

# How Much Morphological Support is Needed to Change a Phylogenomic Based Recalcitrant Node?

Rob DeSalle\*

*Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA*

## Abstract

In this paper we examine the relative contribution of information to nodes in a phylogenomic analysis combined with a morphological dataset. We examine the behavior of branch support metrics using the partitioned Bremer support or PBS. This metric measure the contribution of a data partition to a node in question and can be easily computed for likelihood (PLS) and parsimony (PBS). In addition, we use an artificial metric associated with phylogenomic matrices that is similar to branch support that we call the “flip weight”. When two competing and incongruent partitions are analyzed the flip weight is the weight of the weaker partition that results in a change in topology in a concatenated analysis. To quantitate our observations about PBS, PLS and flip weight we use a specific case of a recalcitrant node in phylogenomic analysis – the sister of all other metazoans (SOM). Specifically, we assess the ratio of PBS/PLS values of molecular to morphological support at this recalcitrant node in comparison to flip weight. We find that there is a strong correlation between the PBS/PLS ratio with the weight of the weaker partition where a flip in topology ensues. We use this correlation to calibrate the flip weight for competing partitions at a recalcitrant node.

**Keywords:** Evolution • Systematics • Phylogeny • Support • Morphology

## Introduction

Despite being developed around thirty years ago, Bremer support (BS) and partitioned Bremer support (PBS) have been underused by phylogeneticists. These metrics measure the contribution of a data partition to a node in question and can be easily computed for a given tree topology for both parsimony (partitioned branch support; PBS) and likelihood (partitioned likelihood support; PLS). One way to think about BS, PBS and PLS is that they give a researcher an idea of how much new conflicting data need to be added to change or flip the prevailing phylogenetic inference. Large values for these metrics mean that the node in question is robust while small values mean there is weak support for the node. Of course, the problem here is what is “large” and what is “small”. Currently, the BS, PBS or PLS are given as a raw number. The raw number can be scaled by dataset size or some other scaling method, but it is difficult to come up with a reasonable way to compare these metrics across datasets or to even interpret the metric in a reasonable way.

Another approach to measuring the robustness of a molecular partition relative to a morphological one in phylogenomic analysis was suggested by Neumann JS, et al. [1] and is called the “flip weight”. This artificial “tipping point” in phylogenetic analysis is the weight at which a difficult to resolve node flips from one hypothesis (that favored by molecules) to a second competing hypothesis (that favored by morphology). The flip weight tells a researcher how much weight the morphological data set needs to overturn the molecular hypothesis and can be used to characterize incongruence at nodes that are recalcitrant to straightforward biological interpretation because of a fundamental difference in the phylogenetic signal from two opposing partitions. Incongruence often exists between molecular and morphological data sets.

**\*Address for Correspondence:** Rob DeSalle, Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA; Tel: +16462269675; E-mail: [desalle@amnh.org](mailto:desalle@amnh.org)

**Copyright:** © 2022 DeSalle R. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Received:** 01 December, 2022, Manuscript No. JPGE-22-84393; **Editor assigned:** 03 December, 2022, PreQC No. P- 84393; **Reviewed:** 14 December, 2022, QC No. Q-84393; **Revised:** 19 December, 2022, Manuscript No. R-84393; **Published:** 26 December, 2022, DOI: 10.37421/2329-9002.2022.10.252

This incongruence problem was obvious when molecular data were first generated for phylogenetic problems almost thirty years ago [2-17]. One common remedy to the incongruence of molecular and morphological data sets prior to the onslaught of phylogenomics data was to combine morphology data with molecular data; with the logic that they both provide relevant information and that they might better solve difficult nodes together. Another opposed approach was to use congruence criteria to assist in making the decision to combine. More recently though a trend toward accepting the molecular phylogeny over the morphological one and then examining the character evolution of morphology on the molecular topology has been adopted [18,19]. Still, some researchers have called for a better integration of the morphological data in phylogenetic analyses (Giribet; Neumann et al.) [20,1]. In fact, Neumann JS, et al. [1] suggest that the incongruence of a morphological data set with a molecular one could be examined in more detail using what they called the “flip weight”.

This paper attempts to examine the relative contribution of information from molecular and morphological datasets using flip weight, PBS and PLS. To do this we use a recalcitrant node in a well-known phylogenetic problem (the sister of all other metazoan – SOM) as a tool to better understand the phylogenetic implications of PBS and PLS. For the current study, we assess the ratio of PBS values of molecular to morphological support at this recalcitrant node as a function of character support emanating from morphological data partitions. We find that there is a strong correlation between the PBS/PLS ratio of two partitions at odds with each other with how much weight each partition is given to reach the flip weight. We use this correlation to calibrate a “flip weight” (the weight of the conflicting partition that results in a change in topology) for competing partitions at a recalcitrant node. We suggest that this calibration is an improvement over the raw PBS and PLS metrics and gives morphological context to what the values of these metrics might mean in a phylogenetic analysis.

## Material and Methods

### Rationale

Recalcitrant nodes in phylogenetic trees are those where contradicting hypotheses receive high support creating a situation where incongruence persists at such node. Recalcitrant nodes can be “flipped” (changed) from one topology to another by the addition of data in conflict with the prevailing

inference [1]. This flipping allows one to assess the impact of adding data to a phylogenetic matrix which can be measured by the partitioned Bremer support (PBS) or partitioned likelihood support (PLS) measures.

One familiar recalcitrant node in phylogenomics concerns the sister to all other metazoans (SOM) problem where there is significant conflict between morphology vs. molecular inferences; the relationship at this node remains controversial. The recalcitrant node concerns which taxon (Porifera, Cnidaria, Ctenophora, Bilateria or Placozoa) is the sister taxon to the others. Morphological data sets generally support Porifera and in some cases Placozoa as the SOM. Phylogenomic datasets in general support Ctenophora as the SOM, but this inference depends on the kind of analysis done. Our approach here is to use the flip weight in the phylogenetic analysis of this well-known recalcitrant node – the sister of all other metazoans or SOM [1] - as a tool for assessing the phylogenetic meaning of PBS and PLS.

## Matrices

For comparing molecular and morphological datasets, we focus on the sister of all other metazoan (SOM) problem [1], which involves five ingroup taxa – Porifera, Ctenophora, Placozoa, Cnidaria and Bilateria. For molecules, we used twelve phylogenomic data sets (CH1, CH2, CH3, CH4, WH1, WH2, WH3, SI1, SI2, SI3, RYE, RYG). For morphology, we used nine datasets (BAK, BRU, GLE, ERN, SCH, ZRZ, COM, PO1, PL1).

The twelve phylogenomic datasets we use here always find Ctenophora as the SOM when using standard models in likelihood or parsimony. When alphabet reduction methods Dayhoff-6 [21], S&R-6 [22], and KGB-6 [23] and more complex likelihood CAT models are applied, Porifera is observed as the SOM for some of the datasets [24]. Here we focus on parsimony and likelihood analyses with the less complex models. Analyses of the morphological datasets yield two inferences as to the SOM – Porifera or Placozoa. Table 1 shows the data sets used and the relationships they suggest [1].

## Definition of flip weight

The flip weight for this recalcitrant node is simply the weight of the morphological partition where the inference from the analysis for the SOM flips from Ctenophora to Porifera. Finding the flip weight involves analyzing each pairwise combination of morphological and molecular data with weights of 2, 5, 10, 20, 33, 50, 76, 100, 150, 200, 333, 500, 770 and 1000. Trees from each of these weighted analyses are inspected and the flip from Ctenophora as SOM to Porifera as SOM is recorded. The flip is always in the direction of Ctenophora to Porifera because the molecular data sets initially support Ctenophora.

## Correlation of pbs with flip weight

For the comparison of PBS with flip weight we used several matrices that varied the number of taxa from six (one taxon from each ingroup and one taxon from the outgroup) to eleven (two taxa each from the five ingroup taxa and one taxon from the outgroup) to the original full taxonomic makeup of the original matrices. There are 108 (12 × 9) different pairwise combinations of the molecular and morphological matrices we examined. In addition, we applied three-character change matrices to the parsimony analyses (equal weights [EW], WAG and LG) and three different models of amino acid change in likelihood (OneST, WAG, and LG).

TreeRot.v3 Sorenson MD and Franzosa EA [25] was used to compute PBS and PLS (partitioned support for parsimony and likelihood, respectively) for the SOM node. We then used the ratio of molecular support to morphological support for both PBS and PLS as a measure of morphological partition strength. Both ratios under PBS and PLS and flipping weight were log transformed and plotted using linear regression. Since each dataset serves as an individual data point (averaging the values for that dataset) we performed two sets of regressions. The first regression used average MorphologyPBS/MolecularPBS (or Morphology PLS/molecular PLS) and the second simply plotted all MorphologyPBS/MolecularPBS (or Morphology PLS/molecular PLS) within each dataset. The same was accomplished for PLS values. In all cases we compared the raw plots to plots that constrained the regression through the origin. This is a reasonable assumption as a PBS ratio of 0.0 should coincide

with a flip of 0.0 because if there is no morphological support for the node, there will be no flip.

## Results and Discussion

Correlation of PBS and PLS ratio to flip weight: Figure 1 shows plots of the average PBS and PLS ratio vs average flip weight for the nine morphological matrices. The same plots for raw PBS and raw PLS ratio vs flip weight are shown in Supplemental Figure 1 and Supplemental Table 1. Table 1 summarizes the regression analyses for six and eleven taxon datasets.

Figure 1 and Table 1 demonstrate a correlation for the PBS and PLS ratio to the flip weight of the SOM node in the phylogenetic analysis of the SOM. There are five other important observations that accompany this correlation. First the significance of the correlation drops as the number of taxa in the analysis rises, as is further exemplified with the full dataset in Table 2. These results also show that when larger numbers of taxa are added to an analysis the correlation drops as does the slope of the regression. This means that as more taxa are added to an analysis PBS at the recalcitrant node is reduced. This pattern is more than likely due to more homoplasy being introduced to the analysis as a result of adding more taxa. Second, the correlations are higher when averaged values are used as data points for regression. Third, by constraining the regression to pass through the origin (a reasonable constraint because when PBS ratio or PLS ratios are zero, the flip point by definition should be zero) the correlations are higher. Fourth, the model of choice has little if any effect on the regression for likelihood and similarly choice of a character weighting matrix has little if any effect on the regression. Finally, likelihood appears to give reduced slopes when compared to parsimony.

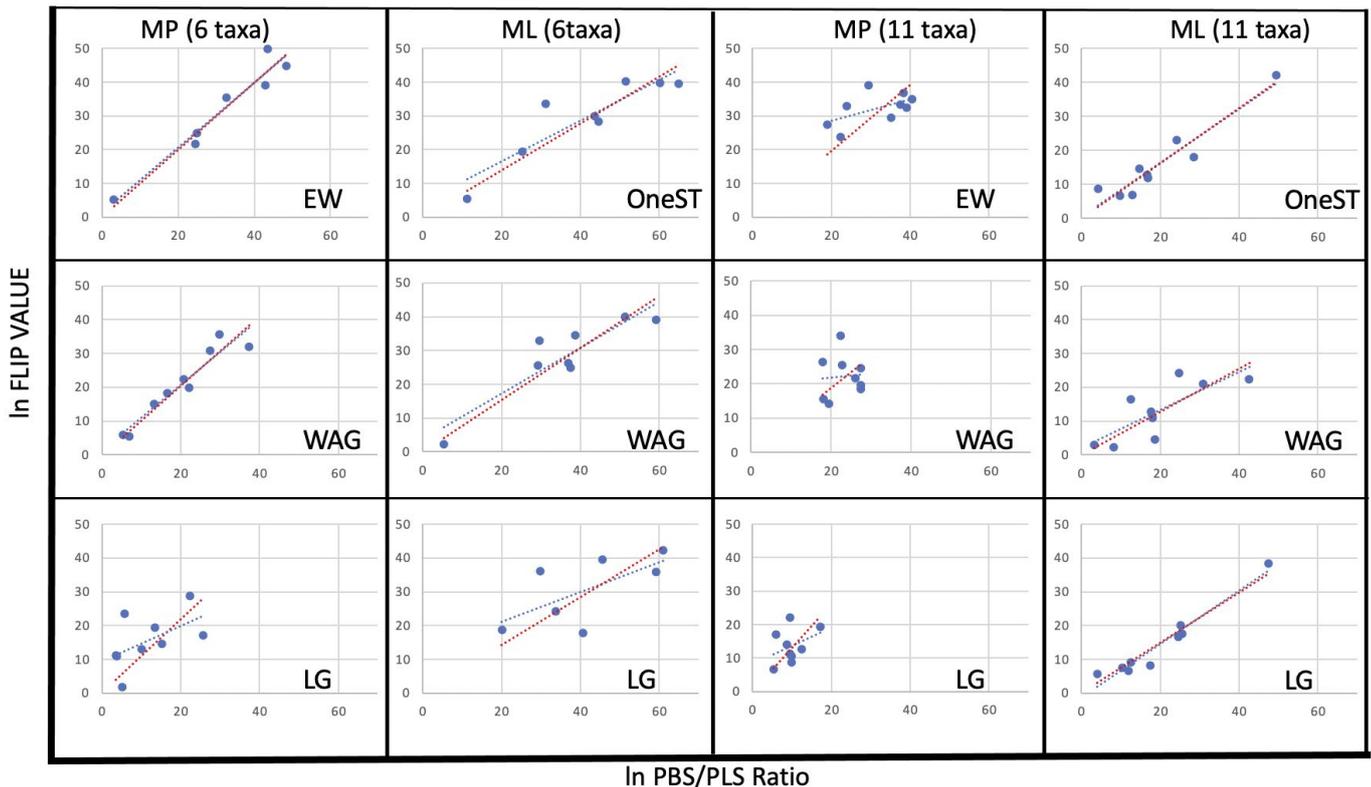
## Incongruence and character weighting

In a concatenated analysis with a recalcitrant node, increased weighting of the weaker of the two partitions relative to the stronger will result in flipping from one hypothesis to the other. In this case “weaker” simply means the inference of that weaker partition is overcome by the other partition in concatenated analysis. The reason for this seems simple. As one upweights a weaker partition that is incongruent to another stronger one, the stronger partition will have an increasingly reduced influence on the overall phylogeny. In this paper we do not advocate weighting just to get the answer the weaker partition infers, but rather offer it as a tool to understand better how character support in concatenated analysis behaves. In addition, the situation is not as simple as one might first think. When characters are combined there is a great deal of interaction of phylogenetic signal and neither partition may win [26].

Our approach of regressing the flip weight on the PBS or PLS ratio allows us to dissect further the support for a node and can be useful for examining other recalcitrant nodes in phylogenetic analysis. The slope of the regressions we show here gives the relationship between the PBS/PLS ratio and the flip

**Table 1.** Results of regression analyses for six and eleven taxa with averaged PBS and PLS ratios. The three models used for computing PLS were oneST, WAG and LG. The three weighting matrices for computing PBS were equal (EW), WAG and LG. Italic values are for no constraint through origin (other values are for analyses that were constrained through origin).

ntax	model	crit	r2	slope	r2	slope
6	EW	MP	0.94	0.96	0.98	0.99
6	WAG	MP	0.91	0.97	0.98	1.02
6	LG	MP	0.7397	0.6236	0.79	1.09
6	oneST	ML	0.81	0.6	0.97	0.69
6	WAG	ML	0.82	0.7	0.97	0.76
6	LG	ML	0.443	0.44	0.94	0.71
11	EW	MP	0.31	0.32	0.96	0.98
11	WAG	MP	0.005	0.1	0.92	0.94
11	LG	MP	0.14	0.55	0.87	1.3
11	oneST	ML	0.9	0.81	0.97	0.81
11	WAG	ML	0.62	0.64	0.89	0.64
11	LG	ML	0.98	0.74	0.98	0.74



**Figure 1.** Plots of the ratio of molecular to morphology under PBS/PLS ratio versus Flip Value for 6 taxa on the left and 11 taxa on the right. MP=maximum parsimony, ML=maximum likelihood. EW (equal weights), WAG and LG for MP are weighting matrices used. OneSt, WAG and LG are the amino acid change models used. Blue line is regression while red lines are regressions constrained through the origin. R2 and slopes are given in Table 1.

weight. In the case of origin-constrained correlations with averaged regressions for PBS the slope averages 1.0 (range 0.94 to 1.09) and for PLS averages 0.73 (range 0.64 to 0.81). The slopes for unaveraged values are similar and indicate that the flip weight of a data set with parsimony has a one-to-one correlation with the PBS ratio (molecular to morphological). So, for instance, if the ratio of morphological to molecular PBS is 8, the flip weight for morphological characters will be 8. Likewise, a PBS ratio of 50 would interpolate to a flip weight for morphological characters of 50. For likelihood, PLS will be 0.7 of the flip weight so that a PLS ratio of 8 would mean a flip weight for morphological characters of 11.4; a PLS of 50 would mean the flip weight for morphological characters would be 70. For larger data sets the flip weight will increase. Table 2 suggests that the flip weight for larger data sets might be as much as twice what it would be for smaller taxa number. In general, these analyses indicate that small PBS and PLS ratios of molecular to morphological support are correlated with flipping ease, supporting the results of Neumann JS, et al. [27].

### Likelihood analyses are harder to flip than parsimony

Another observation we make is that the same data set behaves differently under likelihood vs. parsimony criteria with respect to flip weight. Specifically, it will take more conflicting morphological information to flip a node for likelihood than for parsimony. This holds regardless of character set size of the two kinds of partitions. However, if there are many more molecular characters than morphological then the PBS and PLS ratios will be higher and hence the flip weight larger. This is because the models applied in likelihood are better at correcting for molecular vagaries than the transformation matrices used in parsimony. In other words, the inferences made using likelihood models are stronger than the parsimony inferences, and this strength requires more morphological weight to flip from one hypothesis to the alternative.

### Using flip weight and PBS/PLS to explore node robustness

There are several ways to assess the robustness of phylogenomic inferences. The most common are the bootstrap and Bayesian posteriors. Problems have been pointed out with these measures which were developed

**Table 2.** Regression stats for average PBS ratio versus average flip value for full data matrices. *Italic values* are stats for no constraint through origin and other values are for analyses constrained through origin. The three weighting matrices for computing PBS were equal (EW), WAG and LG.

matrix	<i>r2</i>	slope	<i>r2</i>	slope
EW	0.44	0.39	0.98	0.85
LG	0.41	0.26	0.97	0.48
WAG	0.36	0.3	0.97	0.61

early in the molecular phylogenetics data surge and worked well for small molecular data sets. Specifically, the bootstrap and Bayesian posteriors tend to over inflate support [28-35]. These problems are most evident in the context of modern phylogenomics where orders of magnitude more character data are available to the systematist for analysis.

It should be noted that all of the molecular matrices we use in this analysis result in 100% bootstrap values supporting Ctenophora as SOM. For the opposing morphological matrices, the bootstrap values are also high and support Placozoa (supporting matrices are SCH, BAK and PL1) or Porifera (supporting matrices are BRU, EER, GLE, ZRZ, COM and PO1). In most cases where Bayesian analysis was performed on the molecular matrices the posterior probability for the SOM node is 1.0. These support values indicate strong support for alternate hypotheses and suggest an impasse between molecular and morphological characters in these datasets. However, such bootstrap support and Bayesian posteriors may not discriminate between the relative strength of the two partitions. Comparison of the PBS or PLS ratio and the flipping weight can render more precision to the relative support each analysis will have on an inference as we discuss below.

Narechania A, et al. [36] developed a similar way to evaluate the strength of an inference with the RADICAL algorithm, which can also tease apart different levels of support in data sets when bootstrap is 100% and posterior probabilities are 1.0 at all nodes. While the RADICAL approach can add precision to understanding a node of interest and the support that

nodes accrue from different partitions, it is a relative measure and conveys its precision comparatively. In this sense it is hard to understand the biological meaning of differences detected by RADICAL. The PBS/PLS and flip weight can give some biological meaning to cases where the bootstrap and posterior probabilities are undifferentiable. Specifically, using this approach we can make statements about how much novel support for a hypothesis is needed to flip an alternative one. Since modern phylogenomic inferences will be based on more molecular characters than morphological and sometime with no morphological characters at all, the approach we describe here will tell researchers how morphological data might influence a phylogenomic based hypothesis for recalcitrant nodes.

It is therefore important that congruence measures, robustness metrics (bootstraps, BS, PBS, PLS) and posterior probabilities be better understood by phylogeneticists.

## Acknowledgement

We thank the Korein Foundation and the AMNH Institute for Comparative Genomics for continued support.

## References

- Neumann, Johannes S., Rob Desalle, Apurva Narechania and Bernd Schierwater, et al. "Morphological characters can strongly influence early animal relationships inferred from phylogenomic data sets." *Syst Biol* 70 (2021): 360-375.
- Kluge, Arnold G. "A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes)." *Syst Biol* 38 (1989): 7-25.
- Hillis, David M. "Molecular vs. morphological approaches to systematics." *Annu Rev Ecol Evol Syst* (1987): 23-42.
- Baker, Richard H., Xiaobo Yu and Rob DeSalle. "Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees." *Mol Phyl Evol* 9 (1998): 427-436.
- Nixon, Kevin C., and James M. Carpenter. "On simultaneous analysis." *Cladistics* 12 (1996): 221-241.
- Lee, Michael SY. "Uninformative characters and apparent conflict between molecules and morphology." *Mol Biol Evol* 18 (2001): 676-680.
- Donoghue, Michael J. and Michael J. Sanderson. "The suitability of molecular and morphological evidence in reconstructing plant phylogeny." In *Molecular systematics of plants*, Springer (1992): 340-368.
- Eernisse, Douglas J. and Arnold G. Kluge. "Taxonomic congruence vs. total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology." *Mol Biol Evol* 10 (1993): 1170-1195.
- Larson, Allan. "The comparison of morphological and molecular data in phylogenetic systematics." *Molecular approaches to ecology and evolution*, Springer (1998): 275-296.
- Desalle, Rob and Andrew VZ Brower. "Inferring relationships among closely related hawaiian drosophila from multiple gene regions." *Syst Biol* 46 (1997): 751-764.
- Baker, Richard H. and Rob DeSalle. "Multiple sources of character information and the phylogeny of Hawaiian drosophilids." *Syst Biol* 46 (1997): 654-673.
- Cannatella, David C. and David M. Hillis. "Amphibian relationships: Phylogenetic analysis of morphology and molecules." *Herpetol Monogr* 7 (1993): 1-7.
- Patterson, Colin, David M. Williams and Christopher J. Humphries. "Congruence between molecular and morphological phylogenies." *Annu Rev Ecol Evol Syst* 24 (1993): 153-188.
- Shaffer, H. Bradley, James M. Clark and Fred Kraus. "When molecules and morphology clash: A phylogenetic analysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae)." *Syst Biol* 40 (1991): 284-303.
- Doyle, Jeff J. "Gene trees and species trees: Molecular systematics as one-character taxonomy." *Syst Bot* (1992): 144-163.
- Doyle, Jeff J. "Trees within trees: Genes and species, molecules and morphology." *Syst Biol* 46 (1997): 537-553.
- Page, Timothy J., Satish C. Choy and Jane M. Hughes. "The taxonomic feedback loop: Symbiosis of morphology and molecules." *Biol Lett* 1 (2005): 139-142.
- Lee, Michael SY and Alessandro Palci. "Morphological phylogenetics in the genomic age." *Curr Biol* 25 (2015): R922-R929.
- Wanninger, Andreas. "Morphology is dead—long live morphology! Integrating MorphoEvoDevo into molecular EvoDevo and phylogenomics." *Front Ecol Evol* 3 (2015): 54.
- Giribet, Gonzalo. "A new dimension in combining data? The use of morphology and phylogenomic data in metazoan systematics." *Acta Zoologica* 91 (2010): 11-19.
- Dayhoff M.O., Schwartz R.M. and Orcutt B.C. "A model of evolutionary change in proteins. In: DayhoffMO, editor. Atlas of protein sequence and structure." NBRF (1978): 345-352.
- Susko, Edward and Andrew J. Roger. "On reduced amino acid alphabets for phylogenetic inference." *Mol Biol Evol* 24 (2007): 2139-2150.
- Kosiol, Carolin, Nick Goldman and Nigel H. Buttimore. "A new criterion and method for amino acid classification." *J Theor Biol* 228 (2004): 97-106.
- Simion, Paul, Hervé Philippe, Denis Baurain and Muriel Jager, et al. "A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals." *Curr Biol* 27 (2017): 958-967.
- Sorenson, M. D., and E. A. Franzosa. "TreeRot, version 3. Boston University, Boston, MA." (2007).
- Barrett, Martin, Michael J. Donoghue and Elliott Sober. "Against consensus." *Syst Zool* 40 (1991): 486-493.
- Neumann, Johannes S., Michael Tessler, Rob DeSalle and Bernd Schierwater. "Modern Invertebrate Systematics: The Phylogenetics of Early Metazoa." In *Invertebrate Zoology*, CRC Press (2021): 65-80.
- Siddall, Mark E. "Unringing a bell: Metazoan phylogenomics and the partition bootstrap." *Cladistics* 26 (2010): 444-452.
- Fernández, Rosa, Gustavo Hormiga and Gonzalo Giribet. "Phylogenomic analysis of spiders reveals nonmonophyly of orb weavers." *Curr Biol* 24 (2014): 1772-1777.
- Salichos, Leonidas and Antonis Rokas. "Inferring ancient divergences requires genes with strong phylogenetic signals." *Nature* 497 (2013): 327-331.
- Kvist, Sebastian and Mark E. Siddall. "Phylogenomics of Annelida revisited: A cladistic approach using genome-wide expressed sequence tag data mining and examining the effects of missing data." *Cladistics* 29 (2013): 435-448.
- Karbstein, Kevin, Salvatore Tomasello, Ladislav Hodač and Franz G. Dunkel, et al. "Phylogenomics supported by geometric morphometrics reveals delimitation of sexual species within the polyploid apomictic *Ranunculus auricomus* complex (*Ranunculaceae*)." *Taxon* 69 (2020): 1191-1220.
- Liu, Liang, Christian Anderson, Dennis Pearl and Scott V. Edwards. "Modern phylogenomics: Building phylogenetic trees using the multispecies coalescent model." *Methods Mol Biol* (2019): 211-239.
- Shen, Xing-Xing, Chris Todd Hittinger and Antonis Rokas. "Contentious relationships in phylogenomic studies can be driven by a handful of genes." *Nat Ecol Evol* 1 (2017): 1-10.
- Suzuki, Yoshiyuki, Galina V. Glazko and Masatoshi Nei. "Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics." *Proc Natl Acad Sci* 99 (2002): 16138-16143.
- Narechania, Apurva, Richard H. Baker, Ryan Sit and Sergios-Orestis Kolokotronis, et al. "Random addition concatenation analysis: A novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria." *Genome Biol Evol* 4 (2012): 30-43.

**How to cite this article:** DeSalle, Rob. "How Much Morphological Support is Needed to Change a Phylogenomic Based Recalcitrant Node?." *J Phylogenetics Evol Biol* 10 (2022): 252.