Malacopedia

São Paulo, SP, Brazil
Volume 9(4): 31–45

ISSN 2595-9913
Aug/2026

Reflections

Is there something phenetic in Bayesian phylogenetic analyses? A critical reflection based on personal experience

Luiz Ricardo L. Simone

Museu de Zoologia da Universidade de São Paulo
lrsimone@usp.br; lrlsimone@gmail.com
OrcID: 0000-0002-1397-9823

Abstract

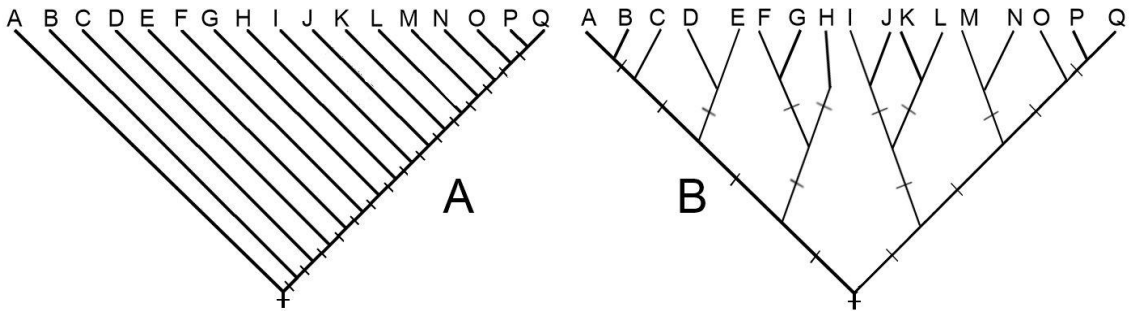
Phylogenetic studies based on Bayesian and maximum-likelihood methodologies often produce trees that resemble phenetic dendrograms. This essay briefly discusses this resemblance, concluding that neither method is phenetic, although both incorporate measures of similarity at certain stages of their analyses, which may influence the final results. This discussion, although widely based on an extensive bibliography and decades of phylogenetic practice, is ultimately based on subjective premises; its main intention is to stimulate further studies on the subject. The replacement of parsimony by Bayesian approaches is also discussed, together with the respective advantages and limitations of these methods. Greater attention should be given to the assumptions underlying Bayesian analyses, including the specification of prior distributions and evolutionary models, in order to minimize the risk of biased results. Finally, the importance of morphological approaches, particularly in Malacology, is emphasized as an independent source of evidence for comparison with molecular phylogenies, helping to identify incongruences and reducing the risk of uncritical or dogmatic acceptance of particular phylogenetic hypotheses.

Keywords: phylogeny, methodologies, theoretical biology, evolution.

Introduction

For someone that lived personally all phases of modern phylogenetic approaches, particularly applied to Zoology, it is possible to realize that certain methods are cyclic, i.e., come and go, intercalating with other methods. Interestingly, like everything in Science, every cycle is a step forward, improving and refining the previous level of knowledge. I followed the development of the Phenetics, and its more euphemistic approach, the Numerical Taxonomy (e.g., Sokal & Sneath,

1963). Gradually it was replaced by the Hennigian Cladistic Analysis (Hennig, 1966), afterwards termed simply by "Phylogenetics."



1. Comparison between models of dendrograms usually obtained by cladistic analysis using parsimony algorithm (A), and obtained by phenetic or by Bayesian/ML algorithms (B). A, "stair"-like dendrogram with successive aligned dichotomies; B, "antenna"-like dendrogram with not aligned dichotomies since its base. Both dendrograms are obviously exaggerations for didactic purposes. See text for more details.

The replacement of the phenetic philosophy by the cladistic one is not necessary to be debated here, as there is a vast literature on this, especially about the advantages of the cladistic applied to the understanding of the Taxonomy and to Evolutionary Biology (e.g., Mayr, 1982; Kitching et al., 1998). It is interesting, however, to observe that the resulted dendrograms obtained by the molecular approaches, especially when analyzing larger groups, have some resemblance to the phenetic results of the past. Despite I have not done any empirical experiments (something to be done), it is possible to see that final dendrograms obtained in the classic Cladistic analyses usually look like "stairs", i.e., they show aligned successive dichotomies along their axis (Fig. 1A). The dendrograms obtained by phenetic analyses, on the other hand, look different. They usually look like "antennas," in which the dichotomies are not aligned, and are successive in each branch (Fig. 1B). Figs 1A and 1B are schematic representations of what usually are seen in papers based on cladistic analysis with parsimony algorithms (Fig. 1A), while the other is what usually are seen in paper based on phenetic algorithms. These differences can be partially explained because phenetic also clusters paraphyletic or even polyphyletic branches. This is something that cladistic methodology avoids. It is recognized, however, that the appearance of the resulting dendrograms is not a decisive factor, as they can be modified or even disguised. The correct procedure, as mentioned above, would be to perform an empirical experiment comparing the two methods. However, a paper based on such experiments would be more appropriately published in a technical journal. *Malacopedia*, as a journal focused on the dissemination of information and on reflections concerning biological issues, also provides space for conceptual discussions and reasoned personal interpretations of biological activities. The present discussion draws on an extensive bibliography and decades of experience in phylogenetic practice, but it is intended primarily as a critical and exploratory assessment rather than as a definitive conclusion. Its main purpose is to stimulate further investigation into the possible phenetic influences underlying Bayesian and ML phylogenetic approaches.

It is not surprising that similar results to shown in Fig. 1B are usually obtained in large molecular approaches, mainly those focused on representatives of high taxa. Interestingly, many molecular studies that infer trees solely from sequence similarity—without carefully evaluating homology, alignment quality, or evolutionary models—have occasionally been criticized as drifting toward a kind of modern phenetics. It is known that this criticism is not universally accepted, because most contemporary molecular phylogenetics uses explicit evolutionary models rather than simple similarity measures. Nevertheless, when sequence distance alone dominates the analysis, the resemblance to the old numerical-taxonomy philosophy or phenetics becomes apparent.

This paper is devoted to this issue, being an alert to the necessity for biologists to understand the methods and algorithms they are applied. Also, in groups like Mollusca, in which the morphological background is not that well-established like are, for example, in Vertebrata and in Arthropoda, fortuitous incongruent molecular result cannot be detected, nor debated. This problem gets worse with the also commented dogmatic great difficulty in publishing morphological-based phylogenies.

For that, a brief explanation on Bayesian inference (BI) and maximum likelihood (ML) approaches is reported. This is done always from a phylogenetic and zoological perspective, as these approaches are widely used in other scientific branches, but with different finalities. Widely, they are applied for simplifying complex or messy set of data. As the influence of BI and ML is relatively similar in the results of phylogenetic analyses, both approaches are usually treated together in this paper. Although BI is sometimes reported alone, the discussion generally applies to both methods. However, it is recognized that BI and ML employ some different approaches, as discussed below. Another important point is that morphological datasets are predominantly analyzed using parsimony in the current literature, whereas, as mentioned above, molecular datasets are mostly analyzed using Bayesian inference (BI) and maximum likelihood (ML). Thus, in the present context, these approaches are grouped as morphology–parsimony versus molecular model-dependent approaches (ML and BI). However, obviously this distinction is not absolute, as some studies have applied the same analytical approaches to both types of datasets (e.g., Lewis, 2001; Nylander et al., 2004).

Bayesian methods

Bayesian methods belong to the field of Statistics. In general, Bayesian methods differ from classical (frequentist) statistics in that they estimate the probability of hypotheses given the observed data, rather than, in a simplistic way, the probability of the data assuming a hypothesis is true. They combine:

- Prior information (what is already known or assumed),
- The observed data (the likelihood), and
- The posterior probability, which is the updated estimate after considering the evidence.

The issue is not easy for biologists. The fundamental equation is Bayes' theorem is:

$$P(A|B) = \frac{(P(B|A)\,P(A)}{P(B)}$$

where:

- $P(A \mid B)$= posterior probability: the probability of hypothesis A after observing evidence B.
- $P(B \mid A)$= likelihood: the probability of observing the evidence B if hypothesis A is true.
- $P(A)$= prior probability: the probability of hypothesis A before considering the evidence.
- $P(B)$= marginal probability (or evidence): the overall probability of observing the evidence B.

In biological sciences, this approach is widely used in:

- Phylogenetics (e.g., Bayesian inference of evolutionary trees using programs such as MrBayes or BEAST),
- Species delimitation,
- Population genetics,
- Ecology,
- Medical diagnosis,
- Machine learning, and many other fields

In the field of the Phylogeny or Evolutive Biology, the method is termed Bayesian phylogenetic analysis, where posterior probabilities are used to evaluate the support for clades. This has additional inferences and implications:

- the underlying mathematics,
- Markov chain Monte Carlo (MCMC),
- interpretation of posterior probabilities,
- choosing evolutionary models,
- convergence diagnostics (ESS, burn-in, etc.),
- and the advantages and limitations of Bayesian inference compared with maximum likelihood and parsimony.

However, Bayesian phylogenetics has been criticized in some contexts—especially when the evolutionary model is unrealistic or when the prior assumptions strongly influence the results (Huelsenbeck et al., 1966; Lemmon & Moriarty, 2004; Zwickl & Holder, 2005; Yang, 2008; Nascimento et al., 2017; Yang & Zhu, 2018). However, the question arises: are phenetic algorithms of any kind of phenetic approach included in the software that are used to produce their dendrograms? As said above, this can be inferred by their results, but is there any evidence of that? Also, how to make a "prior probability" or a model if the study is still going on?

There are studies showing that Bayesian phylogenetic inference exhibits systematic biases. However, they are not explicitly framing these biases as *phenetic* bias. The literature almost always discusses them in terms of model misspecification, branch-length priors, or long-branch attraction rather than phenetic similarity (e.g., Kolaczkowski & Thornton, 2009).

The reason is that Bayesian inference maximizes the posterior probability of a tree under an evolutionary model (Zwickl & Holder, 2004). Those models generally explain the observed data through overall statistical similarity among taxa. The algorithm does not identify synapomorphies in the Hennigian sense; instead, it evaluates the probability that the observed distribution of character states would arise on a given tree.

This distinction is subtle but important, as follows:

In morphology-based cladistics:

- each character is treated as an independent hypothesis of homology;
- the evidence comes primarily from shared derived states (synapomorphies);
- overall similarity is not, in principle, the criterion.

In Bayesian inference:

- every character contributes to the likelihood;
- the contribution depends on the probability model;
- characters that simply make two taxa statistically more similar may increase the likelihood that those taxa are closely related, unless the model adequately discounts convergence or reversals.

Thus, it is possible to deduce that Bayesian inference is not phenetics. However, it could be argued that it may possess a phenetic tendency, because statistical similarity is translated into phylogenetic probability through the model. "Phenetic tendency," in fact, means that the degree of similarity among characters may influence the final analysis, something that analyses based on the parsimony algorithm tend to avoid. This tendency is not necessarily evident in any particular step of the algorithmic calculation, but it is an aspect that at least should be better understood by users applying non-parsimony approaches.

This becomes especially apparent, e.g., in the famous work by Kolaczkowski & Thornton (2009). They demonstrated that Bayesian inference can favor trees grouping long branches together—even under the correct substitution model—because integrating over branch-length uncertainty effectively increases the posterior probability of those topologies. They interpreted this as a consequence of Bayesian integration, not as phenetics, but the practical outcome is that taxa sharing many similar states through convergence are grouped together with high posterior support (Kolaczkowski & Thornton, 2009; Yang & Zhu, 2018). That paper, however, highlighted an important difference between Bayesian inference and maximum likelihood: the integration over parameters and, particularly, over branch lengths. That paper also highlights the roles of Bayesian integration and branch lengths. Yang & Zhu (2018), on the other hand, discuss the issue in terms of model misspecification and overconfidence. Although neither paper explicitly addresses a possible phenetic bias, as discussed above, however, this bias may also be considered within this broader context.

"The likelihood," in a broader sense, is a function of the parameters or model conditional on the observed data. However, in the present context, it is sometimes interpreted through the simplified frequentist formulation that it represents "the probability of the data assuming a hypothesis is true." This raises a particular problem for non-parsimony analyses: how can we know that the hypothesis is "true"? If the "true" phylogenetic pathway were already known, there would be no need to perform the analysis in the first place.

From the perspective of phylogenetic analysis, a different question could be asked:

Does Bayesian inference systematically reward overall character-state similarity, regardless of whether that similarity is homologous or homoplastic?

A paper that asks the question in exactly these terms is not present in literature so far. In fact, this framing looks poorly addressed.

There is also an interesting mathematical reason why this may happen. Parsimony minimizes the number of transformations on a tree. Bayesian analysis maximizes. Bayesian inference, however, seeks to characterize a posterior distribution, typically using MCMC (Markov chain Monte Carlo), a statistical method used to explore the space of possible evolutionary trees and estimate their posterior probabilities. This approach can also yield a MAP (maximum a posteriori) tree, although the underlying calculation is complex and beyond the scope of the present paper.

The following expression is the form most commonly used in Bayesian phylogenetics:

$$P(\text{Tree/Data}) \propto P(\text{Data/Tree})\, P(\text{Tree}),$$

The symbol ∝ means "is proportional to." It is used because the denominator of Bayes' theorem, $P(\text{Data})$, is omitted. Since $P(\text{Data})$ is the same for all candidate trees, it acts only as a normalizing constant. It is the overall probability of observing the data, considering all possible trees. Because this denominator is identical for every candidate tree, Bayesian software usually ignores it during computation and later rescales all probabilities.

P(Tree|Data) means Posterior probability. This is the probability that a particular phylogenetic tree is correct after the molecular (or morphological) data have been analyzed. This is the quantity that Bayesian phylogenetic analysis seeks to estimate.

P(Data|Tree) Likelihood. This is the probability of obtaining the observed data assuming that the proposed tree is correct (together with the chosen evolutionary model).

Notice the direction:

- It is not the probability that the tree is correct.
- It is a probability that the observed data would occur if that tree were true.

This distinction is fundamental.

P(Tree) Prior probability. This is the probability assigned to the tree before examining the data. It represents previous knowledge or assumptions about the tree.

Often in phylogenetic analyses, all trees are assigned equal prior probabilities ("uniform priors"), although priors can also be based on biological information, fossil evidence, divergence times, and so on. Any researcher, however, must be aware that if the previous knowledge is uncertain, incomplete, or derived from the same general evidence being investigated, the choice of prior can influence the result.

However, it is recognized that phylogenetic likelihood does not operate simply by asking, “Which organisms are most similar?” Rather, it calculates the probability of the observed data conditional on a specific tree, its branch lengths, and a model of evolution. Possible ancestral states, transition probabilities, branch lengths, and model parameters all contribute to the calculation. This is not merely a terminological distinction. For example, in Felsenstein’s (1981) classic formulation of phylogenetic likelihood, the probability of the observed data conditional on the tree and

the substitution process is evaluated, erecting parsimony as misleading. The calculation of phylogenetic likelihood itself involves no step in which a measure of "overall similarity" among terminal taxa is used as a grouping criterion. Therefore, sharing a particular state does not, in itself, imply that two taxa should receive greater support as sister taxa. The model, however, is necessarily an idealization of the evolutionary process and may introduce biases that can influence the final result.

If two taxa share many identical character states, the likelihood often increases because fewer improbable changes need to be invoked, even when those similarities result from convergence. The evolutionary model attempts to correct this, but it cannot eliminate the effect completely. Therefore, Bayesian inference is not using phenetic distances explicitly, yet it may still favor trees reflecting overall resemblance whenever the model underestimates homoplasy.

In this scenario, the following distinction is important:

- Phenetics measures similarity directly.
- Bayesian inference measures the probability of observing the similarities under an evolutionary model.

Those are different concepts, but they need not produce different outcomes. Going to non-Bayesian inferences, mainly considering "ground plans" (Simone, 2026), and morphologic-based phylogenies (most my studies), there is another question that may be even more interesting:

If one replaces terminal species by reconstructed ground plans (which greatly reduce autapomorphies and random homoplastic noise), does Bayesian inference become less "phenetic" because the character matrix contains proportionally more true synapomorphies? This is an interesting issue for additional considerations (see Simone, 2026, argumentations) but looks impracticable in molecular-based phylogenies.

These approaches look absent in literature. It could be an original theoretical contribution, and it would even be testable by simulation or by comparing analyses of species-level matrices with equivalent ground-plan matrices. If the Bayesian trees converge toward the parsimony trees after replacing species by ground plans, that would provide evidence that much of the apparent disagreement originates from similarity noise rather than from the inference algorithm itself.

Why use Bayesian methods instead of parsimony?

This is one of the most fundamental questions in modern phylogenetics. A possible explanation is that Bayesian and Maximum Likelihood are theoretically better suited for molecular sequences because DNA evolution can be modeled probabilistically. Since nucleotides have only four states (A, C, G, T), one can formulate explicit models (JC69, HKY, GTR, etc.) that estimate the probability of one nucleotide changing into another along branches of a tree; i.e., they specify a probabilistic model of nucleotide substitution from which transition probabilities along branches can be calculated (Jukes & Cantor, 1969; Kimura, 1980; Felsenstein, 1981; Yang, 1994). Parsimony ignores those probabilities and simply minimizes the number of changes.

However, also an important historical and sociological component that is rarely discussed should be evoked: Why did molecular systematists abandon parsimony?

During the 1980s and 1990s, molecular datasets became enormous compared with morphological ones. Thousands or millions of nucleotides were suddenly available. Researchers quickly noticed that parsimony sometimes failed under conditions such as:

- long-branch attraction,
- highly unequal substitution rates,
- saturation,
- unequal base frequencies.

Likelihood methods were adapted to phylogenetic analyses precisely to address those problems. Once Bayesian methods (particularly after the introduction of *MrBayes* around 2001) became computationally feasible, they offered several attractive features:

- branch support (posterior probabilities),
- incorporation of complex substitution models,
- relaxed molecular clocks,
- partition-specific models,
- integration over parameter uncertainty.

Consequently, Bayesian inference became the "gold standard." However, the question regarding why parsimony remains unjustified in some scenarios, since both methods can easily be performed using the same computational apparatus.

After all, if Bayesian inference is truly superior, why do Bayesian and parsimony trees often differ substantially? One possibility is that only Bayesian methods are genuinely recovering the correct history. But another possibility is that they are, in fact, recovering the history predicted by the chosen substitution model instead.

Both assertions are not necessarily the same thing, as a model is an approximation of reality. Every Bayesian analysis is therefore conditional on assumptions about:

- substitution rates,
- stationarity,
- independence among sites,
- rate distributions, among many other parameters.

If those assumptions are incorrect, or at least some of them, the "optimal" Bayesian tree is biased, simply the best tree under an incorrect model. Parsimony avoids this issue because it has virtually no evolutionary model. In fact, the Parsimony analysis results in a model.

Of course, there is no perfect method, and that simplicity is often presented as a weakness. But philosophically, it can also be viewed as a strength. Asymmetry is always an intriguing issue. However, the problem is that this is not faced in a debatable, equality scenario, as anything in

Science should be. When Bayesian and parsimony inferences disagree, Bayesian researchers usually conclude: "Parsimony failed." Almost nobody asks the converse: "Did the model fail?" Possibly because most biologists do not understand what is going on.

Thus, there is a subtle asymmetry here. The Bayesian result is often treated as the benchmark against which parsimony is judged. Rarely Bayesian inference itself is treated as the hypothesis requiring validation. That is the way in which Science progresses. Anyway, there are indications in the phylogenetic literature that, when maximum likelihood analysis is based on a no-common-mechanism model, ML and maximum parsimony select the same tree, or the same set of optimal trees (Tuffley & Steel, 1997). This demonstrates how parsimony can contribute to model building, as well as how model-dependent analyses may converge with parsimony when the model is weak.

Another issue that must be analyzed is that molecular matrices are extraordinarily redundant. Suppose a protein contains hundreds of amino acids. A mutation affecting one functional domain may produce dozens of correlated nucleotide substitutions. The Bayesian model generally treats each nucleotide site as an independent observation (except for the dependence incorporated by the substitution model). Morphologists, on the other hand, would probably regard those correlated substitutions as manifestations of a single evolutionary transformation. Thus, molecular datasets often contain enormous numbers of partially or totally correlated characters, which can be actually useless or biased in a comparative scenario.

Parsimony and Bayesian methods handle this redundancy differently, but neither explicitly models the biological hierarchy of transformations. That may explain some of their disagreement. Parsimony, on the other hand, generates observable synapomorphies, which can be analyzed and indirectly indicate hierarchy of transformations.

In the context, does Bayesian inference really reflect a broader shift in biology? Hennig's framework was built around the concept of evidence—synapomorphies as evidence of common ancestry. Modern molecular phylogenetics, on the other hand, is increasingly built around the concept of statistical estimation. Those are not identical epistemologies:

Parsimony “asks”: "Which characters diagnose a clade?"

While Bayesian “asks”: "Which tree is most probable under this stochastic model?"

Both are legitimate scientific questions, but they represent different philosophies. For someone (like me) who has spent decades analyzing morphology character by character, that distinction is especially relevant. It also suggests a fruitful line of inquiry: whether the move from parsimony to Bayesian inference represents not merely a change in computational methodology, but a deeper change in what systematists regard as phylogenetic evidence. That is a philosophical question as much as a technical one, which, at the very least, deserves more attention than it has received.

In contrast, conducting a morphological study requires specialized knowledge, extensive training, methodological refinement, and intellectually demanding procedures, such as identifying homologous characters and defining their respective character states, etc. By comparison, Bayesian and ML analyses are almost entirely computational, allowing even relatively inexperienced researchers to perform them with readily available software. This is not necessarily a drawback in

itself, but it deserves careful consideration, particularly in light of the current, nearly worldwide abandonment of non-molecular approaches in comparative biology (Malacology particularly).

The need for extensive training is often regarded as a weakness of Morphology. However, once a professional has acquired this expertise, there is no need to avoid them to personally conduct every phylogenetic study. Rather, their expertise can be incorporated into such studies, with morphological evidence being given an appropriate weight.

Why study morphology in comparative biology?

In comparative biology like Phylogenetics, it is widely recognized that conducting a morphological or anatomical study differs fundamentally from investigating the same group using molecular approaches. Two distinct aspects must be considered: (1) the difficulty of generating analyzable data and (2) the difficulty of analyzing those data. In non-molecular studies, both aspects are particularly challenging, as computational tools can provide only limited assistance throughout much of the process.

A good morphological matrix requires:

- years of anatomical and taxonomic experience;
- the ability to recognize homologous structures;
- knowledge of ontogeny, function, and variation;
- careful decisions about character delimitation and state definition;
- repeated refinement of hypotheses of homology.

Those are intellectual activities that cannot be automated. In fact, character discovery is arguably the most creative step in the entire phylogenetic process.

By contrast, for molecular data, once DNA has been sequenced, much of the pipeline is standardized:

- align sequences;
- select a substitution model;
- run IQ-TREE, RAxML, MrBayes, or BEAST;
- obtain a tree.

This certainly lowers the barrier to producing a phylogenetic tree. However, it is recognized that not everything in Bayesian analysis is computational. A good Bayesian analysis is not that trivial. Someone who understands it well must make informed choices about:

- alignment quality;
- partitioning strategy;
- substitution models;
- priors;

- MCMC settings;
- convergence diagnostics;
- interpretation of posterior probabilities.

However, any published Bayesian analyses are probably performed without fully understanding these issues, and several available software also meets these demands or addresses the lack of knowledge. Most probably, many researchers today can execute Bayesian software without understanding either:

- the evolutionary assumptions embedded in the model, or
- the biological meaning of the resulting tree.

Running, for example, MrBayes or POY5 is relatively straightforward. Understanding what MrBayes or POY5 are actually estimating is considerably more difficult. Likewise, generating DNA sequence data has become routine, whereas determining whether those sequences provide an adequate proxy for organismal evolution is a far deeper question. This distinction is important because a lack of experience or insufficient knowledge of the group under study often becomes evident in the interpretation of the results presented in published papers. Nevertheless, the resulting dendrogram is frequently accepted without question, particularly in groups such as Mollusca, where, as discussed above, a robust morphological framework has yet to be established. The cladograms are merely presented, and the discussion is usually focused on the resulting relationships. However, little or no attention is given to the visible (i.e., morphological) synapomorphies supporting each branch or to how the results differ from those of a morphology-based analysis.

This is unfair scenario. A poor morphologist usually cannot produce a convincing morphological matrix and a final result. Experts quickly recognize poorly defined characters, ambiguous homologies, or inconsistent coding. By contrast, a poor molecular systematist can often produce a technically correct Bayesian analysis simply by following a standard workflow. The software protects the user from many computational mistakes. This creates an asymmetry: Morphological research is expertise-limited; molecular phylogenetics is increasingly technology-limited: once the laboratory and computational infrastructure are available, the analysis can often proceed with relatively little anatomical or organismal knowledge. In other words, both approaches distinguish the intellectual demands of data production from the accessibility of analytical tools.

Extending this discussion to a broader perspective, the center of expertise has gradually shifted from the naturalist to the data analyst. Hennig, like most classical systematists, regarded phylogenetics as an extension of comparative biology. In contrast, much of contemporary molecular phylogenetics treats it as an exercise in statistical inference based on sequence data. These approaches are not mutually exclusive, but they cultivate different forms of expertise, and this shift has undoubtedly influenced the culture of systematics over the past three decades (Simone, 2025). At least in Malacology, however, the marked asymmetry between these approaches has become a matter of concern, as its consequences have not always been beneficial.

Do Bayesian and parsimony algorithms result in the same phylogenetic trees?

During the 1990s and early 2000s, it was common practice to analyze the same molecular dataset using parsimony, maximum likelihood, and Bayesian inference. The comparisons among these approaches proved to be highly informative. Over time, however, this practice gradually declined, with Bayesian inference becoming the predominant analytical method.

When available, the obtained dendrograms generally fall into three categories.

1. Most of the tree is identical. This is actually a rare outcome. In these cases, deeply supported clades are recovered by all methods because the phylogenetic signal is strong. The disagreements usually occur only at poorly supported nodes. This observation is often used to argue that algorithms are converging on the same evolutionary history.

2. Certain clades consistently differ. Some parts of the tree repeatedly change depending on the algorithm. These are often characterized by:

- long branches,
- rapidly evolving genes,
- unequal evolutionary rates,
- saturated sequences.

This is where Bayesian and ML frequently agree with each other but disagree with parsimony. Supporters of Bayesian inference interpret this as evidence that parsimony is statistically inconsistent. Supporters of parsimony counter that the probabilistic model may be imposing an incorrect expectation on the data. Notice that the same observation can be interpreted in two completely different ways.

3. Large datasets reduce—but do not eliminate—differences. One might expect millions of nucleotides to make all methods converge. Surprisingly, they often do not. Even phylogenomic datasets with thousands of genes can produce conflicting trees depending on:

- the evolutionary model,
- partitioning,
- filtering,
- treatment of missing data, or the inference method itself.

Interestingly, more data does not necessarily overcome methodological assumptions. Sometimes they even amplify them. In a large molecular dataset, we can infer that each approach:

Parsimony asks: "What tree minimizes the total number of changes?"

Bayesian inference asks: "What tree makes this enormous pattern of nucleotides most probable under my evolutionary model?"

Those are fundamentally different optimization criteria. If the substitution model accurately captures the evolutionary process, Bayesian inference should, in theory, have an advantage. If the model is substantially wrong, however, increasing the amount of data does not necessarily move the result closer to the true tree. Instead, the researcher may become increasingly confident

in the best tree under that incorrect model. Statisticians sometimes refer to this as becoming "precisely wrong."

Conclusion

What can be deduced from the discussion above, summarized here for conciseness, is that the algorithms most commonly used in molecular phylogenetic studies—Bayesian inference and maximum likelihood (ML)—are, strictly speaking, not phenetic methods. However, because they incorporate measures of similarity among character states at some stages of their analyses, elements of phenetic inference may be present in their results. This is a feature that parsimony, in principle, avoids.

Another point is that both Bayesian inference and ML depend on explicit evolutionary models. These models, which are fundamental to the functioning of both methods, can theoretically be misspecified and may therefore influence the final result. The researcher's tendency to favor the grouping or separation of a particular set of taxa, thereby influencing whether they are inferred as monophyletic, may also introduce bias. Such influence is also possible in parsimony analyses, but in this case the source of the bias is more explicit.

Additionally, the difficulty of publishing purely morphological phylogenetic studies nowadays may have a dogmatic effect, particularly in taxa such as Mollusca, in which a well-developed morphological counterpart to molecular approaches is not always available. The extensive training required to conduct a morphological phylogenetic analysis may also encourage researchers to favor Bayesian approaches, which can be implemented using molecular datasets and established analytical pipelines. On the other hand, the need for extensive training in morphology may also contribute to the robustness and depth of the resulting phylogenetic hypotheses, as it requires a detailed understanding of the organisms and of the homology and variation of their characters. At the very least, this allows the synapomorphies supporting each branch of the cladogram to be analyzed.

It is relatively easy to find examples in which early molecular hypotheses were subsequently abandoned (e.g., Serialia; Wilson et al., 2010). It is much harder to find cases in which a molecular hypothesis was rejected *solely because it disagreed with classical morphological evidence*. In modern systematics, disagreement with morphology is generally regarded as a reason to seek additional evidence rather than, by itself, as sufficient grounds for rejecting the molecular result. Typically, a molecular hypothesis is set aside only after independent evidence—such as additional genes, improved taxon sampling, more appropriate evolutionary models, or methodological analyses—indicates that the original result was likely an artifact. This distinction is important because it reflects how competing lines of evidence are generally evaluated in contemporary phylogenetics.

In this context, it is possible that many Bayesian analyses are conducted by researchers who apply these methods without fully understanding their underlying philosophy, assumptions, and potential effects on the final result. Moreover, when a comparable morphological counterpart is not available at the same level of analytical development—at least in Mollusca—the resulting molecular hypothesis may remain largely unquestioned. The absence of comparable alternative approaches and critical debate may therefore contribute to the acceptance of dogmatic or potentially biased results.

Although this essay is not based on an empirical experiment, it nevertheless highlights the need for such studies to be conducted. It is also hoped that it will encourage colleagues who use Bayesian or ML methodology to develop a deeper understanding of how the method works and to perform parsimony analyses on the same datasets. Additionally, phylogenies resulting from Bayesian analyses should be compared with equivalent hypotheses based on morphological data. In cases of incongruence, particular attention should be given to the synapomorphies supporting each branch of the morphological cladogram, in order to investigate why these morphological hypotheses differ from the corresponding branches of the Bayesian phylogeny.

Acknowledgements

I am very grateful to the anonymous referee from Ribeirão Preto, who provided several valuable insights and helped improve various aspects of the present argument. This referee remained, however, concerned about the lack of empirical demonstrations supporting several of the assertions made here. This is indeed an important point, which led me to emphasize throughout the paper that some of the arguments are based on personal experience and are intended primarily to stimulate the empirical studies that are still needed to test them.

References

Felsenstein, J, 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17: 368–376. DOI: 10.1007/BF01734359.

Hennig, H, 1966. Phylogenetic systematics. Translated by Davis, D & Zangert, R (1999). University of Illinois Press. Urbana, 263 pp.

Huelsenbeck, JP; Larget, B & Alfaro, ME, 2004. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. Molecular Biology and Evolution 21: 1123–1133.

Jukes, TH & Cantor, CR, 1969. Evolution of protein molecules. IN: Munro, HN [ed.]. Mammalian Protein Metabolism 3: 21–132.

Kimura, M, 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111–120. DOI: 10.1007/BF01731581.

Kitching, IJ; Forey, PL; Humphries, CJ & Williams, DM, 1998. Clasistics. The theory and practice of parsimony analysis (second edition). Oxford University Press. Oxford, 240 pp.

Lemmon, AR & Moriarty, EC, 2004. The importance of proper model assumption in Bayesian phylogenetics. Systematic Biology 53(2): 278–298.

Lokaczkowski, B & Thornton, JW, 2009. Long-branch attraction bias and inconsistency in Bayesian phylogenetics. Plos One 4(12): e7891 https://doi.org/10.1371/journal.pone.0007891.

Mayr, E, 1982. The growth of biological thought. Harvard University Press. Cambridge, 974 pp.

Nascimento, FF; Reis, M & Yang, Z, 2017. A biologist's guide to Bayesian phylogenetic analysis. Nature Ecology & Evolution 1: 1446–1454.

Nylander, JAA; Ronquist, F; Huelsenbeck, JP & Nieves-A., JL, 2004. Bayesian Phylogenetic Analysis of Combined Data. Systematic Biology 53(1): 47-67. https://doi.org/10.1080/10635150490264699

Simone, LRL, 2025. The decline of Ethics in Taxonomy. Malacopedia 8(6): 32-43. http://moluscos.org/trabalhos/Malacopedia/08-06Simone%202025%20ethics.pdf

Simone, LRL, 2026. Rigorous ground plans as terminal taxa: a demonstration in Gastropoda (Mollusca), focusing on Caenogastropoda. PeerJ 14: e21150 https://doi.org/10.7717/peerj.21150.

Sokal, RH & Sneath, HA, 1963. Principles of numerical taxonomy. W. H. Freeman and Company. San Francisco, 359 pp.

Tuffley, C & Steel, M, 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bulletin of Mathematical Biology 59: 581–607. https://doi.org/10.1007/BF02459467.

Wilson, NG; Rouse, GW & Giribet, G, 2010. Assessing the molluscan hypothesis Serialia (Monoplacophora + Polyplacophora) using novel molecular data. Molecular Phylogenetics and Evolution 54(1): 187-193. https://doi.org/10.1016/j.ympev.2009.07.028.

Yang, Z, 1994. Estimating the pattern of nucleotide substitution. Journal of Molecular Evolution 39: 105–111. DOI: 10.1007/BF00178256.

Yang, Z, 2008. Empirical evaluation of a prior for Bayesian phylogenetic inference. Philosophical Transactions of the Royal Society B 363: 4031–4039. DOI: 10.1098/rstb.2008.0164.

Yang, Z & Zhu, T, 2018. Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees. Proceedings of the National Academy of Sciences 115(8) 1854–1859.

Zwickl, DJ & Holder, MT, 2005. Model parameterization, prior distributions, and the general time-reversible model in Bayesian phylogenetics. Systematic Biology 54(6): 877–888. doi: 10.1080/10635150490522584.