Quantitative Phylogenetic Analysis in the 21st Century

Análisis Filogenéticos Cuantitativos en el siglo XXI

Daniel R. Brooks*[1], Jaret Bilewitch[1], Charmaine Condy[1], David C. Evans[2], Kaila E. Folinsbee[2], Jörg Fröbisch[2], Dominik Halas[1], Stephanie Hill[2], Deborah A. McLennan[1], Michelle Mattern[1], Linda A. Tsuji[2], Jessica L. Ward[1], Niklas Wahlberg[3], David Zamparo[1], and David Zanatta[1]

[1] Department of Ecology and Evolutionary Biology. University of Toronto, Toronto, Canada
[2] Department of Biology, University of Toronto-Mississauga, Mississauga, Canada
[3] Department of Zoology, Stockholm University, Stockholm, Sweden and Department of Biology, Laboratory of Genetics, University of Turku, Finland
*Correspondent: dbrooks@zoo.utoronto.ca

Abstract. We review Hennigian phylogenetics and compare it with Maximum parsimony, Maximum likelihood, and Bayesian likelihood approaches. All methods use the principle of parsimony in some form. Hennigian-based approaches are justified ontologically by the Darwinian concepts of phylogenetic conservatism and cohesion of homologies, embodied in Hennig's Auxiliary Principle, and applied by outgroup comparisons. Parsimony is used as an epistemological tool, applied a posteriori to choose the most robust hypothesis when there are conflicting data. Quantitative methods use parsimony as an ontological criterion: Maximum parsimony analysis uses unweighted parsimony, Maximum likelihood weight all characters equally that explain the data, and Bayesian likelihood relying on weighting each character partition that explains the data. Different results most often stem from insufficient data, in which case each quantitative method treats ambiguities differently. All quantitative methods produce networks. The networks can be converted into trees by rooting them. If the rooting is done in accordance with Hennig's Auxiliary Principle, using outgroup comparisons, the resulting tree can then be interpreted as a phylogenetic hypothesis. As the size of the data set increases, likelihood methods select models that allow an increasingly greater number of a priori possibilities, converging on the Hennigian perspective that nothing is prohibited a priori. Thus, all methods produce similar results, regardless of data type, especially when their networks are rooted using outgroups. Appeals to Popperian philosophy cannot justify any kind of phylogenetic analysis, because they argue from effect to cause rather than from cause to effect. Nor can particular methods be justified on the basis of statistical consistency, because all may be consistent or inconsistent depending on the data. If analyses using different types of data and/or different methods of phylogeny reconstruction do not produce the same results, more data are needed.

Key words: phylogenetics, quantitative phylogenetics, maximum likelihood, parsimony, Bayesian likelihood, Hennig, information theory, data congruence.

Resumen. Se revisa la sistemática filogenética Hennigiana y se compara con las aproximaciones de Máxima Parsimonia, Máxima Verosimilitud y verosimilitud Bayesiana. Todos los métodos utilizan el principio de la parsimonia en alguna forma. Las aproximaciones con bases Hennigianas se justifican ontológicamente con los conceptos Darwinianos de conservacionismo filogenético y cohesión de las homologías, representados en el Principio Auxiliar de Hennig, y aplicado en la comparación con el grupo externo. La Parsimonia se utiliza como una herramienta epistemológica, aplicada a posteriori en la elección de la hipótesis más robusta cuando hay datos en conflicto. Los métodos cuantitativos utilizan la parsimonia como un criterio ontológico: los análisis de Máxima Parsimonia utilizan la parsimonia sin pesaje, la Máxima Verosimilitud les asigna un peso igual a todos los caracteres que explican los datos, mientras que la verosimilitud Bayesiana depende del pesaje de cada una de las particiones de caracteres que explican los datos. Las diferencias en los resultados derivan de un muestreo insuficiente de datos, en cuyo caso cada método trata las ambigüedades de manera diferente. Todos los métodos cuantitativos producen redes. Las redes pueden convertirse en árboles al ser enraizadas. Si el enraizamiento se efectua de acuerdo con el Principio Auxiliar de Hennig, utilizando la comparación con un grupo externo,

el árbol resultante puede considerarse como una hipótesis filogenética. Al incrementarse el número de datos, los métodos de verosimilitud selccionan modelos que permiten un número cada vez mayor de posibilidades a priori, convergiendo en la perspectiva Hennigiana de que nada está prohibido a priori. Por lo tanto, todos los métodos producen resultados similares independientemente del tipo de datos, especialmente cuando las redes se enraizan utilizando grupos externos. Las invocaciones a la filosofia Popperiana no pueden justificar ningún tipo de análisis filogenético, ya que sus argumentos van del efecto a la causa y no de la causa al efecto. Tampoco se puede justificar el uso de un método en particular con base en la consistencia estadística, ya que todos pueden ser consistentes o incosistentes dependiendo de los datos. Si los análisis con diferentes tipos de datos y/o métodos de reconstrucción filogenética no producen igual resultado, significa que es necesario reunir datos adicionales.

Palabras clave: filogenia, filogenia cuantitativa, máxima verosimilitud, parsimonia, verosimilitud bayesiana, Hennig, teoría de la información, congruencia de datos.

Introduction

Phylogenetic analysis has become an essential element unifying a broad range of research programs investigating the patterns and processes of evolution. The diversity of perspectives inherent in the traditions which researchers represent, however, has led to a multitude of quantitative methods for phylogeny reconstruction, the efficiency and validity of which are fiercely debated. That debate may stem largely from the nature of phylogeny reconstruction. Unlike the physical and chemical sciences, in which explanations aim to be predictive with respect to spatio-temporally invariant laws, phylogenetic analysis deals with a singular history of events, many of which involve a degree of contingency.

We believe the fierceness of the debate about quantitative methods for phylogeny reconstruction obscures an essential point – there is considerably more agreement than disagreement in results produced by those methods. Our intention in this review is to show how such convergence of outcome could arise from such seeming divergence of methodologies and philosophies.

A Precis of Hennigian Phylogenetics
The primacy of homology

As Darwin suggested it should be, phylogenetic analysis is about analyzing characters, finding homologies and basing classifications (phylogenetic hypotheses) on them.

> [w]e are forced to trace community of descent by resemblances of any kind…we choose those characters which are least likely to have been modified, in relation to the conditions of life…
> Rudimentary structures on this view are as good as, or even better than, other parts of organization… let it be the inflection of the angle of the jaw, the manner in which

an insect's wing is folded… if it prevail throughout many and different species… it assumes high value… for we can account for its presence in so many forms with such different habits, only by inheritance from a common parent. (Darwin 1872: 403)

Almost all systematists agree that to be phylogenetically useful, a character must possess certain properties, the most fundamental of which is **that it is inherited**. Many systematists consider a phylogenetic character to be any inherited attribute (e.g., Colless 1985). Recently, Grandcolas et al. (2001) argued this point cogently. They suggested that because we assume homologies exist *a priori,* but we cannot know with certainty whether a particular complex of traits are homologues, we cannot exclude any kind of attribute, from the simplest of nucleotides to the most complex of behaviors, for any reason other than non-inheritance.

Characters must also exhibit varying degrees of **evolutionary conservatism**, such that some characters will indicate ancient phylogenetic relationships and others, more recent ones. Some traits may, in fact, be homologous despite appearing to be quite different, because they are causally related. The older, ancestral, or *plesiomorphic* (Hennig 1950), character arose in an older period of evolutionary history than the younger, descendant, or *apomorphic* (Hennig 1950) character and together, the two form a *transformation series* (Hennig 1966). Whereas Hennig (1966) used the term "character" to mean a discrete stage – a singular heritable unit – in a transformation series, contemporary workers often use the term "character-state" in lieu of Hennig's "character" and "character" for Hennig's "transformation-series" (Wagner 2001; Brooks and McLennan 2002; Richards 2002, 2003; but see Grant and Kluge 2004 for objections to this terminology).

> "Homologous parts tend to vary in the same manner, and homologous parts tend to cohere" (Darwin 1872:158)

If we are to use "co-variation and coherence" among characters as evidence of phylogeny, we must assume that what we call different characters are not merely repetitions of the same thing. They must exhibit some degree of **independence**, i.e., they are potentially capable of evolving at different rates and to different degrees in different lineages in different times and places. Some phylogeneticists suggest that because homologous traits are, in a sense, non-independent with respect to phylogeny, we cannot distinguish non-independence *a priori*. Character selection is guided by *Kluge's Auxiliary Principle* (Brooks and McLennan 2002): "always presume character independence in the absence of evidence to the contrary".

The fact that different organisms share similar features was noted long before any scientific theory of evolution was codified. One of the first attempts to codify the manner in which characters are shared amongst organisms was by Richard Owen, who coined the term *homology* in 1843. Owen's (1843) definition was structural: "the same organ in different animals under every variety of form and function". He contrasted this with *analogy*, defined as "a part or organ in one animal which has the same function as another part or organ in a different animal". This specification of homology as sameness in structure rather than function had significant impact on evolutionary thought, so that not until one hundred years later was it stated that behaviours could be considered homologues (Hubbs 1944).

Owen (1847) later expanded his definition of homology, splitting it into three types. *Special homology* was the correspondence of a part in one animal with a part in a different animal, for example, the foreleg of a lizard and that of a mouse. *General homology* was "a higher relation of homology", "that in which a part or series of parts stands to the fundamental or general type"; consider as an example one of the vertebrae of a shrew and one of the vertebrae of a whale, which are generally homologous as vertebrae. *Serial homology* was defined as the serial repetition of segments, such as the foreleg and the hind leg of a salamander.

The Darwinian revolution easily assimilated special homology; the sharing of traits among species because they were inherited from a common ancestor is a simpler concept to understand than similarity based on correspondence to an ideal type. The function of classifications thus became describing the phylogeny of life. Producing such classifications seemed straightforward – document enough homologies and you will have documented phylogeny. Owens' special homologies were indicators of phylogeny and provided an evolutionary criterion for classifying species. Initial attempts to implement this "Darwinian

Imperative" were problematical because they were tautological; the similarity of characters in different taxa was a result of the common ancestry of those taxa, but this common ancestry could only be recognized by means of the similarities. In other words, homologous characters were both defined by, and used to delineate, evolutionary relationships.

Henry Fairfield Osborn was among the first to stress the primacy of the historical aspect of homology, stating that the only decisive test of homology was historic community of derivation (Osborn 1902). Likewise, some founders of the New Synthesis supported the primacy of the historical concept of homology when they accepted Owens' special homology category, but questioned whether general and serial homologies have any applicability in the realm of phylogeny (Haas and Simpson 1946). Other neo-Darwinians, however, expressed support for pre-Darwinian homology concepts. Boyden (1947) objected to homology being defined as similarity due to common ancestry, claiming that we could not know ancestry independently of the analysis of presumptive homologies. This resurrected ahistorical approach for determining homology was codified by Remane (1956, 1961). The entry of molecular biology into evolutionary studies and systematics in the 1960s all but ignored the common descent criterion in definitions of homology; for example, Neurath et al. (1967) defined homology among proteins simply as a degree of structural similarity greater than might be expected by chance alone.

Echoing Darwin's (1872) assertion that homologies tend to vary in the same manner, **to cohere**, the *evolutionary homology criterion* (e.g., Wiley 1981; Patterson 1982, 1988; Roth 1984, 1988, 1991, 1994; Gould 1986; Rieppel 1992; McKitrick 1994) assumes that homologous traits covary with phylogeny, and also that non-homologous traits co-vary only under special circumstances. The proposal of this criterion marked the path to integrating both the special and the structural homology criteria, implying that phylogeny reconstruction could be undertaken by examining large numbers of characters, seeking to distinguish those that "varied in the same manner and cohered" from those that did not.

Wiley (1981) proposed a two-part protocol for implementing the evolutionary homology criterion. He reasoned that the relationships among species are not self-evident, but must be discovered by finding characters that are shared among species on the basis of common ancestry. In order to do this, we must first have protocols for identifying similarities that will serve as candidate markers of phylogeny. Such protocols can be considered the *discovery* criteria for homology. For this task, Wiley suggested using ahistorical criteria,

such as Remane's, to pinpoint similar traits that we then hypothesize to be homologous. Not all similarity is due to common ancestry, however, so the phylogenetic signal in the putative homologies discovered using ahistorical protocols must be assessed by constructing a phylogeny using many characters, and checking to see which ones are logically consistent with the relationships expressed in the phylogenetic tree. In this way, phylogenetic congruence among characters is the *evaluation* criterion for homology. Wiley (1981) suggested that this integration of discovery and evaluation of homologies represented an example of *reciprocal illumination* (Hennig 1966). Patterson (1988), in a comprehensive overview of definitions of homology, codified Wiley's approach. He listed three separate criteria for assessing homology: congruence (shared history), conjunction (two different character states cannot be homologues if they are found together in the same organism), and similarity. He concluded that all three criteria must be satisfied for a homology to be real.

Despite its generality, there is a drawback to the evolutionary homology criterion. If the only way to determine mistaken presumptions of homology is via the discovery that some potential homologues are not congruent with other characters, the evolutionary homology criterion is relatively weak. Changes in our estimates of phylogeny, resulting from additional data, may change previous interpretations of homologous and non-homologous characters. Despite this limitation, however, most phylogeneticists agree that avoiding circularity is paramount, justifying reliance on the relatively weak evolutionary homology criterion.

The Auxiliary Principle

The most important concept introduced by Hennig was the stipulation that we should assume homology in the absence of contradictory evidence, now known as *Hennig's Auxiliary Principle*. The idea that similarity in traits even among distantly related species was due to homology (i.e., plesiomorphy) rather than independent evolution (homoplasy) was, however, established long before the development of Hennigian systematics

> …it would in most cases be extremely rash to attribute to convergence a close and general similarity of structure in the modified descendants of widely distinct forms. The shape of a crystal is determined solely by the molecular forces and it is not surprising that dissimilar substances should sometimes assume the same form; but with organic beings we should bear in mind that the form of each depends on an infinitude of complex relations, namely on the variations that have arisen, these being due to causes far too

intricate to be followed out, --on the nature of the variations that have been preserved or selected, and this depends on the surrounding physical conditions, and in a still higher degree on the surrounding organisms with which each being has come into competition, --and lastly, on inheritance (in itself a fluctuating element) from innumerable progenitors, all of which had their forms determined through equally complex relations. It is incredible that the descendants of two organisms, which had originally differed in a marked manner, should ever afterwards converge so closely as to lead to a near approach to identity throughout their whole organisation. If this had occurred, we should meet with the same form, independent of genetic connection, recurring in widely separated geological formations; and the balance of evidence is opposed to any such admission. (Darwin 1872: 127-128).

Making the Auxiliary Principle Operational: Outgroup Comparisons

> "Mr. Waterhouse has remarked that, when a member belonging to one group of animals exhibits an affinity to a quite distinct group, this affinity in most cases is general and not special." (Darwin 1872: 409)

This statement indicates how comparisons of similar traits found in members of a group being studied (the "ingroup") and species outside that group ("outgroups") could be used to implement what we now call Hennig's Auxiliary Principle. By referring to a colleague, this passage also indicates that at least some systematists during Darwin's day used a form of this concept in their work. Ironically, there was no codification of what we now call outgroup comparison until more than a century after Darwin wrote the above passage. Engelmann and Wiley (1977) provided the rationale for outgroups in phylogenetic reconstruction, demonstrating how the same data are differentially treated in what they called "closed" versus "open" systems. Engelmann and Wiley pointed out that reference to species outside the ingroup (their open systems approach) permits a researcher to distinguish between traits that truly conflict with phylogeny (homoplasies), from those that only appear to conflict (plesiomorphies).

The application of outgroup analysis to real datasets proved problematical in early studies because even by those who claimed to use the method never treated outgroups explicitly.(Colless 1967). Stevens (1980) voiced the frustration of the day, stating that although phylogenetic systematists claimed their assumptions and

procedures were explicit, there was in fact little discussion of the crucial early step, namely assignment of character state polarity. Watrous and Wheeler (1981) were the first to suggest operational rules for outgroup comparison.

Farris (1982) objected to some of Watrous and Wheeler's general rules, arguing that they failed to recognize that direct application of parsimony both explained their rules and overcame all putative limitations in the application of outgroup analysis. This was one of the first instances in which ontological and epistemological parsimony were conflated by systematic theorists. We believe that the published exchange of opinions about the function, significance, and relevance of outgroup comparisons is the clearest manifestation of a lack of clarity between epistemological and ontological uses of parsimony in phylogeny reconstruction.

The Auxiliary Principle and the Principle of Parsimony

The principle of parsimony (Latin *parcere*, to spare), also known as the principle of simplicity, is often connected with the English philosopher and Franciscan monk William of Ockham (ca. 1285-1349), who advocated the use of the principle so forcefully that it is also known as 'Ockham's razor': "*Pluralitas non est ponenda sine neccesitate*" ("plurality should not be posited without necessity") and "*non sunt multiplicanda entia praeter necessitatem*" ("entities should not be multiplied unnecessarily"). In this sense, the principle represents an epistemological tool that obliges us to favor theories or hypotheses making the fewest unwarranted, or *ad hoc*, assumptions about the data. **This does not necessarily imply that nature itself is parsimonious, or that most parsimonious theories are true**. Aristotle (350 B.C.E.) articulated a different view of the principle of parsimony, that "nature operates in the shortest way possible" and "the more limited, if adequate, is always preferable" (Charlesworth 1956). This postulates that nature itself is parsimonious, using the principle in an ontological rather than epistemological manner. Phylogeneticists have used the term "parsimony" in both senses, resulting, in our estimation, unnecessary confusion and conflict.

Hennig clearly intended to maximize hypotheses of homology and minimize hypotheses of homoplasy, which invokes the principle of parsimony by avoiding the assumption of unnecessary *ad hoc* hypotheses of parallelism. In the Hennigian system, if evolution (or nature) were parsimonious as Aristotle suggested, all traits would be logically consistent with the true phylogeny – there would be no conflicting relationships suggested by any set of traits, that is, there would be no homoplasy. The Auxiliary Principle implies that there will often be conflicts in the data, which should be resolved in favor

of the hypothesis postulating the fewest number of assumptions of multiple origins (homoplasy) over single origins (homology).

Contemporary Hennigians assert that both the Auxiliary Principle and the use of parsimony are logical requirements of any attempt to reconstruct phylogeny; if one were to assert that all similarities were due to homoplasy, there would be no evidence of common descent, and thus no evidence of evolution. Likewise, if one is going to invoke the Auxiliary Principle, one must invoke it for all traits, thereby choosing the phylogenetic hypothesis that minimizes the total number of violations of the Auxiliary Principle for a given set of data. Wiley (1981) suggested four main assumptions of phylogenetics: (1) evolution has occurred, documented by the characters of different species; (2) each species is a historically unique mosaic of plesiomorphic, synapomorphic, and autapomorphic traits; (3) we do not have foreknowledge about which characters are homologous and homoplasious; and (4) we do not have foreknowledge of the phylogenetic relationships, or of the relative or absolute rates of divergence. The presumption of homology in Hennig's Auxiliary Principle assumes only that evolution is **conservative**, not parsimonious, and we have good empirical reasons to believe that presumption, most notably that replication rates are higher than mutation rates.

As noted above, the Auxiliary Principle is an ontological criterion, suggesting that evolution has been conservative, not necessarily parsimonious. Outgroup comparison is thus an operational tool used to satisfy the Auxiliary Principle with respect to distinguishing plesiomorphies from apomorphies. If there is no conflict among the apomorphies, there is only a single phylogenetic hypothesis supported by the data – there would be no "most parsimonious" tree, only "the" tree. When outgroup comparison does not resolve all conflicts in the data, phylogenetic analysis requires an epistemological tool to make a contingent decision about the preferred hypothesis based on empirical robustness. This tool is the principle of parsimony. The extent to which we need to implement the principle of parsimony depends on many factors (Brooks 1996), but even a cursory survey of published studies shows that no type of data is free of homoplasy.

The first algorithm to determine ingroup relationships with reference to multiple outgroups was presented by Maddison et al. (1984), who showed that the most robust outgroup comparison relied on two or more paraphyletic outgroups. They proposed a two-step procedure that first assesses character states "locally" among a number of outgroups; when there is ambiguity, parsimony is used to make a decision about the preferred plesiomorphic state. These ancestral states are used in performing Hennigian

analysis, or for rooting a network. This produces phylogenetic trees that are most parsimonious "globally", i.e. most parsimonious in the context of related groups, in the same sense that Engelman and Wiley (1977) proposed.

Quantitative Approaches

The development of quantitative methods for phylogeny reconstruction parallels the emergence and development of Hennigian phylogenetic systematics during the 1960s. Three major classes of quantitative methods differ philosophically from Hennigian principles by invoking some form of parsimony as an ontological, rather than epistemological criterion.

Maximum Parsimony

In September 1965, two seminal articles on phylogeny and parsimony appeared. Wilson (1965) introduced a "consistency test for phylogenies based on contemporaneous species." His null hypothesis was that all characters are unique and unreversed. In order to pass the consistency test, the taxa defined by these characters must be nested and these conditions must persist as new species are added to the analysis. Colless (1966) was concerned that more than one phylogenetic tree might pass the consistency test, that a character might mistakenly be regarded as unique and unreversed, and that the taxa are, in the first place, grouped solely on the basis of similarities. Wilson (1967) asserted that his consistency test was internally sound, but that he shared one of Colless' main concerns, which "is the lack of efficient methods for selecting the character states".

That concern was discussed in the second article, in which Camin and Sokal (1965) presented the first algorithm for applying the parsimony criterion to phylogenetics and first applied the term "parsimony" to a method of phylogenetic inference. They used a group of imaginary animals ("Caminalcules") possessing a number of morphological characters that could change according to particular rules. Thus, the 'true phylogenetic tree' was known and could be compared to trees that were achieved by different methodologies. Camin and Sokal found that the trees that most closely resembled the "true phylogeny" required the least number of changes in the morphological characters, which seems to invoke an epistemological use of parsimony. They claimed that their technique examined "the possibility of reconstructing cladistics by the principle of evolutionary parsimony" (Camin and Sokal 1965), but then qualified it by stating that their approach assumed nature is parsimonious, an appeal to ontological parsimony. Camin and Sokal produced a computer program implementing their method, demonstrating for

the first time that quantitative phylogenetic analysis could be operational. Their original algorithm, however, was unwieldy and inefficient for larger data sets, and never effectively programmed.

Soon afterward, Kluge and Farris (1969, also Farris 1970), presented "Wagner parsimony", named in honor of W.H. Wagner (Wagner 1952, 1961, 1969), who formalized the *groundplan divergence method* (Mitchell 1901, 1905; Tillyard 1921; Sporne 1949; Danser 1950, 1953) on which Kluge and Farris' algorithm was based. Wagner parsimony minimizes the Manhattan distance between members of a set of taxa via the creation of hypothetical taxonomic units or archetypal ancestors. One year later, Farris (1970) argued that it was not necessary to have an ancestor to begin tree construction because the choice of an ancestor could change the topology of the tree. He concluded that a rootless network would reduce the dependency of tree topology on *a priori* assumptions about the nature of the ancestor. To do this, he used a method for creating networks that minimized the length of the intervals between taxa (symbolized by nodes), using the shortest network connections (Manhattan distance) method of Prim (1957; Sokal and Sneath 1963). Farris' method differed from previous phenetic applications by the use of shared, derived characters rather than characters connected by only "similarity". The subsequent network can be converted into a phylogenetic tree by rooting it at one of the taxa within the tree, or at an interval within the network. Phylogenies constructed using this method are completed by optimizing the characters onto the tree. In the decade following Farris' (1970) contribution, a number of algorithms were developed (e.g., Fitch parsimony, Fitch 1971); Dollo parsimony, Farris 1977), which were incorporated into the existing programs as alternatives to Wagner parsimony. These algorithms differed primarily in their assumptions and restrictions regarding character evolution, and are discussed in more detail by Wiley et al. (1991).

Converting a Wagner Network into a Phylogenetic Tree

Converting a Wagner network into a phylogenetic tree requires rooting the network in some manner. Increasingly, published studies convert the network into a tree by rooting it with an arbitrarily chosen single taxon not included in the group being analyzed. This protocol should not be mistaken for the method of outgroup comparison that emerged in phylogenetics during the 1970s. The distinction is slight, but significant, and must be understood in light of Hennig's perspective on the issue of ancestors. Hennig objected strongly to the notion that phylogeny reconstruction could be achieved by reconstructing a series of archetypal ancestors from which particular descendant

species could be derived. He argued that each species was a unique mosaic of plesiomorphic and apomorphic traits so archetypes, defined as ancestral species exhibiting only plesiomorphic traits, did not exist. In other words, no single taxon could be used as an outgroup to determine the plesiomorphic and apomorphic traits for any analysis. Given this, rooting a network with a single outgroup taxon would be sufficiently robust in the Hennigian system only if that taxon were the archetypal ancestor of the ingroup, something the Hennigian system disavows.

As can be seen from the above discussion, the early development of the Wagner algorithm was not informed directly by Hennigian reasoning. Rather, it relied on the groundplan divergence method based on *a priori* recognition of an archetypal ancestor. When Farris (1970) abandoned the *a priori* reliance on an ancestor, the Wagner algorithm became a method for producing an unrooted network. Lundberg (1972) linked the results of Wagner analyses with Hennigian analyses by differentiating ancestors from outgroups. He proposed that the structure of a network might make certain character states more likely to be ancestral, helping to determine which interval should form the root of the tree of a parsimony-based network. The shift in emphasis from searching for ancestors to identifying outgroups was critical in linking Wagner with Hennig.

Farris explicitly proposed that parsimony, rather than the Auxiliary Principle, be considered the ontological criterion for phylogeny reconstruction. Parsimony analysis, however, produces a network, which can only be converted into a tree by rooting it. Any network has many possible roots, all producing equally parsimonious trees, so parsimony cannot serve as an ontological criterion for phylogenetic analysis. Using outgroup comparisons, as the corollary of the Auxiliary Principle, to root Wagner networks produces high degrees of consistency between Wagner algorithm, groundplan-divergence, and Hennigian analyses of the same data (Churchill et al., 1984), returning us to the Hennigian perspective that the Auxiliary Principle is the ontological principle, and parsimony is an epistemological complement to it.

Character Weighting – Segue to Model-Based Phylogenetic Analysis

Not all phylogeneticists believe that robust phylogeny reconstruction can be achieved solely through the application of Hennigian principles. If the evolutionary homology criterion (EHC) is violated to such an extent that the number of co-varying homoplasies equals the number of homologies, Hennigian phylogenetics produces ambiguous results, in the form of multiple most parsimonious trees (MPTs). If co-varying homoplasies

outnumber homologies, Hennigian phylogenetics produces an unambiguous, yet incorrect, result. An array of "character weighting" protocols for giving some characters more significance than others have been formulated in an effort to compensate for presumptive cases in which the EHC is violated.

A posteriori weighting methods are alternatives to consensus trees and bootstrapping/ jackknifing for reducing ambiguity caused by homoplasy, testing for phylogenetic signal in different characters, or selecting a preferred tree from among multiple MPTs. The underlying assumption is that the researcher does not know a priori which characters are likely to exhibit co-varying homoplasy, but once those traits have been identified by non-weighted phylogenetic analysis, their influence on determination of the preferred phylogenetic hypothesis can be minimized. Farris (1969) provided the first numerical algorithmic approach to character weighting with the successive approximations algorithm for character weighting (SAW), developed from the concept of "cladistic reliability", defined as the fit between a character and the phylogeny (Farris 1969). The most parsimonious tree(s) derived by standard phylogenetic analysis (Farris's 'estimated tree') become(s) the foundation for subsequent parsimony analysis. The consistency index (or the rescaled consistency index, the retention index, or the best consistency index: Quicke, 1993) of each character is determined; when there are MPTs, consistency indices are averaged over the set of trees. Each character is then reweighted by multiplying it by the index value, and a new phylogenetic analysis performed until two successive iterations remain the same. While most phylogeneticists use tree topology to obtain initial weights, Farris (1969) suggested using a modified compatibility technique of LeQuesne (1969) that recodes multi-state characters using additive binary coding to negate biasing. Farris tested his method by inputting a hypothetical phylogenetic tree (Farris's 'true tree') with 31 nodes and 30 completely consistent characters. Inconsistent (homoplasious) characters were then assigned to nodes by a random number generator and the characters successively weighted. Comparing his resulting trees to the 'true tree', Farris concluded that the algorithm almost always improves 'true tree' estimates. Four types of functions (concave bounded, concave unbounded, linear, and convex) relating weight to the probability of character change were tested, of which the unbounded concave weight function was the most effective. Carpenter (1988, 1994) expanded Farris's rationale, stating that successive approximations weighting is meant to allow characters in a data set to judge themselves in terms of their cladistic reliability. The intention is to downweight less reliable characters.

Various authors have criticized successive approximations weighting. Felsenstein (2004) pointed out that the successive weighting methodmakes it is difficult to detect ties. The method will not select between two equally parsimonious solutions and, therefore, will not always result in one tree. Kluge (1998a) suggested that independence is lost if the consistency of other characters determines the inconsistency of the down-weighted characters and that, paradoxically, character independence can be retained only if weights were applied arbitrarily (Kluge 1998a). Goloboff (1993, 1995) argued that Farris's method uses pooled data to determine the fits of characters to trees and when trees are compared during a search, the implications of character reliability from a tree found in a previous analysis will affect the search. Swofford and Olsen (1990; also Cunningham 1997) argued that successive approximations weighting is circular as it always increases support for one or more of the trees produced by the initial phylogenetic analysis. Carpenter (1994) countered that they confused circularity with recursion, and that Farris's (1969) simulation analysis showed that the final tree might not be one of the original MPTs.

Goloboff's (1993, 1995) *implied weighting* is a non-iterative method for weighting characters according to their reliability. It aims to maximize the 'total fit' of characters to a tree – among all possible trees, the tree that implies higher weights is assumed to be maximally reliable. The fit for each character is determined independently from other characters and the total fit is the sum of the fits of the individual characters. Goloboff (1993) pointed out that his method would not necessarily produce most parsimonious trees, but would produce maximally reliable (self consistent) trees, claiming that self consistency is a "necessary but not a sufficient condition" for reconstructing evolutionary history.

A priori weighting methods give differential weights to characters prior to phylogenetic analysis. That molecular data should be an important source of phylogenetic information was recognized more than 30 years ago (e.g., Neyman 1971; Sokal and Sneath 1973). Today, the ease with which large amounts of nucleotide sequence data can be collected makes them very attractive. The very simplicity of DNA/RNA, however, results in high levels of "built-in" homoplasy (Brooks, 1996), and consequently, several *a priori* weighting methods have been introduced specifically to compensate for the peculiarities of nucleotide sequence data.

Nucleotide substitutions occur via transitions and transversions. Transitions are substitutions between pyrimidines (C – T), or between purines (A – G); they occur with little cost and are more common. Transversions are substitutions between pyrimidines and purines; they

are costly and less likely to occur (Brown et al. 1982). This has led some authors to postulate that transitions are more likely to reach saturation and become less phylogenetically reliable than transversions (Broughton et al. 2000). Transitions are thus frequently down-weighted relative to transversions (Hickson et al. 1996; Milinkovitch et al. 1996; Murphy et al. 2002). In the most extreme version, transversion parsimony (Swofford et al. 1996), transitions receive zero weight. Murphy et al. (2002) found that a tri-fold weighting of transversions versus transitions provided greater resolution of rattlesnake data, but other authors have found transitions more useful than transversions; consequently, differential weighting of transitions versus transversions as a general procedure may be unwarranted (Kraus and Miyamoto 1991; Reeder 1995; Kallersjo et al. 1998, 1999; Broughton et al. 2000; Simmons et al. 2006).

Analyses of protein coding genes may employ *codon position weighting* (Björklund 1999; Sennbald and Bremer 2000). A protein-coding gene is structurally divided into codons, each composed of three base pairsencrypting eitheran amino acid or a stop message. The functional position of each base pair (1st, 2nd, and 3rd) is proportional to the impact it has on the amino acid for which it codes, which corresponds to the probability of the base pair changing. Second codon positions have historically been given greater weight in phylogenetic analyses because they evolve slowly. 3rd codon positions, which have less impact on the amino acids for which they code and thus are free to evolve at much higher rates , are often down-weighted or excluded in phylogenetic analyses because they are less likely to have a favourable signal to noise ratio (Björklund 1999). Commonly, the three codon positions are weighted inversely to their variability (i.e., $2^{nd} > 1^{st} > 3^{rd}$) (Björklund, 1999), although several studies have challenged the general usage of *a priori* differential weighting based on generalized assumptions of character state evolution (Björklund, 1999; Sennbald and Bremer, 2000). For example, Murphy et al. (2002) found that the transformations in the third position were phylogenetically more informative compared with 1st and 2nd codon positions. Then Murphy (2004) demonstrated that extreme weighting of the 2nd position drastically changed the hypothesis of relationships in the data set. If this effect is general, it occurs because changes in the 2nd position affect the functioning of the encoded proteins so only those changes that produce functional proteins will survive and, since this is likely to be a small subset of all possible substitutions, the chance of homoplasy is increased.

When secondary structures of RNA sequences are analyzed as part of phylogenetic analysis, stems and loops are often differentially weighted (Hickson et al., 1996). Stems, double stranded regions sustaining a

greater number of compensatory mutations because of complementary base pairing, often violate the assumption that a change in one nucleotide does not affect the probability of change in another (Dixon and Hillis, 1993). This lack of independence has prompted some authors (for example, Wheeler and Honeycutt, 1988) to argue that stem nucleotides are not a meaningful source of phylogenetic information and to recommend either eliminating nucleotides associated with stem regions or down-weighting them by one-half, assuming that loop regions carry twice as much weight. Other authors believe that both stem and loop characters are phylogenetically informative and recommend down-weighting stem characters by no more than 20% relative to single-stranded loop characters (Dixon and Hillis, 1993). To make matters more complicated, loop regions can undergo frequent base pair substitutions because these changes have little or no consequence to the secondary structure. Consequently, loops often experience "transition saturation" resulting in problematical alignments. Following Hennig's Auxiliary Principle, difficulty in aligning loop regions results in uncertainty concerning the homologous nature of base pairs; thus, it is common practice to exclude these regions from analysis if alignment is not possible (Gatesy et al., 1993; Leache and Reeder, 2002; Hertwig et al., 2004). In other words, arguments can be made to down-weight both loops(Tang et al., 1999) and stems. Not surprisingly then, Hickson et al. (1996) observed patterns of conservation and variability in both stem and loop regions in their analysis of mitochondrial sequences in small subunit rRNA and concluded that differential weighting these regions would prove unsatisfactory.

Character weighting remains a controversial topic in phylogenetics. Some researchers argue that all weighting lacks objective criteria for choosing which characters to weight and how much to weight them (Allard and Carpenter, 1996; Vidal and Lecointre, 1998; Allard et al., 1999). Others argue that *a priori* weighting of nucleotide sequence data inevitably discards evidence because general assumptions (e.g., rates of evolution) do not apply in every case and specific assumptions cannot be generalized (Farris, 1983; Carpenter, 1992; Wheeler, 1992). It is also unclear how character weighting affects the character independence that is essential for quantitative phylogenetic analysis (see discussion above).

The proliferation of multiple approaches to character weighting, none of which has become generally accepted, was not directly responsible for the emergence of model-based methods of phylogenetic analysis. That emergence was more subtle, and was based on suspicions about the nature of particular data that prompted thoughts about weighting. Those suspicions provided fertile ground for the growth and development of model-based methods, which we discuss next.

Maximum Likelihood
A Precis of Maximum Likelihood in Phylogenetics

Edwards and Cavalli-Sforza (1963) explored the idea that likelihood could be applied to phylogeny reconstruction using blood-group allele frequency data in human populations (Edwards and Cavalli-Sforza, 1964; Cavalli-Sforza and Edwards, 1967). They called their approach 'Method of Minimum Evolution', but the original algorithm did not work because it was based on the assumption that evolution has been parsimonious (Edwards, 1996). Neyman (1971) applied likelihood analysis to nucleotide sequences, and presciently suggested that this approach might become important in the future. Farris (1973) and Felsenstein (1973) published likelihood algorithms for phylogeny reconstruction; however, problems of computational difficulties limited practical applications. Felsenstein (1981) introduced the first computationally efficient maximum likelihood algorithm for discrete character nucleotide sequence data. Since then, maximum likelihood methods have become increasingly popular in phylogenetic studies (Swofford et al., 1996; Huelsenbeck and Crandall, 1997; Tuffley and Steel, 1997; Felsenstein, 2004). These approaches are most commonly used in molecular phylogenetics (Swofford et al., 1996; Huelsenbeck and Crandall, 1997; Huelsenbeck et al., 2002; Ronquist, 2004), but morphology-based likelihood methods have been proposed and are being refined (Lewis, 2001; Nylander et al., 2004; Ronquist, 2004).

The idiosyncrasies of nucleotide sequence data have spawned several methods for inferring phylogenies (Goldman, 1990; Penny et al., 1992; Swofford et al., 1996; Huelsenbeck and Crandall, 1997; Steel and Penny, 2000). Maximum likelihood methods evaluate hypotheses of evolutionary relationships using a presumed model of the evolutionary process and evaluate the probability that it would give rise to the observed data, which are typically DNA sequences of the terminal taxa (Felsenstein, 1973, 1981, 2004; Swofford et al., 1996; Huelsenbeck and Crandall, 1997). It should be noted that there are several different types of likelihood (Steel and Penny, 2000; Goloboff, 2003). Most maximum likelihood approaches in phylogenetics use maximum average likelihood, a form of maximum relative likelihood, which is discussed below.

The likelihood of an hypothesis (Fisher, 1922) is a function of the probability, P, of the data (D), given the hypothesis (H). Likelihoods are calculated for each possible tree topology, given the data and assuming a particular model of molecular evolution (Felsenstein, 1973, 1981, 2004; Swofford et al., 1996). The hypothesis, H, contains three distinct parts: 1) a mechanism or model of sequence

evolution, 2) a tree and 3) branch lengths (Penny et al., 1992). For a given data set, likelihoods are calculated for each of the possible tree topologies, or a sample of them, and the tree topology with the highest overall likelihood is the preferred phylogenetic hypothesis.

Calculating the likelihood can be computationally laborious if the data set is large, especially if the maximum likelihood model uses rooted trees in its calculus. The most general and most commonly used models in molecular analyses are, however, time reversible (Rodriguez et al., 1990; Swofford et al., 1996). With a time reversible model the probability of character state change from state i to state j is the same as the probability of change from state j to state i (Felsenstein, 1981). Under this condition the likelihood of the tree does not depend on the position of the root, and the use of unrooted networks greatly reduces the total number of trees to be evaluated, and decrease computation time (Rodriguez et al., 1990; Swofford et al., 1996). The network with the highest overall likelihood is the preferred phylogenetic hypothesis; the network topology thus maximizes the likelihood function for the data given the specified model (Felsenstein, 1973). The network is converted into a tree by rooting it with an outgroup or a molecular clock (Swofford et al., 1996; Felsenstein, 2004). It is always possible, however, that the network represents only a local maximum, or that it is one of a larger number of equally likely networks (Felsenstein, 1973; Chor et al., 2000; Salter and Pearl, 2001).

Models of Molecular Evolution

Likelihood analyses involve the same assumptions about the evolutionary process as other methods, including that evolution occurs in a branching pattern and is independent in different lineages (Swofford et al., 1996). The character change probabilities are calculated using a specified model of molecular evolution, which requires further assumptions about the nucleotide substitution process, including the assumption that sequence evolution can be modeled as a random, or stochastic, process (Rodriguez et al., 1990). Substitution models are typically based on a homogeneous Markov process (Rodriguez et al., 1990; Swofford et al., 1996) that assume that the probability of a state change at one site does not depend on the history of that site and that probabilities of substitution do not change significantly in different parts of the tree (Felsenstein, 1981, 2004; Swofford et al., 1996).

A DNA substitution model is expressed as a table of rates (substitutions per site per evolutionary distance unit) in which nucleotides are replaced by alternate nucleotides known as the Q matrix (Rodriguez et al., 1990; Swofford et al., 1996; Huelsenbeck and Crandall, 1997). In the instantaneous rate matrix, Q_{ij} represents the rate of change

from base i to base j over an infinitesimal evolutionary time period dt (Swofford et al., 1996). The rates defined in the Q matrix are per instant of time dt; in order to calculate the likelihoods of each site, the probabilities (P_{ej}) of the possible state changes along a branch length of t (Swofford et al., 1996) must be determined. For the simple Jukes-Cantor model, these values are easily evaluated because there are only two probabilities -the probability of a state change and the probability of stasis-such that the transition probability matrix consists of two values.

The substitution probability matrix that corresponds to the most general model has twelve values, one for each different substitution rate. The branch lengths are unknown prior to the analysis and must be estimated in the course of the likelihood calculation (Goloboff, 2003). Estimation of branch lengths involves an iterative algorithm in which each branch is optimized separately (Felsenstein, 1981; Swofford et al., 1996). Unlike the rate and frequency parameters, branch lengths are specific to a particular tree topology. For each tree, multiple different branch lengths need to be evaluated, and branch lengths must be recalculated for each network considered (Penny et al., 1992).

Models employed in likelihood analyses make explicit assumptions regarding sequence evolution (Swofford et al., 1996). The General Time Reversible Model (GTR) is the most general stochastic model of nucleotide substitution presently in use. It models base substitution as a random Markov process in which substitution rates are independent among sites, constant in time, equal in two lineages, and in which the ancestral sequence base frequencies represent the equilibrium frequencies (Rodriguez et al., 1990). The GTR model has a maximum of 12 different substitution rates (estimated from the data and using the aforementioned assumptions in their calculus) and at least seven parameters (Rodriguez et al., 1990). Because of its greater complexity, nearly all models (including JC, K2P, K3ST, L, TK, GIN, and TN) can be considered special cases of the GTR model (Rodriguez et al., 1990). For example, the Jukes-Cantor (often abbreviated JC69) model is the simplest model and assumes that all base substitutions are equally likely (i.e. all rate parameters are equal) and that the base frequency parameters are equal. The K2P model has two rate parameters since it considers differences in rates between transition and transversion type substitutions (Rodriguez et al., 1990). The K3ST model considers three substitution rates, one for transitions and two for each of two types of transversions.

The mathematical procedures of likelihood methods for phylogeny reconstruction have one critical component that is not met in the standard calculus of maximum likelihood. For probabilities to be multiplicative, the change

probabilities must be independent. The base compositional frequency parameters in the Q matrix are derived from the terminal taxon base sequences over *all* characters in the analysis (Siddall and Kluge, 1997). Rate parameters (relative rate and mean rate) are calculated using the Q matrix and the assumption that base frequencies remain constant over evolutionary time (Rodriguez et al., 1990). All sites use the Q matrix to calculate the P_{ej} values and therefore the probability for character *i* is dependent on all other characters through the frequency parameters in Q. Characters and their associated probabilities are thus not independent quantities, even though they are assumed to be so in the calculus of the method (e.g., Felsenstein, 1973, 1981; Rodriguez et al., 1990; Swofford et al., 1996). The non-independence of site change probabilities may be one factor responsible for the fact that the total likelihood of the universe of possible trees does not sum to unity (Felsenstein, 1981). The true probabilities for character changes should be calculated on an individual basis because they are connected with unique and historically contingent events (see below) (Farris, 1973). But this is clearly impossible, as it not only requires knowledge of the true history before undertaking an analysis, it also requires an objective and consistent way of determining the probability of a novel, context- specific evolutionary event (Farris, 1973); which is computationally impossible (Felsenstein, 1973, 1981; Siddall and Kluge, 1997). So, as currently and commonly employed in phylogenetic maximum likelihood methods, basic assumptions of frequency probability theory are violated (Yang, 1996; Siddall and Kluge, 1997).

Choosing a Model – More Ontological Parsimony

The choice of an appropriate model is a critical aspect of a phylogenetic likelihood analysis. There are many models of molecular evolution, and determining which to use can significantly influence the results of an analysis. Models range in complexity from the relatively simple Jukes-Cantor model, through the most complex GTR model. Currently there are at least 16 models that are commonly used in molecular systematics, most of which are special cases of the GTR model (Rodriguez et al., 1990). Each of the 16 basic models is varied with regard to *G* (gamma distribution), *I* (proportion of invariable sites), and both (*G+I*), for a total of 56 different options (Posada and Crandall, 1998). The overall likelihood score of a tree increases with increasing complexity of the model, but the accuracy of the model decreases with the increased number of estimated parameters (Huelsenbeck and Rannala, 1997b). The model that best fits the data while minimizing its complexity is chosen through pair-wise comparison of the maximum likelihood trees generated under each model using hierarchical likelihood ratio tests (Huelsenbeck and

Crandall, 1997; Huelsenbeck and Rannala, 1997b; Posada and Crandall, 1998; Johnson and Omland, 2004). When no statistically significant difference between two trees is found, the simplest model is selected. Recently several researchers have noted that the models being tested are not necessarily nested within each other, which is an assumption of the likelihood ratio test. These researchers advocate the use of the *Akaike Information Criterion* or the *Bayesian Information Criterion* when choosing the most parsimonious model (e.g. Posada and Buckley, 2004). Model selection based on the *relative* likelihood values is an ontological appeal to the principle of parsimony, because choosing the least complex explanation of the data rules out the possibility that evolution proceeded in a more complex manner (Huelsenbeck and Rannala, 1997b).

Criticisms of the Models

Many criticisms of maximum likelihood methods are directed at its *a priori* dependence on a model. Evolutionary realism of the models employed in likelihood analyses is often compromised by approximations designed to improve the computational efficiency of the algorithms. For example, Lockhart et al. (1994) suggested that a modified GTR model, in which time-reversibility is relaxed, across site rate variation is considered, and the nucleotide compositional frequencies are flexible, allows more evolutionary 'freedom' than any other model, and best considers the historical ambiguity and contingency of the evolutionary process. They suggested that this complex, parameter rich, and computationally intensive model should be logically preferred over all other models, if inferring phylogeny using the most realistic conception of evolution (i.e., evolution is complex) is the goal of the analysis. Relaxing the time-reversibility assumption, however, introduces the need for rooted trees, and is accompanied by additional computational problems (Swofford et al., 1996). Relaxing the assumption that rates are equal across all sites can be accomplished by adding another relative rate parameter to the matrix, which commonly involves modeling rate heterogeneity using the Gamma distribution (Swofford et al., 1996). If this distribution is modeled as continuous (as it should be), it again becomes computationally laborious, and a discrete distribution typically serves as a computationally more efficient approximation (Swofford et al., 1996).

Maximum likelihood also requires that numerous parameters be approximated using the data, and relies heavily on the frequency parameters that are taken directly from the observed sequences and the assumption that base frequencies are at equilibrium (Swofford et al., 1996). In this sense, likelihood methods require that the processes maintaining systems today were persistent throughout the

entire evolutionary history of the clade being investigated (Brooks and McLennan, 2002). Siddall and Kluge (1997) and Lockhart et al. (1994) provided empirical examples in which the nucleotide frequencies differ across terminal taxa, showing that the assumption of equilibrium base frequencies is not always tenable.

Criticisms of the Method

Use of maximum likelihood in phylogenetics relies on three assumptions: evolution is independent in (1) different lineages and (2) different sites for a given tree (Felsenstein, 1981, 2004; Rodriguez et al., 1990; Swofford et al., 1996), and (3) the same stochastic process of substitution applies in all lineages (Felsenstein, 1981). Some believe the assumptions are unrealistic and/or violated in the calculus (Siddall and Kluge, 1997; Huelsenbeck and Nielsen, 1999; Kluge, 2001; Goloboff, 2003), but likelihood users appeal to simulations to argue that the method is generally robust to violations of these assumptions (e.g., Felsenstein, 1978, 1981, 2004; Goldman, 1990; Penny et al., 1992; Yang, 1994; Swofford et al., 1996; Yang, 1996; Huelsenbeck and Rannala, 1997b; de Queiroz and Poe, 2001).

By relying on a specified model of sequence evolution to infer phylogenetic relationships, interpretation of maximum likelihood results comes with the caveat "if the model is true, then…" We may know which of the models best fits the data according to a model selection procedure, but how can the validity of the model itself be independently tested? Testing the validity of models, although it has been recognized as important (Goldman, 1990), is rarely done in practice (Siddall and Kluge, 1997).

Bayesian Likelihood

This claim [that the simplest hypothesis is more likely to be true] is generally defended by appeals to the Bayesian account of theory confirmation… (McAllister, 1996: 107)

Reverend Thomas Bayes, living in the early 18th century, was an English mathematician who was interested in the concept of using *a priori* knowledge to predict future events. His paper, 'An Essay Towards Solving a Problem in the Doctrine of Chances', published two years after his death in 1761, introduced what would become known as Bayes' theorem (Barnard and Bayes, 1958), in which the posterior probability, [P (H | D)], is the probability of the hypothesis given the observations, or data (D). Note that this differs from likelihood, which is the probability of the data given the hypothesis. However, the likelihood, P (D | H), is a parameter in the calculation of the posterior probability. P (H) is the *prior* probability of the hypothesis before the observation, data, or analysis, and reflects the original beliefs regarding the problem. P(D) is the probability of the

data, equal to the sum of the nominator for all considered hypotheses, and acts as a normalizing factor to ensure the sum of all posterior probabilities equals 1 or 100%. Bayes' Theorem describes the relationship between the prior and posterior probabilities. The prior probability of the hypothesis is updated to take into account the observations, producing a new estimate of the hypothesis that may form the prior probability for subsequent calculations if more observations are then considered. Bayes' Theorem thus acts in an iterative way, altering the posterior probability to reflect the effects, or likelihood, of all available data.

It was not until the latter half of the twentieth century that Bayes' ideas would be applied to phylogenetics. Felsenstein (1968) briefly discussed Bayesian ideas as they could apply to phylogeny reconstruction in his Ph. D. thesis, but the statistical and computational framework with which to derive reliable approximations of posterior probabilities was not available at the time (see Huelsenbeck et al., 2002). Harper (1979) also recognized the usefulness of Bayes' Theorem for choosing between competing phylogenetic hypotheses, although his method was largely conceptual and differed significantly from the Bayesian likelihood approach discussed below. His version of the Bayes' Theorem sought to determine the probability that some taxa were monophyletic given the observation of a synapomorphy between them. Harper's calculation was unique in including estimates of error due to misinterpreting plesiomorphies or homoplasies as synapomorphies, but was plagued by the need to subjectively estimate the likelihoods of null and alternative hypotheses. In 1996, three independent groups introduced working Bayesian methods for phylogenetics that are similar to those currently in use (Li, 1996; Mau, 1996; Rannala and Yang, 1996). All three evaluate phylogenetic hypotheses using the posterior probabilities of different trees.

The likelihood parameter, P (Data | Tree), is calculated using the same general methodology and models of molecular evolution described above for the maximum likelihood approach. The prior probability of the tree, P (Tree), is usually considered to be equal for all trees *a priori* (Archibald et al., 2003). The use of equal prior probabilities (=1/# possible trees) implies that no particular topology is *a priori* preferred over any other and eliminates the sometimes-difficult task of calculating complex prior probabilities when hypotheses vary with respect to their preconceived probabilities. However, the prior probability for any given tree or set of trees can be set to reflect researcher experience, the results of previous analyses, or taxonomy (Huelsenbeck et al., 2002). The denominator, simplified here as P (Data), is the normalizing factor involving summation over all trees (Yang and Rannala, 1997). The resulting posterior probability, P (Tree | Data),

can be interpreted as the probability that the tree is 'correct', *given* the data, the priors, and the model of character change (Huelsenbeck et al., 2000). There are several ways to present the results of a Bayesian analysis. The tree with the maximum *a posteriori* probability can be selected as the preferred phylogenetic hypothesis, this is also known as the MAP, or maximum *a posteriori* estimation of phylogeny (Rannala and Yang, 1996). Alternatively, one may construct a 95% credibility consensus tree by starting with the MAP tree and consecutively adding the next most probable trees until the probabilities total 0.95 (Altekar et al., 2004).

Posterior Probability Estimation using Markov Chain Monte Carlo

Calculating posterior probability of a tree is computationally expensive because it involves summation over all possible trees, and for each tree requires integration over all possible permutations of branch lengths and substitution-model parameters (Larget and Simon, 1999). This is not possible in most practical applications and requires that posterior probabilities be approximated (Huelsenbeck et al., 2002). Markov chain Monte Carlo (MCMC) methods are used to approximate the distribution of posterior probabilities and substitution parameters, allowing contemporary Bayesian likelihood methods to be computationally feasible (Hastings, 1970; Tierney, 1994). The application of the MCMC to phylogeny inference is discussed in detail by Mau and Newton (1997), Yang and Rannala (1997), Mau et al. (1999), Larget and Simon (1999), and summarized in Huelsenbeck et al. (2001, 2002), Pagel et al. (2004), and Kelly (2005). First, a random tree is selected and evaluated. Another tree is proposed by usng dependent sampling from the approximated distribution to change one variable of the original tree (*e.g.* topology, branch length, model parameters, etc.). The two trees are then compared using the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970; Green, 1995; Huelsenbeck et al., 2002): if the second tree represents an improvement, it is accepted and sampled, if not, the tree is accepted or rejected proportional to the likelihood ratio between it and the previous tree (Pagel et al., 2004). An accepted tree and its parameters are recorded and it then becomes the prior hypothesis to which the next change is compared. Since the MH algorithm results in a generalized increase in the posterior probability of successive accepted hypotheses, it will eventually converge on the most-likely range of model parameters and thereafter sample tree hypotheses in proportion to their frequency in the actual posterior density (Tierney, 1994; Pagel et al., 2004). The tree hypotheses sampled by the chain prior to convergence are generally discarded from the final posterior probability

calculations as part of the 'burn-in' since their acceptance is somewhat dependent upon sub-optimal (non-maximum likelihood) alterations of model parameters. The longer the chain is run, the greater precision with which the actual posterior distribution of trees is approximated (Pagel et al., 2004). Thus the frequency with which any particular tree is sampled while at convergence is proportional to its posterior probability. Likewise, the frequency with which a particular clade is seen in any hypotheses is proportional to its posterior probability. Extrapolating the actual posterior probability distribution from the MCMC chain is, however, only valid if the chain has reached convergence at the global maximum in the distribution (Altekar et al., 2004). To prevent entrapment of chains at sub-optimal 'hills' in the distribution multiple, simultaneous Markov chains are used, which periodically swap information. This Metropolis-coupled MCMC process improves mixing and convergence and allows the analysis of exceedingly large datasets that are beyond the scope of conventional single-chain MCMC Bayesian likelihood methods (Geyer, 1991; Huelsenbeck et al., 2001; Altekar et al., 2004).

Advantages of Bayesian Likelihood

A major advantage of the Bayesian likelihood method is the ease with which posterior probabilities can be interpreted (Huelsenbeck et al., 2002). Under the assumption that the evolutionary model is true and that the MCMC has accurately sampled the posterior probability distribution, the posterior probability value represents the probability that the tree is correct given the data. Similarly, the proportion of trees in the MCMC sample in which a monophyletic group appears represents the probability that the clade is 'true', given the caveats of model and data.

One of the most appealing aspects of Bayesian phylogenetic inference is its presentation and comparison of multiple optimal hypotheses. While maximum likelihood usually converges on a single hypothesis and maximum parsimony attempts to produce the shortest topologies, Bayesian likelihood produces a range of solutions, each with a corresponding overall posterior probability as well as comparable node support values for alternative topologies within each tree hypothesis (Li, 1996; Mau et al., 1999). Another major difference between Bayesian and maximum likelihood methods is that Bayesian likelihood calculation not only involves summation over all possible combinations of model parameters and branch lengths, but also includes a prior probability density distribution of these latter variables (Huelsenbeck et al., 2002), allowing the values of parameters to be adjusted according to MCMC sampling and MH selection. Therefore, although the parameters of distance correction models are specified *a priori* as in maximum likelihood, the *values* of these

parameters are allowed to vary and attain maximal states dependent on the topology in consideration.

Some view it as an advantage that Bayesian likelihood analysis requires the incorporation of previous knowledge or beliefs in terms of prior probabilities. The mechanics of formulating a starting prior can be difficult if one chooses to base it on the results of previous analyses or taxonomy ('complex priors', Huelsenbeck et al., 2002). Thus few priors are specified in practice. However, several authors (Li, 1996; Yang and Rannala, 1997; Larget and Simon, 1999; Huelsenbeck and Ronquist, 2001; Altekar et al., 2004) have explored the effects of different starting priors and found chains will nevertheless converge on consistent samples. Li (1996) further found that informative starting priors reduced the burn-in period by reducing the number of generations needed for maximum likelihood estimation. Making the initial prior probabilities of each tree equal eliminates complex priors, as well as any *a priori* assumptions that one hypothesis is more probable than any other in light of prior beliefs; clearly, this approach is not in the true Bayesian spirit (see Archibald et al., 2003).

Criticisms of Bayesian Likelihood

Bayesian likelihood approaches to phylogeny require a likelihood value for a given tree topology (i.e., phylogenetic hypothesis) to calculate the posterior probability of that evolutionary scenario. The likelihood calculation used in the Bayesian method requires the same models of evolution and their associated assumptions as the maximum likelihood methods described above, and thus all of the cautions inherent in maximum likelihood phylogeny estimation also apply to Bayesian likelihood analysis (Larget and Simon, 1999).

Analogous to maximum likelihood, the posterior probability of a tree involves summation over all possible trees (to calculate P (D)), including all their possible permutations in terms of branch lengths and substitution-model parameters (Larget and Simon, 1999). This is impossible to perform in most practical applications because of computational and time constraints, and necessitates approximation of posterior probabilities using Markov chain Monte Carlo techniques (Hastings, 1970; Tierney, 1994). Markov chains may fail to provide an accurate estimate of posterior probability distributions if they are not allowed to run long enough, or if mixing is a problem due to widely separated peaks in the distribution (Kelly, 2005). The longer the chain is run, the more precise the estimate of posterior probability distribution. However, it is difficult to know when a chain has run long enough to provide an acceptable estimate. Huelsenbeck et al. (2002) propose three recommendations to ensure that the posterior probability is sampled reliably, 1) run several long chains

independently and check for consistency in results, 2) run multiple chains, each starting from a random tree and check for consistency (Metropolis-coupled MCMC), and 3) monitor the model parameters for convergence. The Metropolis-coupling technique promotes good mixing and increases the speed of convergence (Huelsenbeck and Ronquist, 2001; Altekar et al., 2004), however chains must still be run for long periods following convergence to ensure one or more chains have not merely been caught in sub-optimal peaks in the distribution (Huelsenbeck et al., 2002).

The connection between Hennigian and Bayesian Likelihood Approaches

Shannon (1948) founded information theory using the statistical formulation of entropy as a synonym for expected uncertainty in the system, following on the proposition that increases in entropy were associated with losses of information. Shannon's use of entropy in information theory was consistent with its use in statistical mechanics and probability theory (Brillouin, 1951, 1953, 1962), which led Jaynes (1957a, 1957b) to formulate the first entropy maximization principle, in which the maximum entropy state of a system could be formally construed as the *a priori* most probable state (something originally proposed by Van der Waals, 1911). Departures from the most probable/most expected state were designated as "surprisals" (Kullbach,,, 1951; term first introduced by R. Levine). Parenthetically, and with respect to our discussion of maximum likelihood analyses above, Jaynes's use of the maximum entropy principle provided a rationale for choosing the most complex, rather than simplest, model for explaining a complex system. He reasoned that adopting the most complex model among all those that explained a system completely would expose our ignorance of possibilities, while adopting the simplest would give us a false sense of security, leading us to think we had more complete knowledge than we had.

Gatlin (1972) added two forms of redundancy in the context of biological (specifically DNA sequence) evolution to this conceptual framework. R-redundancy results from the repeated occurrence of the same symbol to get a message across. This is one way to ensure proper communication of a message, but since each symbol must be repeated, R-redundancy is also associated with reduced message variety. D-redundancy, or Shannon redundancy, results when multiple observations of the same thing require only one explanation (a single symbol). For example, the same trait occurring in two sister species needs only a single explanation (ie., one origin in the common ancestor). D-redundancy is associated with increased message variety, since no symbol need be repeated, and

with reduced message fidelity; missing the symbol results in a loss of information because it will never be repeated. Gatlin associated D-redundancy with optimal coding in communication systems. Overall then, R-redundancy is associated with low information density per symbol (each symbol represents only itself) and D-redundancy with high information density per symbol (each symbol represents many observations).

Nine years later, Brooks (1981), followed by Brooks et al. (1986), used the statistical concept of entropy as embodied in information theory to establish the conceptual links between Hennigian and Bayesian likelihood methods. Brooks (1981) showed that Hennigian phylogenetics operationally produced the lowest possible informational entropy configuration for a set of observations over a given set of taxa. Brooks et al. (1986) then made an informal link between this and Gatlin's D-redundancy, proposing the so-called D-measure for choosing optimal phylogenetic trees on the basis of maximum information density. Although not presented in those terms, the D measure is Bayesian in nature (as perhaps are all efforts to apply statistical reasoning to historical reconstructions). Following Jaynes (1957a, 1957b), Bayesian approaches in information theory are thus those for which the *a priori* subjective hypothesis is determined by the entropy maximum principle – the *a priori* most probable result is H_{max}, in direct analogy with the maximum entropy state being the most probable for a closed system. This becomes Bayesian if we stipulate that the set of observations we are using in any analysis is a closed subset of all possibilities, i.e., we stipulate that our estimate of H_{max} is based on a subjective sub-sample of an imperfectly known universe of characters, and we will not introduce additional observations during the testing procedure.

The entropy maximum is not only analogous to the *a priori* expected *most probable* state, it is also the state of lowest *information density* of each of the observations, hence least informative, hence least surprising (in a Bayesian sense). This would occur if each trait in each species evolved independently (i.e., there is no phylogenetic conservatism in character evolution). For any set of observations (a subjectively selected subset of all observations, drawn from a universe for which we do not have any sense of the actual size or distribution of variables), we can objectively compute the most probable state (H_{max}). We can also objectively compute the least probable state (H_{min}), which is the state of greatest information density for the observations, and thus the state of greatest surprise. This occurs when each trait evolves only once (i.e, when there is no homoplasy). The most powerful analysis of such data is one that seeks to find the most improbable/highest information density configuration permitted by all

the data at hand. For phylogenetic analysis, H_{max} and H_{min} can be calculated from the basic data matrix (hence H_{max} is *a priori*), whereas H_{obs} is calculated over a set of trees (hence, it is *a posteriori*). The preferred result is the one in which H_{obs} approaches H_{min} as closely as possible, which will also show the greatest difference between H_{max} and H_{obs}.

Applying the D measure leads to a number of conclusions for phylogenetic analysis (Brooks et al., 1986; Brooks and Wiley, 1988): (1) information density is proportional to evolutionary conservatism; (2) dichotomous solutions are preferred over polytomies, because dichotomous partitions of information are more information dense than polytomous ones; (3) there is no *a priori* difference between symmetrical and asymmetrical tree structure in terms of information density, since it is the information that produces the tree, not the reverse; (4) for any data set, the most information dense set of relationships of all taxa over all characters allowed by the data is the shortest tree; and (5) when there are multiple most parsimonious trees, ACCTRAN optimization provides a more information dense summary of the data than DELTRAN optimization.

Missing from this formalism are statistical significance tests capable of answering two questions: first, is the result (H_{obs}) significantly different from the *a priori* expectation (H_{max})?; and second, are less information-dense alternatives (e.g., other equally parsimonious trees or less than most parsimonious trees) for the same set of data significantly different from each other?

While such tests are not yet available for Bayesian Likelihood analyses, either, there is reason to believe that the development of such tests for one approach will suffice for both methodologies. Bayesian Likelihood bears a strong similarity to the D measure. For example, a key operation in the computer program Mr. Bayes (Ronquist and Huelsenbeck, 2003; Ronquist, 2004) is "data compression", which must be related to the most information dense configuration of the observations in the data matrix. As noted above, informational measures are now being used to choose the most parsimonious model for Bayesian likelihood analyses, reinforcing the suspicion that Hennigian and Bayesian likelihood approaches are highly complementary. In addition, Huelsenbeck and Rannala (2004) recently proposed that the best Bayesian likelihood results would be obtained if one chose the most complex model, much in the same sense as the proposals by Lockhart et al. (1994) for maximum likelihood. These views also complement the use of the *Akaike Information Criterion* or the *Bayesian Information Criterion* when choosing the most parsimonious model for maximum likelihood analysis proposed by Posada and Buckley (2004). The most complex model possible is one in which

the evolution of each character state in each taxon is independent, or H_{max} for any data set.

How Do You Decide Which Method(s) to Use?

We have discussed a variety of objective methods for pursuing quantitative phylogenetic analysis. We believe, however, that there are no objective means by which one can choose among these methods. Consequently, it is no surprise that some of the most contentious interactions among phylogeneticists concern the very subjective issue of which methods are "best" or "proper" or "correct". It is common for groups of contending scientists, faced with such a situation, to resort to philosophical arguments in an attempt to claim priority for one viewpoint over another on the basis of some set of first principles. This has certainly been the case with phylogenetics.

Deductive versus inductive approaches

The first salvo fired in this conflict was by Wiley (1975), who, defending phylogenetic systematics against claims that it was not falsifiable, proposed that phylogenetic hypotheses of homology could be seen as an exercise in hypothetico-deductive reasoning

> Once a hypothesis of homology is formulated from the world of experience it is tested in two phases: by its own set of potential falsifiers and by a set of potential falsifiers of the phylogenetic hypothesis to which it belongs as a proper subset (i.e. it is tested by other hypotheses of synapomorphy through the testing of the phylogenetic hypotheses which they corroborate). Both phases of testing must be done under the rules of parsimony, not because nature is parsimonious, but because only parsimonious hypotheses can be defended by the investigator without resorting to authoritarianism or apriorism. (Wiley 1975: 236)

Hypotheses of homology (characters), together with their connected hypothesis of phylogenetic relationships, can be tested by other independently proposed homologies, which then represent 'potential falsifying hypotheses' (*sensu* Popper 1968). Wiley emphasized that such a process is not circular, but represents a case of 'reciprocal illumination' (Hennig 1966). He noted that the preferred phylogenetic hypothesis is the one that has been refuted the least number of times. Shortly thereafter, Engelmann and Wiley (1977) suggested that outgroup comparison made polarization decisions testable in a Popperian sense; that is, such decisions were capable of being falsified. That Hennigian phylogenetics was justified by the hypothetico-deductive approach of Popper quickly gained support (e.g.,

especially Gaffney 1979) and still has strong adherents (e.g., Kluge 2003) who consider Hennigian phylogenetics to be deductive in nature.

Recently, de Queiroz and Poe (2001; also Faith and Trueman 2001) attempted to link Popperian thought with likelihood approaches, suggesting that likelihood is the basis for Popper's notions about the degree of corroboration of a hypothesis. For Popper, corroboration was embedded in a falsificationist context, however these authors sought to decouple Popper's ideas about corroboration from those about falsification. Their degree of corroboration is thus more correctly identified with Popper's degree of confirmation, which Popper associated with an inductivist and verificaiotnist viewpoint, and rejected (Popper 1997). This seems to get us nowhere, since it leads back to the position that if a model is accepted as true, or highly typical, its use is justified. But, no objective means is provided for verifying or falsifying the validity of the model beyond the arguments about statistical consistency whose shortcomings we discuss below.

Regardless of semantic arguments about corroboration and confirmation, and possibly a high degree of revisionist interpretation of Popper's views on the relationship between corroboration and falsification, these arguments do not counter the basic observation that maximum likelihood methods are more inductive than deductive in spirit. And, if the difference between what we have characterized as the epistemological and the ontological parsimony approaches is the difference between a preference for deduction and a preference for induction, the history of science tells us that there is no objective means for choosing between them, despite strong personal convictions on both sides of the issue.

However popular it has been among some systematists, this battle of philosophical perspectives has been criticized by philosophers; best summed up by Sober (1988), who identified phylogenetic analysis as abductive, that is, neither exactly deductive nor exactly inductive. This occurs because the phylogenetic inference is based on a retrodictive analysis of historically unique events. That is, inference goes from from effects to cause(s). As systematists, we observe the effects (phylogenetic trees) under the causal theory of descent with modification (i.e. observable synapomorphies), but there are also other possible causes for conflicting data (reversals and parallelisms), just as there might be multiple cuases for the same phylogenetic outcomes. Multiple conclusions about cause(s) are thus possible in phylogenetic inference. By contrast, true deduction enables inferences from cause to effect(s), with singular conclusions for any given analysis.

The differences between the two types of methodologies

then, is not so much one of deduction versus induction, but one of the preference for using either epistemological or ontological parsimony. Hennigians choose the epistemological perspective which suggests that evolution may have been so complex that we should expect to find conflicts in the data, whose resolution requires a logical decision-making principle (Brooks and McLennan, 2002). An important corollary of this perspective is that there need be no necessary connection between the most parsimonious hypothesis and truth. Hennigians are thus preoccupied with the *robustness* of their results. They do not believe their hypotheses can be verified, but do believe that they can be falsified at least in part using new data. Phylogeny reconstruction is thus an open-ended process involving a potentially endless search for information. If, at some point in the future, the accumulation of data leads to a situation in which the phylogenetic hypothesis for a given group is no longer changing with the addition of new data, Hennigians may express the belief that the hypothesis has approached the truth as closely as possible, but in principle a Hennigian will never claim to have the true phylogeny.

The ontological perspective adopted by likelihood and Bayesian approaches, by contrast, requires first that evolution be parsimonious in some manner, usually as defined by certain assumptions and parameters of a model; and second, that the resulting phylogenetic hypothesis be accepted as true so long as the model is accepted as true, qualified by the parameters of the model and the data. McAllister (1996: 107) stated it thusly,

> The argument from likelihood rests on the claim that, of two theories that fit the data equally well, the simpler has a higher likelihood of being true.

Practitioners of likelihood are thus preoccupied with the *accuracy* of their results, and believe it is possible to develop means by which their preferred hypotheses can be verified with respect to the true phylogeny.

Both perspectives have strengths and weaknesses, both are addressing important evolutionary questions. We believe that it is the failure to distinguish ontological from epistemological uses of parsimony that has led to endless disputes as advocates argue past each other, rather than a failure in the methods themselves.

Statistical Consistency

A method is said to be statistically consistent if it converges on the true tree as progressively more character data are added to the analysis. One reason likelihood approaches have gained popularity is that other methods of phylogenetic inference, namely Hennigian phylogenetics and maximum parsimony, are statistically inconsistent

under certain circumstances (Felsenstein 1978; Penny et al. 1992). The region of statistical inconsistency has been referred to as the 'Felsenstein Zone', and it is the result of a process termed 'long-branch attraction'. The long-branch attraction problem occurs when convergent homoplastic changes are more frequent than non-reversed changes in an informative part of the tree (Felsenstein 1978). This confounds Hennigian phylogenetics because, under the Auxiliary Principle, the convergent homoplasies will tend to be considered as homologies and thus the taxa with their convergent 'long-branches' will be grouped together (Hennig 1966; Felsenstein 1978). In simplest terms, when the *data are lying* about the relationships of the taxa, Hennigian phylogenetics may fail to discover the true relationships. How often this occurs in nature is unknown, but Huelsenbeck (1997) cited one case involving insects as exemplifying the long-branch attraction problem in a real data set (but see Siddall and Kluge 1997: 319-20). Some believe, however, that 'noise', or random data, does not misdirect phylogenetic systematics often enough to be a major concern (Wenzel and Siddall 1999).

Maximum likelihood has been reported to exhibit the favourable property of statistical consistency in the face of these situations (Felsenstein 1978; Penny et al. 1992; Yang 1994). It is true that in statistics the maximum likelihood estimate of a parameter is consistent (Fisher 1922; Edwards 1972). Simulation experiments have shown this to be true in the phylogenetic context (Yang 1996), but *only* when the same random model used to generate the data is used and/or only when a certain correction factor is implemented (Goloboff 2003; Steel et al. 1993; Siddall and Kluge 1997; Steel and Penny 2000). It is an important caveat that maximum likelihood methods are only consistent (i.e. converge on the 'true tree') under a certain set of circumstances which typically requires that the 'correct' model is used, but that the correct model and the true tree are both unknowable for real systems. When the model is insufficient or inappropriate, appeals to statistical consistency are rendered moot (Siddall and Kluge 1997).

In a complementary vein, Farris (1973) suggested a protocol by which parsimony methods could be interpreted as derivatives of statistical estimation methods. This probabilistic view of parsimony was critiqued by Felsenstein (1973, 1978, 1981, 1983), who focused on the statistical deficits of parsimony when viewed as a likelihood method. In general, parsimony and likelihood approaches produce the same results under the assumption of particular parameters for parsimony, i.e. low rates of evolutionary changes or equal rates of evolution among the observed lineages, or low rates of homoplasy (Felsenstein 1983). It has also been demonstrated that parsimony-based

methods can be consistent on their own (Siddall and Kluge 1997; Steel and Penny 2000) or with a correction factor (Steel et al. 1993). Overall then, statistical consistency is a not a property of a method, but the property of a specific data set, the model, and specific situations (Siddall and Kluge 1997; Steel and Penny 2000).

A persistent philosophical objection to likelihood methods derives from the fact that all forms of the method use frequency probability theory (Kluge, 1990, 1997; Siddall and Kluge, 1997). The arguments goes as follows: (1) the aim of phylogenetic systematics is to discover the unique evolutionary history of a group of organisms, to elucidate its past; (2) frequency probability is primarily concerned with prediction of future events (e.g., Fisher 1922); (3) likelihood methods apply frequency probability to a *historical singularity*, which is outside of the realm of future-predictive probability theory. All possible trees are assigned a non-zero probability, but in reality one tree has a probability of 1.0, and all others have a probability of zero. Proponents of this position are faced with a conundrum, namely that the D measure also utilizes frequency probabilities and yet leads to Hennigian/maximum parsimony methods. Luckily, this conundrum can be resolved fairly easily, because it is not a new argument. In a similar vein, Franz Boas, a founder of cultural Anthropology and an early champion of phylogenetic comparative studies, suggested that 19th century science had produced

> "a grand picture of nature in which for the first time the universe appears as a unit of ever-changing form and color, each momentary aspect being determined by the past moment..." Franz Boas, [*History of Anthropology* pp. 515, 524; *Mind of Primitive Man*, 1911, 1938 – 2nd ed, p. 11].

> Discussing the early history of statistical mechanics, Brush (1983: 65) noted the fact that a macrostate can be assigned a certain "probability" does not necessarily mean that its existence results from a random process. On the contrary, the use of probabilities here is perfectly compatible with the assumption that each macrostate is rigorously determined by its previous state and the forces acting on it. We need to use probability measures because we must deal with macrostates corresponding to large numbers of microstates. Boltzmann might have avoided the connotations of the word probability by using a neutral term such as "weighting factor."

If we consider each possible tree a macrostate (one possible outcome of a complex historical process called

phylogenetic diversification) and all traits used in an analysis (e.g. all base pairs) the microstates (Brooks and Wiley, 1988), the use of frequency probabilities for phylogenetic analysis is justified on these grounds. The term "weighting factor', however, is not neutral among systematists, even though all three classes of quantitative methods utilize different types of weighting factors (considering maximum parsimony as using a weighting factor of 1.0 for each trait). Finally, the use of "frequency" has been co-opted for comparative analysis of gene frequency data (Swofford and Berlocher, 1987), so we are left with using "probability" as the "least non-neutral available" term.

A New Strategy

We believe that neither Popper's philosophy nor appeals to statistical consistency can give precedence for one method of quantitative phylogenetic analysis over any others. Is there an objective way to reconcile these subjectively divergent approaches? An appeal to collegial pluralism (e.g., Faith and Trueman 2001) seems like a good idea at first glance. It is becoming common practice for an author to present maximum parsimony, maximum likelihood and Bayesian analyses of the same data, then either arbitrarily expressing a preference for one of them, or presenting a consensus tree of the outcomes of each analysis, and using that as "the phylogeny". We support the sentiment behind this proposal, but do not believe it is the best approach. In the first case, the arbitrarily chosen result inevitably is the one that best supports the evolutionary scenario advocated by the author, which actually weakens the author's case over a situation in which that result emerged uniquely from the data. In the second case, a consensus tree effectively hides precisely the parts of analyses that are in need of additional scrutiny, giving author and audience a false sense of security about the results (Miyamoto 1985).

We believe that individual data analyses presented without reference to an explicit evolutionary model or hypothesis (i.e., epistemological parsimony) are not explanations. They are descriptions, admittedly highly sophisticated descriptions, but just that. Using the principle of parsimony as an epistemological tool ensures that we have the most robust empirical result; adopting the most parsimonious summary of the data with respect to outgroup comparisons ensures that the most robust result can be interpreted phylogenetically. However, because such analyses are based on a weak homology criterion, strong interpretations of phylogenetic trees and their evolutionary significance typically require more information than a branching diagram (Brooks and McLennan 2002). Likewise, fitting data to a model provides explanations,

but only if the model is known or assumed to be true. That is, there is no means by which model-based methods can test the veracity of any given model or its assumptions; as we have shown above, they choose a model for a given set of data based only on an ad hoc preference for ontological simplicity.

Independent description and assessment relative to explanatory models both appear to be necessary but not sufficient for robust explanations. Or,

> If models do not agree with the empirical data, chances are the models, not the data, should be re-evaluated. This is not an antimodel stance. A mutually reinforcing and mutually modifying dialogue between models and empirical discovery enhances progress. (Brooks and McLennan 2002: x).

Kluge (1989, 1991, 1997, 1998a, 1998b, 1999) has argued that historical sciences progress through cycles of discovery and evaluation, both being necessary but neither being sufficient for robust explanations. We believe that Hennigian (epistemological parsimony) analysis is the best discovery method we have in phylogenetics. This is because its results are dependent on a minimum of *a priori* assumptions and thus the range of potential discoveries indicated by the data is greater than for any ontological parsimony approach. At the same time, we believe that this feature of Hennigian phylogenetics renders it relatively weak as an instrument of evaluation. In a complementary manner, it appears to us that the various maximum likelihood and Bayesian likelihood approaches are admirably suited as evaluation methods. We would like to see epistemological and ontological parsimony methods used together in a form of reciprocal illumination, not in the narrow sense of deriving a tree from multiple characters, but in a broader sense of cycles of discovery and evaluation.

To illustrate this point, we offer the following thought experiment: Suppose a Hennigian analysis, a maximum-likelihood analysis, and a Bayesian likelihood analysis produce the same results. We should all celebrate, because we would have a relatively independent discovery (the Hennigian tree) supporting an evolutionary model (the likelihood tree). In this case, no one should have any concerns about using the likelihood model to infer divergence rates on the Hennigian tree. Now, what does it mean if these different analyses do **not** produce the same results? In such cases it is likely that the data on hand do not contain very strong phylogenetic signal. This is evidenced by low support for the nodes that differ between the analyses and/or very short branches at these nodes. Hillis et al. (1992, 1993; Huelsenbeck and Hillis 1993) studied this problem when they produced a phylogeny

for bacteriophages maintained in the laboratory. They discovered that most quantitative methods converged on the same, and true (since it was known) phylogeny as more and more traits were sampled. These results would seem to suggest that the primary response to any situation in which the different approaches to phylogenetics produce different answers should be

> When in doubt, get more data. (Brooks and McLennan 2002: 148)

Maximum parsimony analysis is virtually isomorphic with Hennigian phylogenetics whenever outgroup comparisons are used to root a minimum-length network according to Hennig's Auxiliary Principle. Maximum likelihood and Bayesian likelihood should also converge on Hennigian phylogenetics as more data are sampled and the preferred (most parsimonious possible) model becomes more complex, especially if some form of outgroup comparison is used in rooting the tree. The entropy maximum principle in Bayesian information theory shows clearly that the most complex model *permits all possibilities a priori*, while the Auxiliary Principle *prohibits nothing a priori*, clearly two ways of saying the same thing.

If this is true, then we can assess our progress by asking just how much disagreement there is in published phylogenetic trees based on the same data but derived using the different methods. A survey of volumes 51-53 of *Systematic Biology* revealed 20 studies that reported the results of both Hennigian and model-based (Bayesian and/or maximum likelihood) analyses of the same dataset(s). The phylogenetic hypotheses reported in these studies were based on datasets consisting of one to 11 gene sequences (3 gene sequences on average). Trees constructed using Bayesian and maximum likelihood were generally identical, which is not surprising as the underlying model chosen was identical for both methods. Trees constructed by model-based and Hennigian methods were identical in only 5 out of 28 cases (several studies reported results for each gene region separately as well as combined); nonetheless, they agreed far more than they disagreed. On average, the phylogenetic hypotheses tended to have 88% identical nodes. Although there was no correlation between the number of gene sequences and percent identical nodes (Spearman rank correlation, r= 0.105, P= 0.594), variation in percentage similarity between trees derived with the different methods was greatest in those studies using only one gene sequence, and this variation decreased with the number of gene sequences used. All three studies using 6 gene sequences showed identical trees regardless of method used for analysis. One study using 11 gene sequences did not produce identical trees with the different methods. However, the differences were

weakly supported and branches leading to these nodes were almost nonexistent, suggesting the clades had diverged very quickly and possibly no amount of data would ever resolve the ambiguity in a satisfactory manner.

Despite the small number of genes used on average per study, the survey suggests that (1) there is far more agreement than disagreement among the results using the different methods (which we expect given that all maximize some form of parsimony), and that (2) more data leads to congruent results from all methods, especially when the data are analyzed in a combined analysis framework. It seems clear that one should always attempt to gather as much data as possible and use differences in topology from different methods of analysis as a focus for collecting more data. Increasing the number of genes sequenced will become easier as time passes, but currently a good source of "extra" phylogenetic information appears to be morphology (Baker and Gatesy 2002; Wahlberg and Nylin 2003; Wahlberg et al. 2005). In fact, one could well argue that because morphological traits are generally the result of multi-gene interactions, morphology can be excellent "evolutionary control" data when there are concerns about biased gene sampling as discussed e.g., by Rokas et al. (2003; see also Mattern and McLennan, 2004) It also appears clear that a "total evidence" approach gives the most robust answer, whether each character is allowed to evolve independently (Hennigian approach), different partitions are allowed to evolve according to different models (Bayesian likelihood approaches) or all partitions are forced to evolve according to the same model (current implementations of Maximum Likelihood).

What do we do while we are waiting for enough data to give the same answer with all methods? Hillis et al. (1992, 1993; Huelsenbeck and Hillis 1993) showed that when data are limited, some models generate the correct phylogeny better than Hennigian approaches. Although some interpret this finding as an indication that model-based methods are inherently superior to Hennigian methods, Hillis et al. (1994) pointed out a significant trade-off. Model-based approaches provide a distinct answer based on little data, but the confidence you have in that answer is proportional to your belief that the model used accurately reflects the evolutionary process over extended periods of time for the clade being analyzed.

The issue becomes (again), how do we know the model typically gives the truth? Hillis et al. (1992) took a critical first step by generating an experimental phylogeny. The next step is to ask how typical of evolution is that phylogeny? Remember that the phylogeny involved bacteriophages and was generated in the laboratory according to rules invoked by the researchers (reminiscent of the Caminalcules) to develop one of the first epistemological

parsimony algorithms. Some have suggested that prokaryote evolution has produced not a phylogenetic tree but a highly reticulated network (Doolittle 2000), in which case the experimental phylogeny produced by Hillis et al. (1992) is not typical of evolutionary history for their model organisms. Nonetheless, their results may still be typical of phylogenesis for many groups of eukaryotes.

More important is the question, how large a role has the historical contingency that is such a critical part of Darwinian mechanisms played in phylogenesis? Some believe that such contingencies do not affect phylogenetic reconstructions while others believe the opposite (see Yang and Bielawski 2000 for a review). Seen in this light, is it possible that the reason Hennigian and model-based approaches converge with increasing data is that the more data we consider, the more historical contingencies will play a role, in which case model-based approaches will progressively choose models whose set of "allowed possibilities" most closely approximates the minimal "*a priori* restrictions" of Hennigian phylogenetics, such as the modified GTR proposed by Lockhart et al. (1994). For example, Gissi et al. (2000) reported lineage-specific evolutionary rates for different mammalian mtDNA genes, suggesting that finding the correct phylogeny might require a different model for each gene. They suggested their findings supported contentions by other molecular systematists that, given uncertainty about the true phylogeny, we cannot know which model will give the correct phylogeny, and thus we use analyses of as many genes as possible to help determine the appropriate model (see also Mitchell et al. 2000; Kolaczkowski and Thornton 2004).

Conclusions – A Stable Platform for the Future
To summarize:

• Using the principle of parsimony as an epistemological tool ensures that we have the most robust empirical result given the data; adopting the most parsimonious summary of the data with respect to outgroup comparisons ensures that the most robust result can be interpreted phylogenetically.

• Likewise, fitting data to a model provides explanations, but only if the model is known or assumed to be true. That is, there is no means by which model-based methods can test the veracity of any given model or its assumptions; researchers generally choose a model for a given set of data based only on an ad hoc preference for ontological simplicity (ontological parsimony).

• As more data are added to the study, however, likelihood methods, including Bayesian likelihood analysis, generally move more quickly to the most complex model. The entropy maximum principle in Bayesian information theory shows clearly that the most complex model *permits*

all possibilities a priori, while the Auxiliary Principle *prohibits nothing a priori*, clearly two ways of saying the same thing. This may explain why *Mr. Bayes* behaves so much like "Mr. Hennig."

• The three classes of quantitative methods of phylogeny reconstruction begin to converge on the same answer as more data are added to the analysis. If all three methods produce identical results, there is no way, or even reason, to choose between them. Conflicting results in part of the tree highlight peculiarities in the data set (e.g., long branch attraction, rapid diversification etc). In this situation, choice of method depends upon the resistance of each method to the particular vagaries of the data.

• Hennigian (epistemological parsimony) analysis is the best discovery method we have in phylogenetics because its results are dependent on a minimum of *a priori* assumptions and thus the range of potential discoveries indicated by the data is greater than for any ontological parsimony approach. In a complementary manner, the various maximum likelihood and Bayesian likelihood approaches are admirably suited as evaluation methods, that is as methods allowing us to investigate the processes potentially underlying phylogenetic patterns. The most complete explanation for phylogeny is thus the one that incorporates information from both Hennigian and likelihood approaches.

While it is often true that disputes among scientists are the engines of innovation and development, they can also have the opposite effect. There is no doubt that quantitative phylogenetic analysis has revolutionized evolutionary studies and has had significant impacts in many areas of biology, both basic and applied (for a panoramic overview, see Brooks and McLennan 2002). Much of the academic infighting associated with the development and deployment of different methods for phylogenetic analysis, however, has obscured the real progress that has been made in reconstructing the phylogenetic history of life. This, in turn, has undermined efforts by systematists to gain proper credit for their role in making possible the explosion of comparative evolutionary biology during the past 30 years and, more recently in dealing with the global biodiversity crisis. We hope that, by highlighting the substantial complementarities among these methods, as well as the shortcomings of each, we can lay the groundwork for making the competitive arena of development more collegial than it has been in the past.

Acknowledgements

Literature cited

Adams, E. 1972. Consensus techniques and the comparison of taxonomic trees. Systematic Zoology 21:390-397.

Alfaro, M. E., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Molecular Biology and Evolution 20:255-266.

Allard, M., and J. Carpenter. 1996. On weighting and congruence. Cladistics 2:183-198.

Allard, M., Farris J., and J. Carpenter. 1999. Congruence among mammalian mitochondrial genes. Cladistics 15:75-84.

Altekar, G. S., S. Dwarkadas, J. P. Huelsenbeck, and F. Ronquist. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20:407-415.

Archibald, J. K., M. E. Mort, and D. J. Crawford. 2003. Bayesian inference of phylogeny: a non-technical primer. Taxon 52:187-191.

Baker, A. M., and R. DeSalle. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Systematic biology 46:654-673.

Baker, R., and J. Gatesy. 2002. Is morphology still relevant? *In* Molecular Systematics and Evolution: Theory and Practice, R. DeSalle, W. Wheeler and G. Giribet (eds.). Birkhauser Verlag, Basel. p. 163-174.

Baker, R. H., X. B. Yu, and R. DeSalle R. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Molecular Phylogenetics and Evolution 9:427-436.

Barnard, G. A., and Bayes T. 1958. Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. Biometrica 45:293-315.

Barrett, M., M. J. Donoghue, and E. Sober. 1991. Against consensus. Systematic Zoology 40:486-493.

Björklund, M. 1999. Are third positions really that bad? A test using vertebrate cytochrome b. Cladistics 15:191-197.

Boyden, A. 1947. Homology and analogy: a critical review of the meanings and implication of these concepts in biology. American Midland Naturalist 37:648-660.

Bremer, K. 1988. The limits of amino-acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795-803.

Brillouin, L. 1951. Physical entropy and information. Journal of Applied Physics 22:338-343.

Brillouin, L. 1953. Negentropy principle of information. Journal of Applied Physics 24:1152-1163.

Brillouin, L. 1962. Science and information theory. 2nd edition. Academic Press, New York. 347 p.

Brochu, C. A. 1999. Taxon sampling and reverse successive weighting. Systematic Biology 48:808-813.

Brooks, D. R. 1981. Classifications as languages of empirical comparative biology. *In* Advances in cladistics: Proceedings of the first meeting of the Willi Hennig society, New York Botanical Garden, V.A. Funk and D. R. Brooks (eds.). New York. p. 61-70.

Brooks, D. R. 1996. Explanations of homoplasy at different levels of biological organization. *In* Homoplasy, R. J. Sanderson and L. Hufford (eds.). Academic Press, London. p. 6-36.

Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, ecology and behavior: a research program in comparative biology. University of Chicago Press, Chicago. 434 p.

Brooks, D. R., and D. A. McLennan. 2002. The nature of diversity. University of Chicago Press, Chicago. 668 p.

Brooks, D. R., R. T. O'Grady, and E. O. Wiley. 1986. A measure of the information content of phylogenetic trees, and its use as an optimality criterion. Systematic Zoology 35:571-581.

Brooks, D. R., and E. O. Wiley. 1988. Evolution as entropy: toward a unified theory of biology. University of Chicago Press, Chicago. 415 p.

Broughton, R. E., S. E. Stanley, and R. T. Durrett. 2000. Quantification of homoplasy for nucleotide transitions and transversions and a reexamination of assumptions in weighted phylogenetic analysis. Systematic Biology 49:617-627.

Brown, W. M., E. M. Prager, A. Wang, and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. Journal of Molecular Evolution 18:225-239.

Brush, S. G. 1983. Statistical physics and the atomic theory of matter, from Boyle and Newton to Landau and Onsager. Princeton University Press, Princeton. 277 p.

Camin, J. H., and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. Evolution 19:311-326.

Carpenter, J. 1988. Choosing among multiple equally parsimonious cladograms. Cladistics 4:291-296.

Carpenter, J. M. 1992. Random cladistics. Cladistics 8:147-153

Carpenter, J. M. 1994. Successive weighting, reliability and evidence. Cladistics 10:215-220.

Cavalli-Sforza, L. L., A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Evolution 21:550-570.

Charlesworth, M. J. 1956. Aristotle's razor. Philosophical Studies 6:105-112.

Chippindale, P. T., and J. J. Wiens. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Systematic Biology 43:278-287.

Chor, S., M. Hendy, B. R. Holland, and D. Penny. 2000. Multiple maxima of likelihood in phylogenetic trees: an analytic approach. Molecular Biology and Evolution 17:1529-1541.

Churchill, S. P., E. O. Wiley, and L. A. Hauser. 1984. A critique of Wagner groundplan-divergence studies and a comparison with other methods of phylogenetic analysis. Taxon 33:212-232.

Colless, D. H. 1966. A note on Wilson's consistency test for phylogenetic hypotheses. Systematic Zoology 15:358-359.

Colless, D. H. 1967. An examination of certain concepts in phenetic taxonomy. Systematic Zoology 16:6-27.

Colless, D. H. 1985. On "character" and related terms. Systematic Zoology 34:229-233.

Collins, T. M., F. Kraus, and G. Estabrook. 1994. Compositional effects and weighting of nucleotide sequences for phylogenetic analysis. Systematic Biology 43:449-459.

Cunningham, C. W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Systematic Biology 46:464-478.

Danser, B. H. 1950. A theory of systematics. Bibliotheca Biotheoretica 4:113-180.

Darwin, C. 1872. On the origin of species. Sixth edition. 6 ed. John Murray, London. 476 p.

de Queiroz, K., and S. Poe. 2001. Philosophy and phylogenetic inference: a comparison of likelihood and parsimony methods in the context of Karl R. Popper's writings on corroboration. Systematic Biology 50:305-321.

Dixon, M. T., and D. M. Hillis. 1993. Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Molecular Biology and Evolution 10:256-267.

Doolittle, W. F. 2000. Uprooting the tree of life. Scientific

American 282:90-95.

Edwards, A. W. F. 1972. Likelihood. Cambridge University Press, Cambridge. 275 p.

Edwards, A. W. F. 1996. The origin and early development of the method of minimum evolution for the reconstruction of phylogenetic trees. Systematic Biology 45:79-91.

Edwards, A. W. F., and L. L. Cavalli-Sforza. 1963. The reconstruction of evolution. Annals of Human Genetics 27:105-106.

Edwards, A. W. F., and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. *In* Phenetic and phylogenetic classification, V. H. Heywood and J. McNeill (eds.). Vol. Publication No. 6, Systematics Association, London. p. 67-76

Efron, B. 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7:1-26.

Engelmann, G. F., and E. O. Wiley. 1977. The place of ancestor-descendant relationships in phylogeny reconstruction. Systematic Zoology 26:1-11.

Faith, D. P., and J. W. H. Trueman. 2001. Towards an inclusive philosophy for phylogenetic inference. Systematic Biology 50:331-350.

Farris, J. S. 1966. Estimation of conservation of characters by constancy within biological populations. Evolution 20:587-591.

Farris, J. S. 1969. A successive approximations approach to character weighting. Systematic Zoology 18:374-385.

Farris, J. S. 1970. Methods of computing Wagner trees. Systematic Zoology 19:83-92.

Farris, J. S. 1973. A probability model for inferring evolutionary trees. Systematic Zoology 22:250-256.

Farris, J. S. 1977. Phylogenetic analysis under Dollo's Law. Systematic Zoology 26:77-88.

Farris, J. S. 1982. Outgroups and parsimony. Systematic Zoology 31:328-334.

Farris, J. S. 1983. The logical basis of phylogenetic analysis. *In* Advances in cladistics: Proceedings of the Second Meeting of the Willi Hennig Society, N. I. P. and V. A. Funk (eds.). Columbia University Press, New York. p. 7-36.

Farris, J. S. 1989. The retention index and rescaled consistency index. Cladistics 5:417-419.

Farris, J. S. 2001. Support weighting. Cladistics 17:389-394.

Farris, J. S., V. A. Albert, M. Kallersjo, D. Lipscomb, and A. G. Kluge. 1996. Parsimony jackknifing outperforms neighbour-joining. Cladistics 12:99-124.

Felsenstein, J. 1968. Statistical inference and the estimation of phylogenies. Ph. D. thesis., University of Chicago.

Felsenstein, J. 1973. Maximum likelihood and minimum steps methods for estimating evolutionary trees from data on discrete characters. Systematic Zoology 22:240-249.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27:401-410.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17:368-376.

Felsenstein, J. 1983. Parsimony in systematics: biological and statistical issues. Annual Reviews on Ecology and Systematics 14:313-333.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Inc., Sunderland. 664 p.

Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society, A Series 222:309-368.

Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specified tree topology. Systematic Zoology 20:406-416.

Gaffney, E. S. 1979. An introduction to the logic of phylogeny reconstruction. *In* Phylogenetic analysis and paleontology, J. Cracraft and N. Eldredge (eds.). Columbia University Press, New York. p. 79-111.

Gatesy, J., R. DeSalle, and W. Wheeler. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Molecular Phylogenetics and Evolution 2:152-157.

Gatesy, J, R. T. O'Grady, and R. H. Baker. 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15:271-313.

Gatlin, L. 1972. Information theory and the living system. Columbia University Press, New York. 210 p.

Geyer, C. J. 1991. Markov chain Monte Carlo maximum likelihood. *In* Computing science and statistics: proceedings of the 23rd symposium on the interface, Keramidas (ed.). Interface Foundation, Fairfax Station. p. 156-163.

Gissi, C., A. Reyes, G. Pesole, and C. Saccone. 2000. Lineage-specific evolutionary rate in mammalian mtDNA. Molecular Biology and Evolution 17:1022-1031.

Goldberg, L. A., P. W. Goldberg, C. A. Phillips, E. Sweedyk, and T. Warnow. 1996. Miminizing phylogenetic number to find good evolutionary trees. Discrete Applied Mathematics 71:111-136.

Goldman, N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a poisson process model of DNA substitution and to parsimony analysis. Systematic Zoology 39:345-361.

Goloboff, P. A. 1993. Estimating character weights during tree search. Cladistics 9:83-91.

Goloboff, P. A. 1995. Parsimony and weighting: a reply to Turner and Zandee. Cladistics 11:91-104.

Goloboff, P. A. 2003. Parsimony, likelihood and simplicity. Cladistics 19:19-103.

Gould, S. J. 1986. Evolution and the triumph of homology, or why history matters. American Scientist 74:60-69.

Grandcolas, P., P. Deleporte, L. Desutter-Grandcolas, and C. Daugeron. 2001. Phylogenetics and ecology: as many characters as possible should be included in the cladistic analysis. Cladistics 17:104-110.

Grant, T, and A. G. Kluge. 2004. Transformation series as an ideographic character concept. Cladistics 20:23-31.

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrica 82:711-732.

Haas, A., and G. G. Simpson. 1946. Analysis of some phylogenetic terms, with attempts at redefinition. Proceedings of the American Philosophical Society 90:319-349.

Harper, C. W. J. 1979. A Bayesian probability view of phylogenetic systematics. Systematic Zoology 28:547-533.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97-109.

Hecht, M. K., and J. L. Edwards. 1976. Ther determination of parallel or monophyletic relationships: the proteid salamanders-a test case. The American Naturalist 110:653-677.

Hendy, M. D., and D. Penny. 1982. Branch and bound algorithms to determine minimal evolutionary trees. Mathematical Biosciences 59:277-290.

Hennig, W. 1950. Grundzuge einer Theorie der phylogenetischen Systematik. Deutscher Zentralverlag, Berlin.

Hennig, W. 1966. Phylogenetic systematics. University of Ilinois Press, Urbana. 370 p.

Hertwig, S, R. O. De Sa, and A. Haas. 2004. Phylogenetic signal and the utility of 12S and 16S mtDNA in frog phylogeny. Journal of Zoology, Systematics and Evolutionary Research 42:2-18.

Hickson, R. E., C. Simon, A. Cooper, G. S. Spicer, J. Sullivan, and D. Penny. 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Molecular Biology and Evolution 13:150-169.

Hillis, D. M., J. J. Bull, M. E. White, M. R. Badgett, and I. J. Molineaux. 1992. Experimental phylogenetics: generation of a known phylogen. Science 255:589-592.

Hillis, D. M., J. J. Bull, M. E. White, M. R. Badgett, and I. J. Molineaux. 1993. Experimental approaches to phylogenetic analysis. Systematic Biology 42:90-92.

Hillis, D. M., J. P. Huelsenbeck, D. L. Swofford. 1994. Hobglobin of phylogenetics? Science 369:363-364.

Hubbs, C. L. 1944. Concepts of homology and analogy. American Naturalist 89:289-307.

Huelsenbeck, J. P. 1997. Is the Felsenstein zone a fly trap? Systematic Biology 46:69-74.

Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annual Reviews in Ecology and Systematics 28:437-466.

Huelsenbeck, J. P., and D. M. Hillis. 1993. Success of phylogenetic methods in the four-taxon case. Systematic Biology 42:247-264.

Huelsenbeck, J. P., B. Larget, R. E. Miller, and F. Ronquist. 2002. Potential applications and pitfalls of Bayesian inference in phylogeny. Systematic Biology 51:673-688.

Huelsenbeck, J. P., and R. Nielsen. 1999. Effect of nonindependent substitution on phylogenetic accuracy. Systematic Biology 48:317-328.

Huelsenbeck, J. P., and B. Rannala. 1997a. Maximum likelihood of phylogenies using stratigraphic data. Paleobiology 23:174-180

Huelsenbeck, J. P., and B. Rannala. 1997b. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science 276:227-232.

Huelsenbeck, J. P., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Systematic Biology 53:905-913.

Huelsenbeck, J. P., B. Rannala, and J. P. Masly. 2000. Accomodating phylogenetic uncertainty in evolutionary studies. Science 288:2349-2350.

Huelsenbeck, J. P., B. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310-2314.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Jaynes, E. T. 1957a. Information theory and statistical mechanics I. Physics Review 106:620.

Jaynes, E. T. 1957b. Information theory and statistical mechanics II. Physics Review 108:171.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19:101-108.

Kallersjö, M., J. S. Farris, M. W. Chase, B. Bremer, M. F. Fay, C. J. Humphries, G. Petersen, O. Seberg, and K. Bremer. 1998. Simultaneous parsimony jackknife analysis of 2538 rbcL DNA sequences reveals support

for major clades of green plants, land plants, seed plants and flowering plants. Plant Systematics and Evolution 213: 259-287.

Kallersjö, M., Albert V. A., Farris J S. 1999. Homoplasy increases phylogenetic structure. Cladistics 15: 91-93.

Kelly, C. D. 2005. Understanding mammalian evolution using Bayesian phylogenetic inference. Mammal Review 35:188-198.

Kluge, A. G. 1989. Metacladistics. Cladistics 5:291-294.

Kluge, A. G. 1990. Species as historical individuals. Biology and Philosophy 5:417-431.

Kluge, A. G. 1991. Boine snake phylogeny and research cycles. Miscellaneous publications - Museum of Zoology, University of Michigan 178:1-58.

Kluge, A. G. 1997. Testability and the refutation and corroboration of cladistic hypotheses. Cladistics 13:81-96.

Kluge, A. G. 1998a. Sophisticated falsification and research cycles: Consequences for differential character weighting in phylogenetic systematics. Zoologica Scripta 26:349-360.

Kluge, A. G. 1998b. Total evidence or taxonomic congruence: cladistics or consensus classification. Cladistics 14:151-158.

Kluge, A. G. 1999. The science of phylogenetic systematics: explanation, prediction, and test. Cladistics 15:429-436.

Kluge, A. G. 2001. Philosophy and phylogenetic inference: a comparison of likelihood and parsimony methods in the context of Karl Popper's writings on corroboration. Cladistics 17:395-399.

Kluge, A. G. 2003. On the deduction of species relationships: a precis. Cladistics 19:233-239.

Kluge, A. G., J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. Systematic Zoology 18:1-32.

Knight, A., and D. P. Mindell. 1993. Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of Fea's viper. Systematic Biology 42:18-31.

Kolaczkowski, B. and J. W. Thornton. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431:980-984.

Kraus, F., and M. M. Miyamoto. 1991. Rapid cladogenesis among the percoran ruminants: evidence from mitochondrial DNA sequences. Systematic Zoology 40:117-130.

Kullbach, S. 1951. Information theory and statistics. John Wiley and Sons, New York. 399 p.

Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Molecular Biology and Evolution 16:750-759.

Leache, A. D., T. W. Reeder. 2002. Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Systematic biology 51:44-68.

LeQuesne, W. J. 1969. A method of selection of characters in numerical taxonomy. Systematic Zoology 18:201-205.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50:913-925.

Li, S. 1996. Phylogenetic tree construction using Markov Chain Monte Carlo. Ph.D. dissertation, Ohio State University, Columbus.

Lockhart, P. J., M. A. Steel, M. D. Hendy, and D. Penny. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Molecular Biology and Evolution 11:605-612.

Lundberg, J. G. 1972. Wagner networks and ancestors. Systematic zoology 18:1-32.

Maddison, W. P. 2000. Testing character correlation using pairwise comparisons on a phylogeny. Journal of Theoretical Biology 202:195-204.

Maddison, W. P., M. J. Donoghue, and D. R. Maddison. 1984. Outgroup analysis and parsimony. Systematic Zoology 33:83-103.

Margush, T., and F. R. McMorris. 1981. Consensus n-trees. Bulletin of Mathematical Biology 43:239-244.

Mattern, M., D. A. McLennan. 2000. Phylogeny and speciation of felids. Cladistics 16: 232-253.

Mau, B. 1996. Bayesian phylogenetic inference via Markov Monte Carlo Methods. Ph.D. dissertation., University of Wisconsin.

Mau, B, and M. A. Newton. 1997. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carl. Journal of Computational and Graphical Statistics 6:122-131.

Mau, B., M. A. Newton, and B. Larget. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Biometrics 55:1-12.

McAllister, J.W. 1996. Beauty and revolution in science. Cornell University Press, Ithaca, N.Y. 231 p.

McKitrick, M. C. 1994. On homology and the ontological relationship of parts. Systematic Biology 43:1-10.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21:1087-1091.

Milinkovitch, M. C., R. G. LeDuc, J. Adachi, F. Farnir, M. George, and M. Hasegawa. 1996. Effects of character weighting and species sampling on phylogeny reconstruction: a case study based on DNA sequence

data in cetaceans. Genetics 144:1817-1833.

Mitchell, P. C. 1901. On the intestinal tract of birds. Transactions of the Linnean Society of London, Zoology series 2. 8:173-275.

Mitchell, P. C. 1905. On the intestinal tract of mammals. Transactions of the Zoological Society of London 17:437-536.

Mitchell, P. C., C. Mitter, and J. C. Regier. 2000. More taxa or more characters revisited: combining data from nuclear protein coding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). Systematic Biology 49:202-224.

Miyamoto, M. M. 1985. Consensus cladograms and general classifications. Cladistics 1:186-189.

Murphy, R. W., J. Fu, A. Lathrop, J. V. Feltham, and V. Kovac. 2002. Phylogeny of the rattlesnakes (*Crotalus* and *Sistrurus*) inferred from sequences of five mitochondrial DNA genes. *In* Biology of the vipers, G. W. Schuett, M. Hoggren, M. E. Douglas and H. W. Greene (eds.). Eagle mountain publishing, LC. Eagle Mountain, Utah.

Neurath, H., K. A. Walsh, and W. P. Winter. 1967. Evolution of structure and function of proteases: amino acid sequences of proteolytic enzymes reflect phylogenetic relationships. Science 158:1638-1644.

Neyman, J. 1971. Molecular studies: a source of novel statistical problems. *In* Statistical decision theory and related topics, S. S. Gutpa and J. Yackel (eds.). Academic, New York. p. 1-27.

Nixon, K. C., and J. M. Carpenter. 1996. On consensus, collapsability and clade concordance. Cladistics 12:305-321.

Nylander, J. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53:47-67.

Olmstead, R. .G, and J. A. Sweere J A. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Systematic Biology 43:467-481.

Osborn, H. F. 1902. Homoplasy as a law of latent or potential homology. American Naturalist 36:259-271.

Owen, R. 1843. Lectures on the comparative anatomy and physiology of the invertebrate animals. Longman, Brown, Green and Longmans, London. 392 p.

Owen, R. 1847. Report on the archetype and homologies of the vertebrate skeleton. Report of the16[th] Meeting of the British Association for the Advancement of Science. John Murray, London Report 16: pp. 169-340.

Pagel, M. D., Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Systematic biology 53:673-684.

Patterson C. 1982. Morphological characters and

homology. *In* Problems of phylogeny reconstruction, K. A. Joysey and A. E. Friday (eds.). Academic Press, London. p. 21-74.

Patterson C. 1988. Homology in classical and molecular biology. Molecular Biology and Evolution 5:603-625.

Penny D., Hendy M., Steel M. A. 1992. Progress with methods for constructing evolutionary trees. Trends in ecology and evolution 7:73-79.

Popper K. R. 1968. The logic of scientific discovery. Harper and Row, New York. 544 p.

Popper K. R. 1997. The demarcation between science and metaphysics. *In* The philosophy of Rudolph Carnap, P. A. Schilpp (ed.). Open Court, La Salle. p. 183-226.

Posada D., Buckley T. R. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53:793-808.

Posada D., Crandall K. A. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Prim R. C. 1957. Shortest connection networks and some generalizations. Bell System Technical Journal 36:1389-1401.

Quicke D. L. J. 1993. Principles and Techniques of Contemporary Taxonomy. 1st ed. Blackie Academic and Professional, Bishopbriggs, Glasgow. 328 p.

Rannala B., Yang Z. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. Journal of molecular evolution 43:304-311.

Reeder T. W. 1995. Phylogenetic relationships among phrynosomatid lizards as inferred from mitochondrial ribosomal DNA sequences: substitutional bias and information content of transitions relative to transversions. Molecular phylogenetics and evolution 4:203-222.

Remane A. 1956. Die Grundlagen des natürlichen system der vergleichenden anatomie und phylogenetik. 2d edition. Geest and Portig, Leipzig. 364 p.

Remane A. 1961. Gedanken zum problem: Homologie und Analogie, Preadaptation und Parallelität. Zoologischer Anzeiger 166:447-70.

Richards R. 2002. Kuhnian Values and Cladistic Parsimony. Perspectives on Science 10:1-27.

Richards R. 2003. Character Individuation in Phylogenetic Inference. Philosophy of Science 70:264-279.

Rieppel O. 1992. Homology and logical fallacy. Journal of evolutionary biology 5:701-715.

Rodrigo A. G. 1992. A modification to Wheeler's combinatorial weights calculations. Cladistics 8:165-170.

Rodriguez F., Oliver J. L, Marin A., Medina J. R. 1990. The

general stochastic model of nucleotide substitution. Journal of theoretical biology 142:485-501.

Rokas A., Williams B. L, King N., Carroll S. B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425:796-804.

Ronquist B., Huelsenbeck J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Ronquist F. 2004. Bayesian inference of character evolution. *Trends in ecology and evolution*, 19:475-481.

Roth V. L. 1984. On homology. Biological Journal of the Linnean Society 22:13-29.

Roth V. L. 1988. The biological basis of homology. *In* Ontongeny and systematics, C. J. Humphries (ed.). Columbia University Press, New York. p. 1-26.

Roth V. L. 1991. Homology and hierarchies: problems solved and unresolved. Journal of evolutionary biology 4:167-194.

Roth V. L. 1994. Within and between organisms: replicators, lineages, and homologues. *In* Homology: the hierarchical basis of comparative biology, B. K. Hall (ed.). Academic Press, New York. p. 301-337.

Salter L. A, Pearl D. K. 2001. Stochastic search strategy for estimation of maximum likelihood phylogenetic trees. Systematic biology 50:7-17.

Sankoff D., Cedergren R. J. 1983. Simultaneous comparison of three or more sequences related by a tree. *In* Time warps, string edits and macromolecules: the theory and practice of sequence comparison, D. Sankoff and J. B. Kruskal (eds.). Addison-Wesley, London. p. 253-264.

Sennbald B., Bremer B. 2000. Is there a justification for differential a priori weighting in coding sequences? A case study from rbcL and Apocynaceae s.l. *Systematic biology*, 49:101-113.

Shannon C. 1948. A mathematical theory of communication. Bell System Technical Journal 27:379-423.

Siddall M. E, Kluge A. G. 1997. Probabilism and phylogenetic inference. Cladistics 13:313-336.

Simmons M. P, Pickett K. M, Miya M. 2004. How meaningful are Bayesian support values? Molecular Biology and Evolution 21:188-199.

Simmons M. P, Zhang L. B, Webb C. T, Reeves A. 2006. How can third codon positions outperform first and second codon positions in phylogenetic inference? An empirical example from the seed plants. Systematic Biology 55: 245-258.

Sober E. 1988. Reconstructing the past. Parsimony, evolution and inference. MIT Press, Cambridge. 295 p.

Sokal R. R, Rohlf F. J. 1981. Taxonomic congruence in the Leptopodomorpha re-examined. Systematic Zoology 30:309-325.

Sokal R. R, Sneath P. H. 1963. Principles of numerical classification. W.H. Freeman, San Francisco. 359 p.

Sokal R. R, Sneath P. H. 1973. Numerical Taxonomy: the principles of numerical taxonomy. W. H. Freeman, San Francisco. 588 p.

Soltis P. S, Soltis D. E. 2003. Applying the bootstrap in phylogeny reconstruction. Statistical Sicence 18:256-267.

Sporne K. R. 1949. A new approach to the problem of the primitive flower. New Phytologist 48:259-276.

Sporne K. R. 1953. Statistics and the evolution of dicotyledons. Evolution 8:55-64.

Steel M., Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. Molecular Biology and Evolution 17:839-850.

Steel M. A, Hendy M. D., Penny D. 1993. Parsimony can be consistent! Systematic biology 42:581-587.

Stevens P. F. 1980. Evolutionary polarity of character states. Annual Review of Ecology and Systematics 11:333-358.

Swofford D. L. 1990. PAUP: phylogenetic anlysis using parsimony. Version 3. Computer program and manual. distributed by Illinois Natural History Survey, Champaign, Illinois. 302 p.

Swofford D. L. 1998. Phylogenetic Analysis Using Parsimony* (and other methods). Sinauer Associates, Sunderland, Massachusetts. 144 p.

Swofford, D. L., S.H. Berlocher. 1987. Inferring evolutionary trees from gene frequencies under the principle of maximum parsimony. Systematic zoology 36:293–325.

Swofford D. L, Olsen G. J. 1990. Phylogeny reconstruction. *In* Molecular systematics, D. M. Hillis and C. Moritz (eds.). Sinauer, Sunderland. p. 411-501.

Swofford D. L, Olsen G. J, Waddell P J, Hillis D M. 1996. Phylogenetic inference. *In* Molecular systematics, D. M. Hillis, C. Moritz and B. K. Mable (eds.). 2nd edition, Sinauer, Sunderland. p. 407-514.

Tang K. L, Berendzen P. B, Wiley E. O, Morissey J. F., Winterbottom R., Johnson G. D. 1999. The phylogenetic relationships of the suborder Acanthuroidei (Teleostei: Perciformes) based on molecular and morphological evidence. Molecular phylogenetics and evolution 11:415-425.

Tierney L. 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22:1701-1728.

Tillyard R. J. 1921. A new classification of the order Perlaria. Canadian Entomologist 53:35-43.

Tribus M., McIrivine M. C. 1971. Energy and information. Scientific American 225:179-188.

Trueman J. W. H. 1998. Reverse successive weighting. Systematic biology 47:733-737.

Trueman J. W. H. 2002. Reverse successive weighting. About the MacPerl program RSW1.1. Truemanic programming, Canberra.

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.

Turner H., Zandee R. 1995. The behaviour of Goloboff's tree fitness measure F. Cladistics 11:57-72.

Vidal N., Lecointre G. 1998. Weighting and congruence: a case study based on three mitochondrial genes in pitvipers. Molecular phylogenetics and evolution 9:366-374.

Wagner G. P. 2001. The Character Concept in Evolutionary Biology. Academic Press, London. 622 p.

Wagner P. J. 1998. A likelihood approach for evaluating estimates of phylogenetic relationships among fossil taxa. Paleobiology 24:430-449.

Wagner W. J. J. 1952. The fern genus *Diellia*: structure, affinities, and taxonomy. University of California, Publications in botany 26:1-212.

Wagner W. J. J. 1961. Problems in the classification of ferns. *In* Recent advances in botany. From lectures and symposia presented to the IX International botanical Congress, Montreal, 1959, University of Toronto Press. p. 841-844.

Wagner W. J. J. 1969. The construction of a classification. *In* Systematic biology. National Academy of Science U.S.A. Publication 1692:67-90.

Wahlberg N., Braby M. F., Brower A. V. Z., de Jong R., Lee M.-M., Nylin S., Pierce N., Sperling F. A, Vila R., Warren A. D., Zakharov E. 2005. Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. Proceedings of the Royal Society of London, Series B Biological Sciences, in press.

Wahlberg N., Nylin S. 2003. Morphology versus molecules: resolution of the positions of *Nymphalis*, *Polygonia* and related genera (Lepidoptera: Nymphalidae). Cladistics 19:213-223.

Watrous L. E., Wheeler Q. D. 1981. The outgroup comparison method of character analysis. Systematic zoology 30:1-11.

Wenzel J. W., Siddall M. E. 1999. Noise. Cladistics 15:51-64.

Wheeler Q. D. 1986. Character weighting and cladistic analysis. Systematic zoology 35:102-109.

Wheeler W. C. 1990. Combinatorial weights in phylogenetic analysis: a statistical parsimony procedure. Cladistics 6:269-275.

Wheeler W. C. 1992. Quo vadis? Cladistics 8:85-86.

Wheeler W. C, Honeycutt R L. 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. Molecular Biology and Evolution 5:90-96.

Wiley E. O. 1975. Karl R. Popper, systematics, and classification: a reply to Walter Bock and other evolutionary taxonomists. Systematic zoology 24:233-243.

Wiley E. O. 1981. Phylogenetics: the theory and practice of phylogenetic systematics. John Wiley and Sons, New York. 439 p.

Wiley E. O, Siegal-Causey D J, Brooks D R, Funk V A. 1991. The compleat cladist: a primer of phylogenetic procedures. Museum of Natural History, University of Kansas, Special Publications 19. 158 p.

Wilkinson M. 1994. Common cladistic information and its consensus representation: reduced Adams and reduced cladistic consensus trees and profiles. Systematic Biology 43:343-368.

Williams P. L., Fitch W. M. 1989. Finding the minimal change in a given tree. *In* The hierarchy of life: molecules and morphology in phylogenetic analysis: proceedings from Nobel Symposium 70, B. Fernholm, Bremer, K. and H. Jornvall (ed.). Elsevier Press, Amsterdam. p. 453-470.

Williams P. L., Fitch W. M. 1990. Phylogeny determination using dynamically weighted parsimony method. Methods in Enzymology 183:615-626.

Wilson E. O. 1965. A consistency test for phylogenies based on contemporaneous species. Systematic zoology 14:214-220.

Wilson E. O. 1967. The validity of the "Consistency Test" for phylogenetic hypotheses. Systematic zoology 16:104.

Wu C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14:1261-1295.

Yang Z. 1994. Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. Systematic biology 43:329-342.

Yang Z. 1996. Phylogenetic analysis using parsimony and likelihood methods. Journal of Molecular Evolution 42:294-307.

Yang Z., Bielawski J. P. 2000. Statistical methods for detecting molecular adaptation. Trends in ecology and evolution 15:496-503.

Yang Z., Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain method. Molecular Biology and Evolution 14:717-724.