ac.uk/goldman/tests/SOWHinstr.html`.

8. Stamatakis A (2006) Raxml-vi-hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.

9. Rambaut A, Grass NC (1997) Seq-gen: an application for the monte carlo simulation of dna sequence evolution along phylogenetic trees. Computer applications in the biosciences: CABIOS 13: 235–238.

10. Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A (2009) How many bootstrap replicates are necessary? In: Research in Computational Molecular Biology. Springer, pp. 184–200.

11. Guindon S, Gascuel O. Phyml 3.0 benchmarks @ONLINE. URL `http://www.atgc-montpellier.fr/phyml/benchmarks/index.php?ben=md`.

12. Sullivan J, Holsinger KE, Simon C (1995) Among-site rate variation and phylogenetic analysis of 12s rrna in sigmodontine rodents. Molecular Biology and Evolution 12: 988–1001.

13. Castoe TA, de Koning APJ, Kim HM, Gu W, Noonan BP, et al. (2009) Evidence for an ancient adaptive episode of convergent molecular evolution. Proceedings of the National Academy of Sciences 106: 8986-8991.

14. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Molecular biology and evolution 25: 1307–1320.

15. Lartillot N, Lepage T, Blanquart S (2009) Phylobayes 3: a bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25: 2286–2288.

Figure Legends

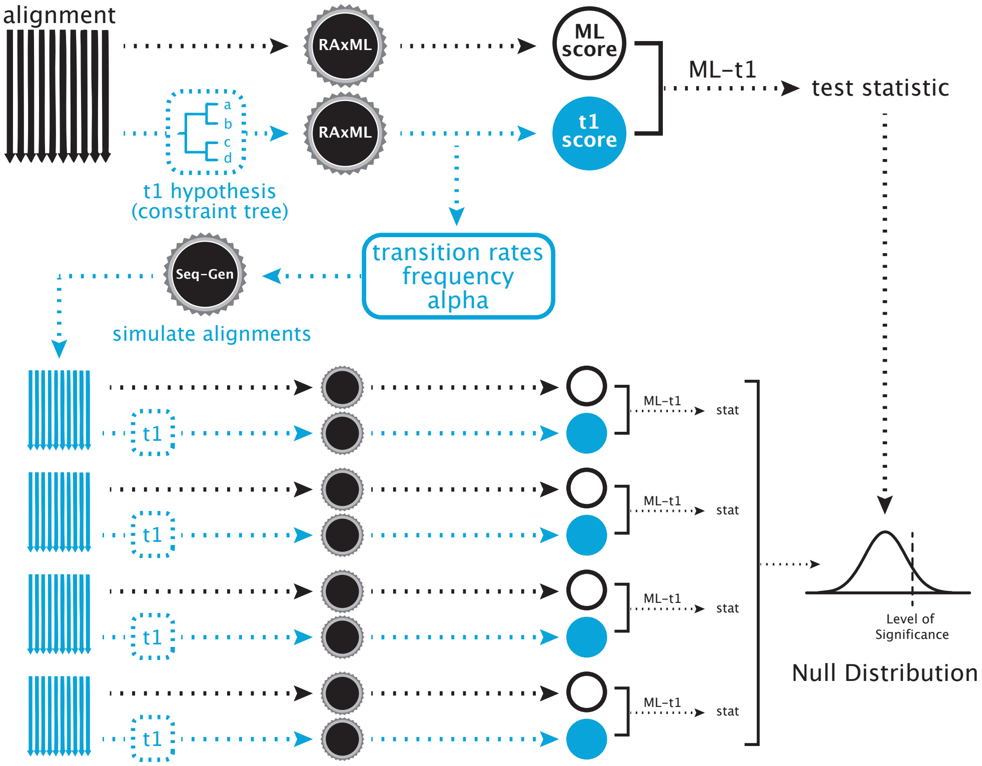

Figure 1. Program flow of SOWHAT The sample size of the distribution is the number of alignments generated based on the estimated parameters.

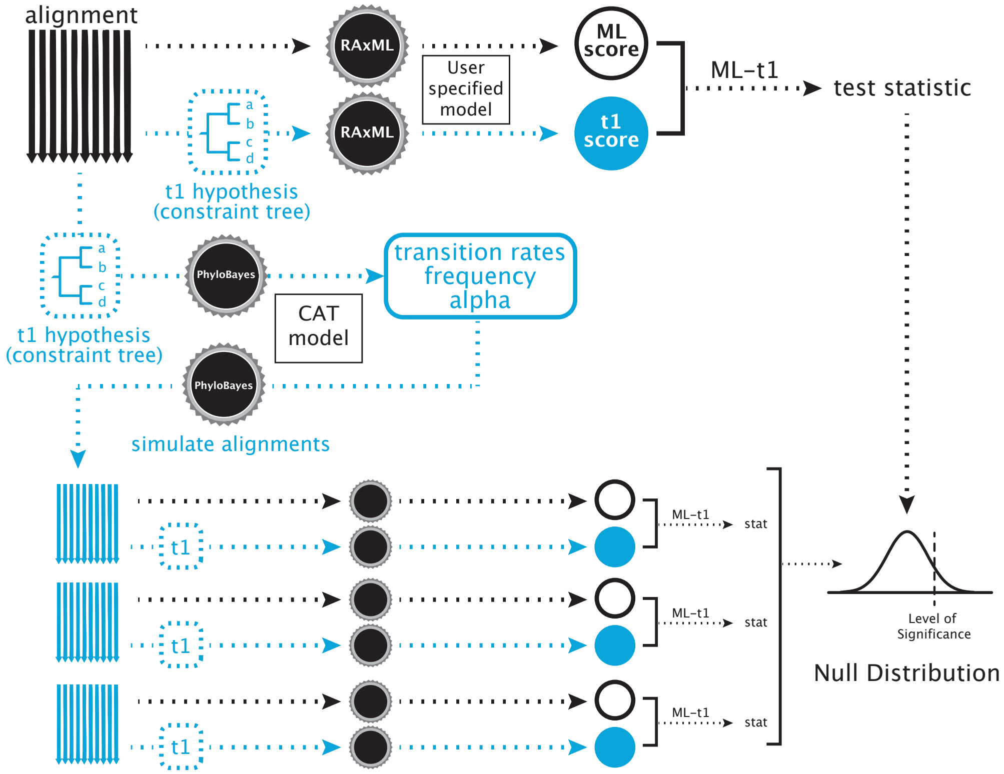

Figure 2. SOWHAT using the CAT-GTR model for parameter estimation The sample size of the distribution is the number of alignments generated by PhyloBayes based on the estimated parameters.

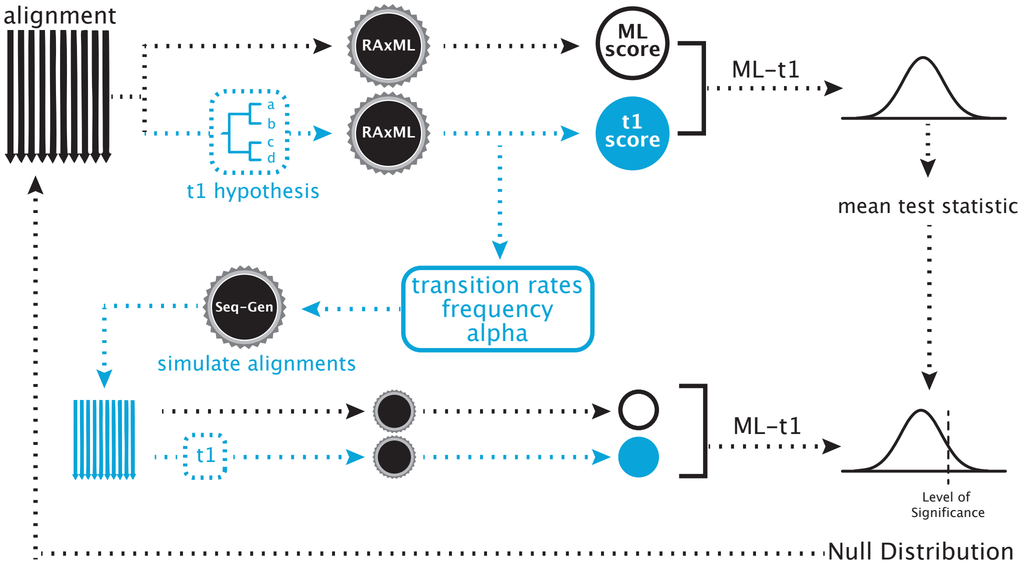

Figure 3. SOWHAT with the test statistic recalculated The entire analysis is independently repeated at each iteration and two distributions are generated. The sample size is the number of repeated analyses performed.

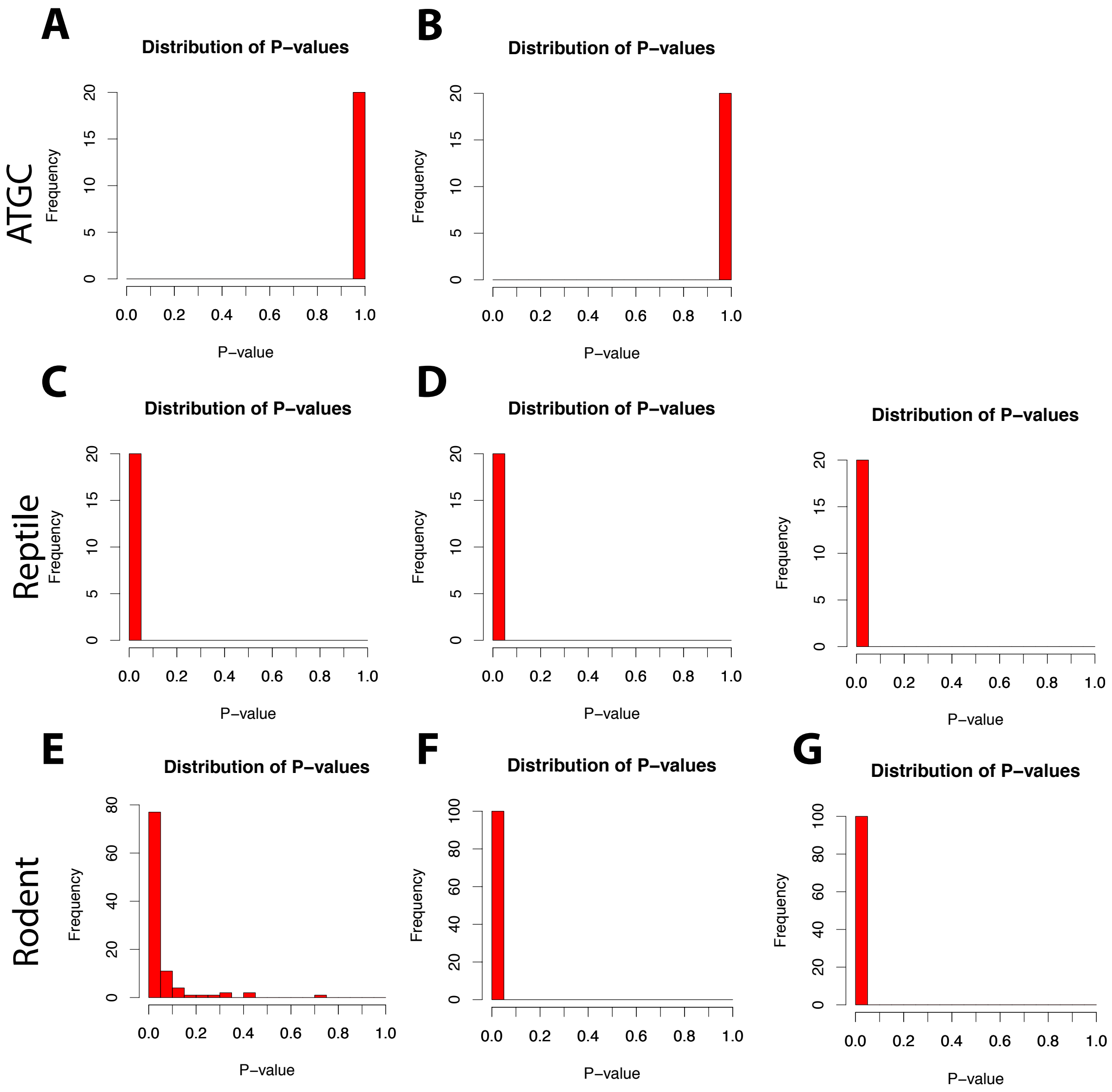

Figure 4. P-values across multiple tests at variable sample sizes A. P-values from 20 tests of the ATGC dataset using a sample size of 100. No test runs reject the hypothesis B. P-values from 20 tests using the stopping criteria. C. P-values from 20 tests of the Reptile dataset using a sample size of 100. All tests reject the null hypothesis. D. P-values of 20 tests using the stopping criteria. E. P-values from 100 tests of the Rodent dataset using a sample size of 100. The tests return inconsistent p-values. F. P-values from 100 tests using the stopping criteria. All tests consistently reject the null hypothesis. G. P-values from 100 tests with a sample size of 500. Past the stopping critieria, more sampling does not change the outcome of the test.

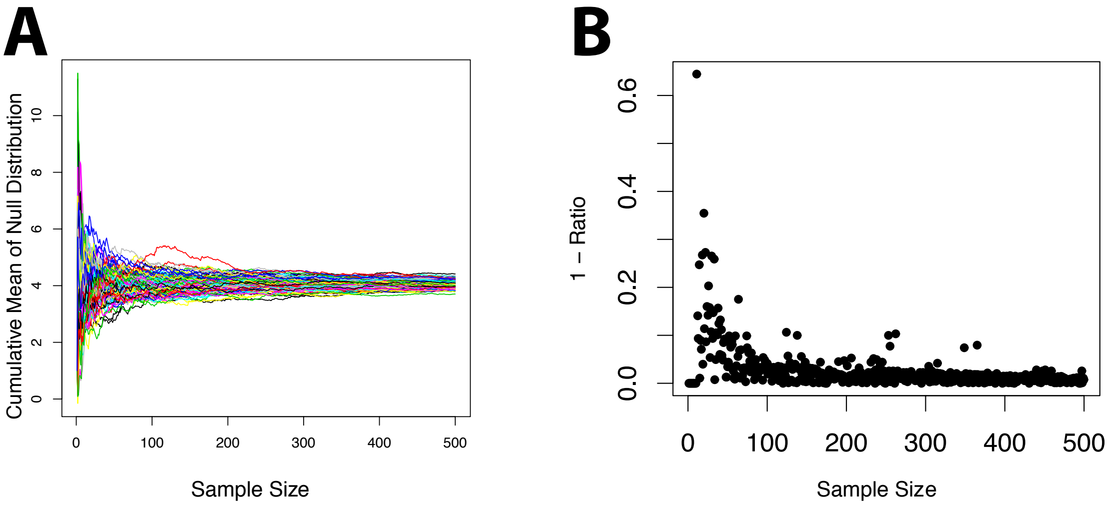

Figure 5. Convergence of tests on the Rodent dataset as sample size increases A. Cumulative mean value of the null distribution of 100 Rodent tests using the GTR+Γ model. B. 1− the ratio of z-scores of one test as the sample size is increased by 1. As the sample size increases, the null distributions of each test converge. This is reflected in the inverse of the ratio of z-scores, which is more likely to be closer to 0 as sample size is increased. At a sample size of 100, it is clear that the null distribution is not well sampled, resulting in inconsistent results across SOWH tests.

Table 1. Test Results

Test	Model (Inference, Simulation)	Number of Tests	Sample Size	P-value	Test Statistic	Mean
Rodent						
Sample 100	GTR+Γ	100	100	$4.980 * 10^{-07}$ - 0.708	4.876	3.334 - 5.098
Stopping criteria	GTR+Γ	100	437	$1.282 * 10^{-14}$ - 0.004	4.876	3.799 - 4.409
Sample 500	GTR+Γ	100	500	$1.302 * 10^{-15}$ - 0.003	4.875864	3.703 - 4.344
CAT model	GTR+Γ, CAT-GTR	1	196	$2.339 * 10^{-256}$	4.876	1.047
Recalculate	GTR+Γ	1	464	$1.030 * 10^{-12}$	4.876	3.776
ATGC						
Stopping criteria	LG	20	100	0.998 - 0.999	$-2.2 * 10^{-05}$	0.239 - 0.582
CAT model	LG, CAT-GTR	1	100	0.999	$-2.2 * 10^{-05}$	0.411
Recalculate	LG	1	100	0.999	$-2.2 * 10^{-05}$	0.387
Reptile						
Sample 100	GTR+Γ	20	100	0	175.638	0.020 - 0.120
Stopping criteria	GTR+Γ	20	183	0	175.638	0.022 - 0.092
CAT model	GTR+Γ, CAT-GTR	1	900	$1.106 * 10^{-87}$	175.638	140.008
CAT model, no partition	GTR+Γ, CAT-GTR	1	358	$1.114 * 10^{-65}$	197.210	122.350
GTR, no partition	GTR+Γ	1	140	0	197.210	-0.034
Recalculate	GTR+Γ	1	198	0	175.638	0.125
With Gaps	GTR+Γ	1	198	0	175.638	0.018

The same model was used for inference and simulation, unless otherwise noted. When more than one test was performed, a range of p-values and a range of the mean of the null distributions is shown.