WILEY-BLACKWELL

Cladistics 25 (2009) 1–34

Cladistics

10.1111/j.1096-0031.2009.00265.x

Comparative cladistics

Paul C. Sereno

Department of Organismal Biology and Anatomy and Committee on Evolutionary Biology, University of Chicago, 1027 E. 57th Street, Chicago, IL 60637, USA

Accepted 4 April 2009

Abstract

Current strategies to compare or synthesize morphology-based cladistic hypotheses do not empower individual cladists to (i) understand the origin, authorship, or structure of character data, (ii) efficiently locate and collate previously published character data, or (iii) effectively compare character data from competing cladistic hypotheses. This paper outlines the requisite terminology, methods and indices to effectively compile and compare morphological character data between competing cladistic hypotheses and to isolate and measure the most important factors behind differing cladistic results—character selection and character-state scoring. When the procedures outlined here are facilitated by appropriate software, morphology-based cladistics may overcome long-recognized limitations in data comparison and synthesis.

Morphology-based cladistics, the birthplace of cladistic methods, has been attacked by molecular systematists as hopelessly ambiguous and inevitably dispensable in an era increasingly awash in molecular data (Hedges and Sibley, 1994; Miyamoto and Fitch, 1995; Hedges and Maxson, 1996). More recently, morphologists have generated an equally scathing critique. ''Character delineation'', ''coding'', and ''scoring'' are fraught with ambiguity (Hawkins, 2000; Stevens, 2000; Wiens, 2001; Scotland et al., 2003; Rieppel and Kearney, 2007; Sereno, 2007); and these flawed character data are only partially resampled in subsequent analyses as a result of unstated preferences in ''character selection'' (Poe and Wiens, 2000; Harris et al., 2007; Sereno, 2007).

Coding and scoring morphological characters usually involves greater interpretational and operational complexity than comparable issues of alignment or substitution stemming from a discrete number of nucleotides or amino acids (Day and McMorris, 1992; Mishler, 1994; Wiens, 2001; Kjer et al., 2007). This inherent complexity, in turn, hinders data comparison between analyses (Rieppel and Kearney, 2002, 2007; Sereno, 2007). Data comparison, nevertheless, is key to character selection. Relevant pre-existing character data, after all, must be recognized by way of comparison prior to any decision regarding their utility.

Morphological characters also may have more limited phylogenetic applicability, on average, when extended to distant species than molecular characters (Fig. 1a). Mammalian cranial or dental characters, for example, are critical to phylogenetics within mammalian clades, but have little or no phylogenetic meaning when applied to a snake, a member of their closest extant sister taxon. Morphological characters, in addition, are more often scored for supraspecific terminal taxa (either without explanation or as a summary of particular exemplar species), whereas molecular character data are almost always assembled from single or multiple individuals within a species.

Morphology-based analyses, as a result, often exhibit fractal patterns (Mishler, 2005), with self-similar branching structure hidden within supraspecific terminal taxa (Fig. 1a), in contrast to molecular analyses that are nearly always based on sequences from individuals within a species. A morphology-based tree of life is built from a patchwork of partially overlapping analyses conducted by a vast array of specialists. Morphology-based analyses not only concern relationships among species within a particular clade, but also can focus on

Corresponding author:
E-mail address: dinosaur@uchicago.edu

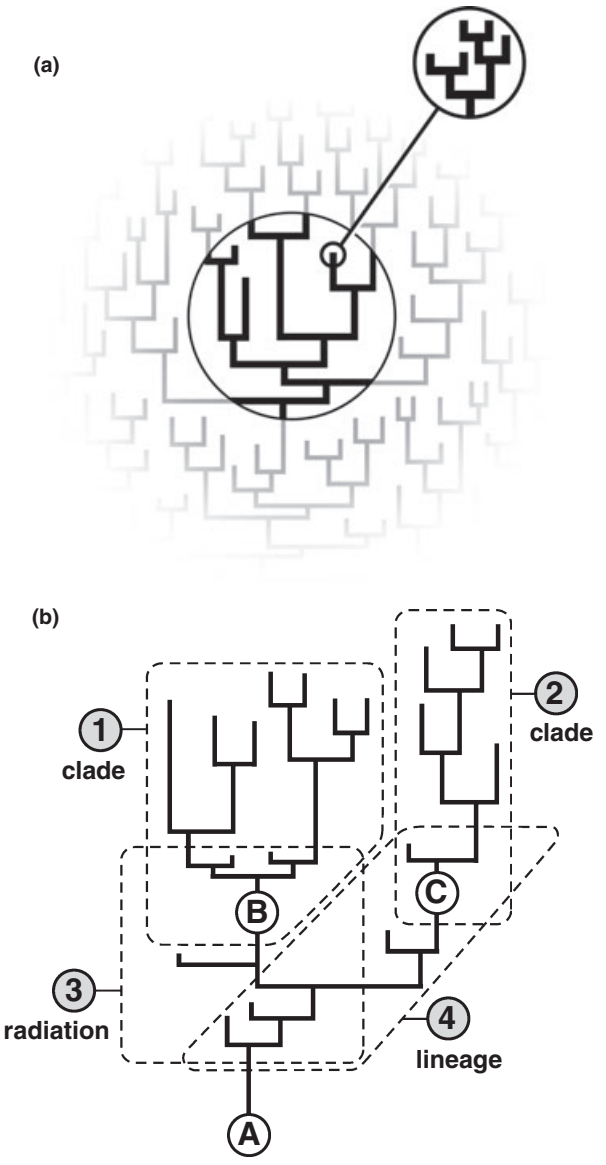

Fig. 1. Nature of morphology-based cladistic analysis. (a) Morphology-based character data apply with least ambiguity to a local area of interest (central circle). Supraspecific terminal taxa are more frequently employed that mask a lower-level (fractal) branching pattern (encircled magnification). (b) Depiction of the complex relationship between four hypothetical morphology-based phylogenetic analyses (shaded 1–4) in a region of the tree of life involving clades (encircled A–C). Vertical lines represent species for simplicity. Analyses 1 and 2 are "clade-focused", using species as terminal taxa to investigate clades B and C, respectively. Analysis 3 is "radiation-focused", using three species exemplars or clades as terminal taxa to maintain focus on the basal radiation of clade A; analysis 4 is "lineage-focused", using two exemplars or clades as terminal taxa to maintain focus on the phylogenetic trajectory from A to C. Analyses 1 and 3 and analyses 2–4 overlap, although none is subsumed within another.

morphological transformations within basal radiations and lineages (Fig. 1b). The resulting cacophony of partially overlapping analyses and terminal taxa in any region of the tree of life constitutes a major challenge for data comparison and, ultimately, character selection. Because of these data limitations and diverse aims, assembling an immense, morphology-based "supermatrix" that would cover major portions of the tree of life at the species level is not possible and has never been a driving ambition for morphology-based systematists.

In this paper, I make no attempt to "solve" the twin problem areas of morphology-based cladistics—character delineation, coding, scoring, and character selection. Rather, I provide the rationale and means better to characterize and measure them.

First, I describe the two basic hurdles to an effective research cycle, as well as current strategies to overcome them. Second, I present a more complete scheme for morphology-based cladistic analysis that includes data comparison. Third, I outline an appropriate terminology and rationale for the three principal areas of *a priori* cladistic analysis, where the aforementioned problem areas are addressed (data compilation, characterization, and comparison). And finally, I briefly look at the history of morphology-based cladistics, from Darwin to the present, in order to understand data comparison in historical context.

The future of morphology-based cladistics as a rigorous discipline, I conclude, lies not only in enhanced character documentation in online image banks, but also in how effectively we understand the nature of character data, and how well we synthesize pre-existing character data. The goal is to approach more closely a true "research cycle" (Kluge, 1991), one that measurably and progressively resolves phylogenetic pattern.

The *bête noire*

Character delineation, coding, and scoring have been explicitly cited for nearly two decades as the prime "*bête noire*" (Pogue and Mickevich, 1990) or "black box" (Patterson and Johnson, 1997) of morphology-based cladistics. "Our pernicious old black box, evolutionary systematics, [is replaced] with a new one, the matrix" (Patterson and Johnson, 1997, p. 361). In the wake of longstanding, largely unresolved cladistic debates involving morphology—such as the origin of turtles or snakes—it has become *de rigueur* to speak of the "poverty of taxonomic characters" or the "poverty of research cycles", and to cast character data in general as a "black box, subject to uncritical assessment and social influence" (Rieppel, 2006, p. 143; Rieppel and Kearney, 2007, p. 95).

Character selection, likewise, has come to the fore as a major problem area (Stevens, 1991; Poe and Wiens, 2000). "Uncritically compiled" morphological data that incompletely sample pre-existing data, or introduce error, according to some morphologists (Jenner, 2004: 307), threaten the very future of morphology-based cladistics as a vigorous discipline. These problem areas

involve *a priori* operations that are described in more detail below.

Besides these two general categories, several other *a priori* factors contribute to differing phylogenetic results, some of which have greater potential influence or are more easily tested than others (Table 1). These include the choice of ancestral character states for ingroup analysis (derived from outgroups or a hypothetical ancestor), choice of terminal taxa, particular analytical assumptions such as ordering or not of character states, and choice of analytical algorithm(s).

Character delineation, coding, and scoring

"Character delineation", also referred to as "character definition" (Pogue and Mickevich, 1990), involves the structure and composition of characters. What constitutes a morphological character? What, at minimum, constitutes a completely formulated character statement? Surprisingly, these questions have not been adequately resolved and remain areas for significant work and consensus (Scotland and Pennington, 2000; Wagner, 2001; Sereno, 2007). Although closely related to coding methods for character states, the focus here is on the structure and composition of the character statement itself. An intuitive operational approach to morphological characters, summarized briefly later in this paper, aims to generate complete, testable character statements that are as comparable as possible across analyses (Sereno, 2007). Closely linked are recent initiatives to develop controlled anatomical terminologies, or character ontologies, and standardized imaging (Ramírez et al., 2007).

A more philosophical approach views morphological characters as "homeostatic property cluster natural kinds", with proper character conceptualization involving causal relations rooted in development (Rieppel and Kearney, 2007). Differing approaches to character delineation, however, do not necessarily generate conflict. When developmental patterns, such as the formation of the dorsal neural tube among vertebrates or sequential condensation of digital rudiments in tetrapods, conflict with phylogenetic patterns based on adult morphology, both ontogenetic and topological sources of error are considered, because neither ontogeny nor adult morphology is free of homoplasy. Morphological character statements, however formulated, operate under the fundamental assumptions of character independence and the mutual exclusivity of character states (Sereno, 2007).

The coding of character states, like the delineation of characters, has not achieved any level of consensus in rationale (Scotland and Pennington, 2000; Wiens, 2001). Do all characters involve transformations? Are some simply "present" or "absent" (neomorphic, *sensu* Sereno, 2007)? Can we mix the two and include "absent" among transformational character states? After review of various coding protocols for multistate characters and for missing or absent structures, Forey and Kitching (2000, p. 77) concluded that "All have advantages and disadvantages but the important points are that, when used in conjunction with binary characters, they can lead to different systematic conclusions." Can morphological character data justifiably be coded in so many ways? Hawkins (2000, p. 35) concluded otherwise, suggesting that "the theoretical framework informing character conceptualization has yet to be fully explored" and that "until there is a clarification of theory inconsistency will remain". Sereno (2007) recommended parsing all such data into neomorphic or transformational character statements. Clearly there are fundamental issues in character coding to resolve.

Table 1
Differing phylogenetic results between morphology-based analyses are due to *a priori* factors (1–6) and choice of analytic algorithm (7), which differ in their testability

Factors	Testability	Comments
Major		
1. Character selection	Difficult	A complete log of all existing character data is almost never presented, because relevant data are located in disparate analyses of different taxonomic scope; previous character data are usually cited on an *ad hoc* basis
2. Character-state scores	Difficult	Systematic scoring differences are usually not logged, because the relevant scores in an opposing analysis are scattered within a data matrix that includes non-comparable data; differences are cited on an *ad hoc* basis
3. Ancestral condition	Difficult	Systematic scoring differences are usually not logged, because the comparable common ancestor may be within the ingroup in an opposing hypothesis, and because the hypothetical common ancestor is usually not specified but rather based on multiple outgroups; differences are cited on an *ad hoc* basis
Minor		
4. Character coding	Difficult	Variation in character coding and presentation can be significant; cited on an *ad hoc* basis
5. Terminal taxa	Easy	Differences in terminal taxa are easily located and their presence evaluated
6. Analytic assumptions	Easy	Ordering of characters is easily altered and evaluated
7. Analytic algorithm	Easy	Maximum parsimony almost always used; data are easily subject to an alternative analytical procedure

Character-state scoring is another area in need of a strategy for improvement. Many erroneous character-state scores survive unchallenged in the published literature. After enumerating many errors in a widely cited opposing analysis, Patterson and Johnson (1997, p. 361) warned that "If the primary work, studying specimens, is not done with as much care and in as much detail as possible, what follows can be weakened and can be almost useless." Yet, in the same paragraph, they admitted that "Few will have the material, the specialized knowledge, and the incentive necessary to check published matrices." Wiens (2001, p. 690) argued that "systematists should explain clearly, and justify, their criteria for selection of characters and their methods of character analysis (i.e., defining, delimiting, coding, and ordering character states)", yet these operations do not seem to be addressed in recent morphology-based analyses any more commonly than there were in decades past. Rieppel and Kearney (2002, p. 60, 2007, p. 97) posited that modern phylogenetic analysis often involves "a superficial approach to comparative anatomy and morphological characters" that "neglects evidence" and results in "irresolvable debates about characters." We seem to lack a well articulated rationale for measuring such scoring differences and determining their phylogenetic significance.

Character selection

Character selection has long been recognized as a critically important procedure that is often little documented or adequately justified (Poe and Wiens, 2000). Poe and Wiens (2000) and Rieppel and Kearney (2002) sought to identify explicit criteria for character selection that test/reject inadequate character data. A more inclusive set of factors involved in character selection, such as unintentional character rejection, has been reviewed and enumerated (Sereno, 2007). As morphological data sets grow in size, complexity and number, it is increasingly difficult to locate overlapping or unshared character data.

When competing hypotheses vary in the scope of the clade under study or the inclusiveness of terminal taxa, character data that are informative in both hypotheses are intermixed with data relevant to only one hypothesis. As a result, the laudable goal of considering all previous character data relevant to a particular hypothesis is rarely achieved. Character data incorporated or rejected from previous analyses are usually difficult or impossible to identify with precision. Unless reanalysing a particular data set, most studies provide only general citations for data sources, and report few details of the complex process of character selection.

Phylogenetic analysis was originally envisioned, in contrast, as an inclusive iterative process (Hennig, 1966). Kluge (1991, 1998) described the "research cycle", in which all previous character data informative to the hypothesis under consideration are included unless rejected by "sophisticated falsification". Besides honing to the spirit of total evidence, Kluge (1998, p. 350) regarded this process as an essential ingredient of scholarly responsibility:

> "Most taxa have been studied previously and with various kinds of data, and such published prior research cannot be omitted from new cladistic studies. To overlook prior hypotheses relevant to one's research shows a lack of scholarship. To purposefully ignore them without cause is authoritarian. … The issue is how honestly the relevant data are surveyed for those synapomorphies that actually have the potential to refute a cladistic hypothesis."

Remarkably, none of the papers cited above that complain of widespread shortcomings in character formulation, scoring or selection has offered the means to measure, or otherwise quantify, the problem. Yet is it reasonable to expect current practice to improve without effective "data similarity" measurements of some kind? And if improvement did occur, how would we recognize it? "Superficial" or "careless" character data, it seems, will not wane unless we formulate an effective means to isolate, display and quantify differences in character data through data comparison.

Debate stagnation

When opposing parties in a phylogenetic debate publish several rounds of analysis and counter-analysis with increasing numbers of terminal taxa and characters, a gloomy state of "debate stagnation" can take hold (Harris et al., 2007, p. 125). Opposing parties unfailingly find the means to defend their most cherished clades or deconstruct those in opposing analyses. Although vigorous debate is often a sign of health in phylogenetics and other sciences, the seemingly intractable views of opposing parties locked in a stagnant debate more closely resemble trench warfare. In the words of one faction in current debate over snake origins, opposing viewpoints appear to devolve into "potentially irresolvable debates about characters, such as can be seen in controversies over bird, tetrapod, and snake origins, to name a few" (Rieppel and Kearney, 2002, p. 60).

Rieppel (2004, p. 90) summarized the situation well:

> "Stagnant debates in contemporary systematics will remain so if systematics does not solve its basic problem. Ever more and faster algorithms can be devised to generate hypotheses of relationships and ever more statistical measures can be used to place confidence limits on those hypotheses. These do not, however, solve the basic problem of systematics, which is the nature of character hypotheses, and the problem of their critical discussion."

Rieppel (2004, p. 89) argued that "We have arrived at the second juncture where the language of science threatens to break down, as is evident from stagnant debates in systematics." The language of character

statements, indeed, is precisely where morphology-based cladists need to begin a rebuilding process (see Discussion; Sereno, 2007). Hot-button "origin" debates are signposts for a more general malaise in morphology-based cladistics that involves character delineation, coding, scoring, and selection.

Without quantitative data-similarity measures that isolate and quantify data-level differences, it is far too easy to generate contradictory phylogenetic results with modest changes in (i) taxon sampling (Hillis, 1998), (ii) character selection, or (iii) character scoring (Stevens, 1991; Poe and Wiens, 2000). Only a few of the papers addressing snake origins, for example, include terminal taxa regarded by an opposing camp as closest relatives to snakes (summarized by Coates and Ruta, 2000). Characters supporting key nodes in an opposing hypothesis can be reduced by simple omission or a claim of character correlation, as has been the case in snake origins with the intramandibular jaw joint and limb reduction (Rieppel and Zaher, 2000). Character data with conflicting distributions are easily neutralized by selective changes in character-state scores, which are difficult to detect in large data sets or in supraspecific terminal taxa scored on particular species exemplars.

The situation regarding character coding, scoring, and selection is, in many regards, parallel to the early days of quantitative cladistics. Cladograms, like current critiques of character data, were initially qualitative endeavours, until Farris, Kluge and others encouraged adoption of quantitative methods to find the shortest tree(s). Hand-constructed qualitative cladograms fell from grace. The shortest tree(s), furthermore, gained particular significance from associated measures of data congruence (consistency index), nodal support, or the increased length of a competing tree. Now a wide range of quantitative tree metrics and tree-sampling strategies are available to understand better the meaning of the shortest tree(s). These *a posteriori* comparisons of analytical results (cladograms, trees), in fact, remain the primary means to assess underlying differences in character data—and, ironically, have come to be called "data exploration" (Grant and Kluge, 2003). Measuring the nature of conflicting results, nevertheless, does little to isolate and understand the root causes for discordance in character data (e.g. character selection, character coding, character-state scoring).

Perhaps truth in character data, in the words of a reviewer of this paper, will simply "out itself" over time? Do we really need quantitative measures of data similarity? For quantitative cladistic analysis, the advent of convenient software transformed the scoring and analysis of character data. To break the present limitations of morphology-based character data, we need not only to renew attention to the nature of character statements and how they are coded, but also to develop computer-assisted, quantitative comparative measures for character data. To reach greater consensus and progressively resolve differences, we must be able efficiently to isolate and measure differences in character data between overlapping analyses.

A molecular analogy

Imagine two labs independently sequencing a mitochondrial gene from the same individuals of several closely related species for phylogenetic analysis. Comparison of the data matrices reveals that (i) some positions were assigned different nucleotides, and (ii) some positions were simply omitted in each matrix. Despite recognition of these discrepancies, their magnitude (or percentage) remained unknown. This circumstance is difficult to comprehend for a conserved sequence without alignment ambiguity among closely related species (dodging complexities common to most molecular data). Misidentification of nucleotides from laboratory error has been reduced to less than one in a thousand (Clark and Whittam, 1992), and exclusion or down-weighting of base positions is explicit (Day and McMorris, 1992; Knight and Mindell, 1993; Thompson, 1999; Wheeler, 2001, 2003; Brudno et al., 2003; Smythe et al., 2006).

Parallel problems, however, are commonplace in morphology-based cladistic analyses: we do not code and score the same morphology with identical characters and character states; we do not select the same set of characters from the pool of available data; and, worse, we have not developed any methods for assessing the magnitude of either of these variables. The foregoing is not an argument for the superiority of molecular data, much of which must contend with issues of alignment, paralogy, and model parameters. Rather, I draw an analogy with the simplest of molecular data sets to highlight the quandary faced by morphology-based systematists.

Limited solutions

Recommendations to reduce or manage the problems sketched above appear to be limited to (i) increasing "explicitness" during data compilation, and/or (ii) urging collegial consensus over anatomical terms and images, exemplar taxa, characters, and character-state scores. These are discussed briefly below.

Greater explicitness

Greater "explicitness"—meaning more character documentation and character analysis—is the clarion call to tame the *bête noire* (Stevens, 1991; Poe and Wiens, 2000; Wiens, 2001; Rieppel and Kearney, 2002, 2007; Jenner, 2004). Simply calling for increased explicitness, how-

ever, is unlikely to effect systemic change. As the number of characters and taxa continues to increase in typical morphological analyses, character documentation seems less common. As Rieppel and Kearney (2002: 60) observed, ''more and more emphasis is being placed upon methods and programs for analyzing data, and less and less on the critical evaluation of the data themselves''. Before the availability of online publishing, lengthy expositions on characters and character states were sometimes difficult or expensive to publish. This is no longer the case. Yet the majority of morphological cladistic analyses incorporate little, if any, individualized character description, documentation, or analysis.

Online image banks and anatomical ontologies have emerged as important developments to document and standardize, respectively, morphological character data. Cell-linked documentation, which at first consisted of a few lines of text linked to a Nexus file (Maddison et al., 1997), has grown to include images and an assortment of auxiliary character information in applications such as Mesquite and in online image banks (Maddison and Maddison, 2003, 2005; Thacker, 2003; Pol, 2004; Blanco et al., 2006; Datta et al., 2007). Image standardization and morphological ontologies (Ramírez et al., 2007) are promising developments that directly addresses the issue of ''explicitness'' with effective documentation of morphological character data.

Data comparison, however, is equally important in achieving greater ''explicitness'' in phylogenetics. How does the most recent matrix compare with that used in a previous competing hypothesis? How much character data was shared between previous competing analyses, and does this shared data yield the same relationships? Or is the unshared portion of data responsible for differing results? And how comparable are character-state scores in shared data between previous competing analyses, and which character-state scores are in conflict? Are scoring differences responsible for differing results? Systematically isolating the underlying causes for differing results necessitates the juxtaposition of the comparable, or overlapping, portions of competing analyses. When we speak of ''combining data sets'', however, we usually are referring to data sets lacking shared characters (e.g. morphological versus molecular; Wiens, 2001).

A recent exemplary attempt at data comparison underscores the persistent shortcomings of current comparative methods. Hill (2005) assembled a matrix of character data compiled from four incongruent hypotheses and original data to evaluate the position of turtles among amniotes. He settled scoring differences with previous analyses by first-hand observations, incorporated character-state scores from previous analyses that differed in the inclusiveness of terminal taxa (e.g. Mammalia versus mammalian genera or subclades), and incorporated his own expanded set of terminal taxa and characters. This exemplary review, nevertheless, leaves unanswered and unmeasured the fundamental comparative questions cited above regarding character selection and character-state scoring. A unique matrix has been generated, one that is doubtless superior in size and accuracy. Yet we evaluate this matrix primarily by *a posteriori* analysis—by comparing its output trees with those from previous analyses.

Consensus

Large-scale collaboration. Large-scale collaborative projects, such as the AToL program of the National Science Foundation (Watanabe, 2002; Pennisi, 2003), often establish gated data sets for collaborators. Significant progress may be achieved in this manner, as the approach encourages experts with overlapping expertise to discuss and reach consensus over exemplar taxa, anatomical ontologies (Mabee et al., 2007; Ramírez et al., 2007), character selection, coding methods, and character-state scores. Historically, phylogenetics weaves consensus over time by comparing, revising, and expanding morphological data sets. Well funded organized collaboration can accelerate these processes.

The vast majority of systematists, nevertheless, are not funded by collaborative programmes and are not contributing directly to collaborative data sets, such as those in the AToL program. And despite the existence of collaborative groups for some clades, independent analyses within these clades continue unabated—as well they should. Morphology-based cladistic analyses, in general, tend to be local in focus (Fig. 1a), with overlapping relations to other analyses (Fig. 1b). Sorting out the phylogenetic relationships of the recently discovered basal toothed turtle, *Odontochelys* (Li et al., 2008), for example, will require a basal amniote matrix similar to that of Hill (2005), incorporating basal data from at least two AToL collaborations that are focusing on clades to which turtles have been allied (archosaurs, squamates).

A supermatrix of thousands of species for millions of base pairs, or even complete genomes, is a goal for some molecular systematists. A matrix at that scale is impractical for morphological data. The 40 000 or so extant species of vertebrates (and increasing numbers of extinct vertebrates), for example, will never be subsumed in a single matrix at the species level for morphological data; the applicability of morphological characters is often more local and human expertise too limited to allow coordinated, efficient coding and scoring of characters across such a vast morphological landscape.

Rather, many independent, morphology-based studies will continue to examine the phylogeny of vertebrates and other organisms across a wide range of scales, from which we derive an estimate of the tree of life (Fig. 1b). Kluge (1998, p. 350) has aptly described phylogenetic analysis as a ''plexus of research cycles … due to the fact

that testable phylogenetic hypotheses occur at different levels of taxonomic inclusiveness''. Large-scale, collaborative character matrices risk being exclusionary or static if the data cannot be efficiently partitioned, modified, compared, and reanalyzed by independent researchers at varying phylogenetic scales.

Consensus data. ''Consensus data'' are based on sequence consensus methods (Day and McMorris, 1992). Overlapping data matrices are decomposed into individual taxon–character entries and then recomposed into a single composite matrix, with character-state conflicts resolved by one of several consensus criteria (Harris et al., 2007). Applying a strict consensus criterion, for example, retains in the final composite matrix only those character states that are in complete agreement across sampled matrices, with other cells scored as indeterminate.

There is a lot to recommend in this approach (Harris et al., 2007), which was used to make a direct comparison between the matrices of protagonists in the debate over turtle origins (Rieppel and Reisz, 1999; Lee, 2001). Like Hill (2005), Harris et al. (2007) compared available matrices to locate disparate character-state scores. They created a strict consensus matrix that included only taxa and characters present in both studies, and were able to show that only a few character-state scores were in conflict. With these few conflicting scores left as indeterminate, the shortest trees exhibited a polytomy favouring neither of the competing hypotheses. Their method thus begins to assess the degree to which the characters and character states of competing hypotheses are in agreement (or conflict).

As currently developed, however, the consensus data method is applicable only to ''scores for exactly the same characters and taxa'' (Harris et al., 2007, p. 129), a severe limitation in comparing most overlapping phylogenies, including the pair of analyses they compared. This paper is devoted in part to overcoming that limitation, allowing recognition and measurement of differences in character selection and character-state scoring with terminal taxa at differing hierarchical levels.

Scheme for morphology-based cladistic analysis

Previous depictions

''Character analysis'' (Kluge and Farris, 1969, p. 9) has long been the preferred term among cladists for the observation, discussion, coding, scoring, weighting, and ordering of characters and character states prior to parsimony analysis and tree construction (Mickevich, 1982; Kitching et al., 1998; Brower, 2000; Rieppel and Kearney, 2002; Wägele, 2004; De Laet, 2005). Cladistic analysis thus was conceived as moving from ''character analysis'' to ''cladistic analysis'', the former including the aforementioned *a priori* procedures surrounding the generation or manipulation of character data and the latter focused on *a posteriori* operations that generate the preferred phylogenetic hypothesis(es) (Neff, 1986; Bryant, 1989) (Fig. 2). The literature that specifically discusses and diagrams these procedures is remarkably thin.

De Pinna (1991) followed Neff's two-part division of cladistic analysis, describing the establishment of ''primary homology'' in character analysis, followed by an analysis of congruence by parsimony to establish ''secondary homology'' (Patterson, 1982; Brower and Schawaroch, 1996). Wheeler (1986, fig. 4) and Wägele (2004, fig. 5.8) have produced more elaborate diagrams for cladistic analysis, but these also can be divided into an initial phase comparable with ''character analysis'' and a secondary phase involving tree reconstruction (Fig. 3).

The principal confusion in the literature involves the second phase, which has been specifically identified as ''cladistic analysis'' (Neff, 1986), ''test of congruence'' (De Pinna, 1991), ''phylogenetic analysis'' (Rieppel and Kearney, 2002), and ''parsimony analysis'' (De Laet, 2005). ''Cladistic analysis'' and ''phylogenetic analysis'', however, are commonly used for the entire process. Kitching et al. (1998, p. 19), for example, stated that

> ''Cladistic analysis consists of three processes: discovery or selection of characters and taxa, coding of characters, and determination of cladograms that best explain the distribution of characters over the taxa.'' ''Parsimony analysis'' and a parsimony-based ''test of congruence'', in turn, are not the only means of generating preferred hypotheses, nor are they the only procedures invoked after ''character analysis''.

The alternative terms *a priori* and *a posteriori* have also been used to describe procedures in phylogenetic analysis. Initially they were applied to character weighting: *a priori* character weighting is implemented prior to tree construction, whereas *a posteriori* character weighting takes into account phylogenetic results (Farris, 1969; Hecht and Edwards, 1976; Kirsch, 1982; Neff, 1986; Wheeler, 1986; Kitching et al., 1998). Other procedures besides character weighting have been invoked before and after tree construction. Neff (1986) and Wägele (2004), for example, favour *a priori* determination of character polarity, a controversial position. Consensus trees and bootstrap analyses, likewise, are routine *a posteriori* operations that involve more than just parsimony analysis.

Although there are historical ambiguities in the use of all of these terms, in this paper I adopt the terms *a priori*, *a posteriori*, and ''phylogenetic computation'', respectively, to describe data-focused operations, tree-focused operations, and the application of a criterion or program to reconstruct phylogenetic trees (Table 2).

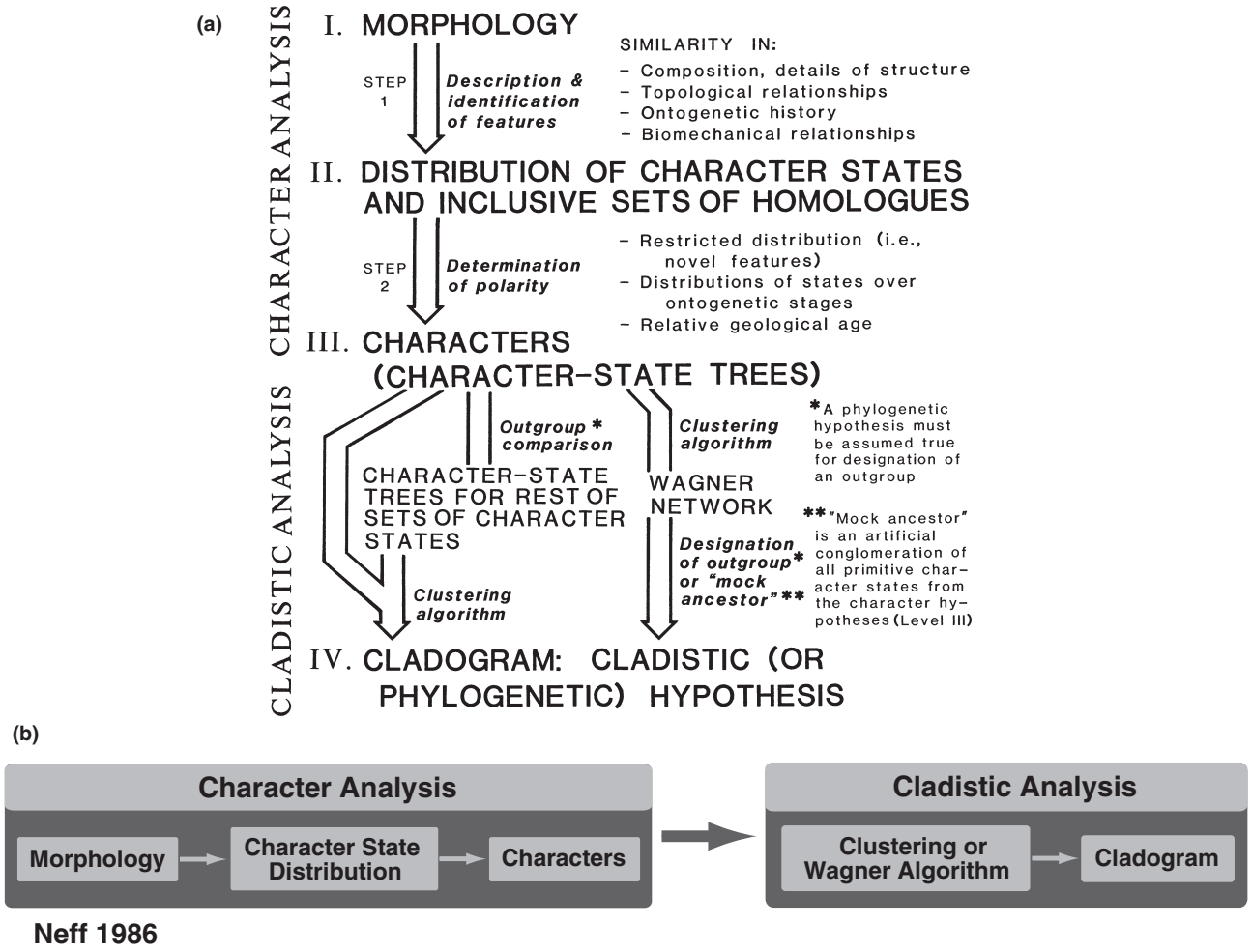

Fig. 2. First scheme for morphology-based phylogenetic analysis. (a) Original diagram (from Neff, 1986). (b) Simplified interpretation of Neff's scheme.

Present scheme

The generalized scheme outlined here recognizes *a priori* and *a posteriori* analysis as distinct procedures within cladistic analysis that have characteristic outputs (Fig. 4; Table 2). Because both procedures can utilize computational phylogenetic algorithms for tree reconstruction, as described below, I place "phylogenetic computation" in an intermediate location rather than regarding it as an *a posteriori* function, as it has long been conceived. The focus of *a priori* analysis is to understand character data, whereas the focus of *a posteriori* analysis is to understand trees. Most or all of what has previously been called "character analysis" is here termed "data compilation", which includes *a priori* procedures that result in the conceptualization and coding of character data, construction of a taxon–character data matrix, and linkage of additional data or graphics to the terminals or cells in that matrix.

Two additional *a priori* procedures are here termed "data characterization" and "data comparison" (Fig. 4). Data characterization seeks to understand the inherent nature of the data assembled for phylogenetic analysis. How much is original to the analysis? In what year were the characters first used? Where are the characters located? How many are binary? These are questions easily answered by a database designed to log pertinent auxiliary information for each character. Data characterization is distinct from the laudable effort in recent years to link images, specimen vouchers, and geographical and other kinds of data to the cells of a taxon–character matrix (summarized by Dettai et al., 2004), or to recent developments that use ontologies and standardized images as the fundamental organizing principle for morphological data (Ramírez et al., 2007). These efforts enhance data compilation, as their primary purpose is the documentation, standardization, and retrieval of character data

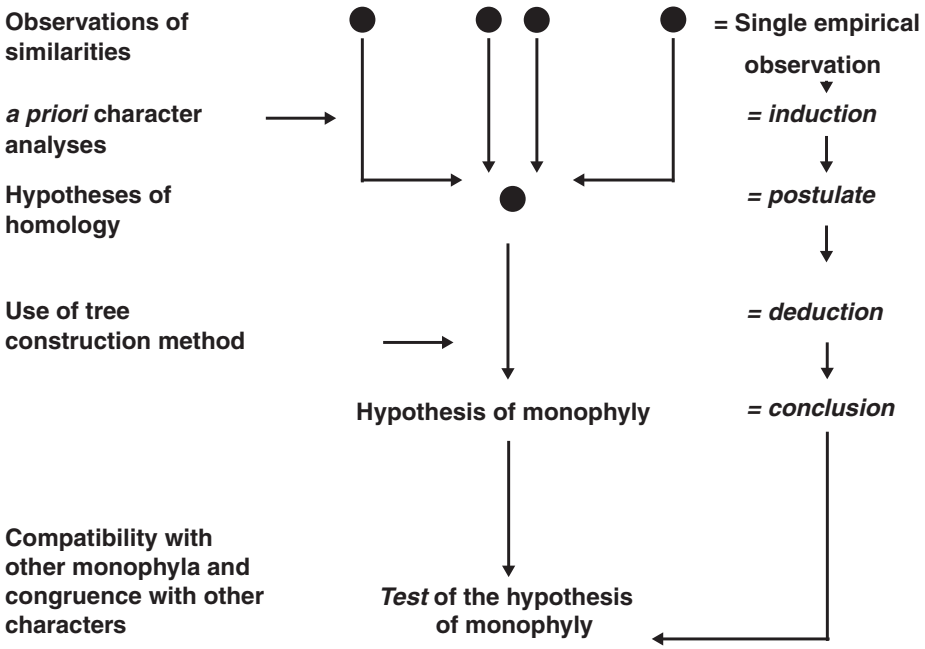

Fig. 3. Recent scheme for morphology-based phylogenetic analysis (after Wägele, 2004).

Table 2
Terms and definitions for phylogenetic analysis as a process (italicized words and acronyms in definitions are defined elsewhere in Tables 1–3)

Term	Definition
Phylogenetic analysis	Analytical, comparative and computational procedures that allow quantitative assessment of the branching history of life based on a taxon–character data matrix
Phylogenetic computation	Computer-assisted computation of phylogenetic pattern with adjustable parameters based on a taxon–character matrix
A priori analysis	*Data compilation, data characterization* and *data comparison* focusing on the generation, nature, and comparison of character data
A posteriori analysis	Evaluation of results from *phylogenetic computation* focusing on the generation, nature, and comparison of phylogenetic trees
Data compilation	Procedures resulting in the discovery, delineation, and coding of character data, construction of a taxon–character data matrix, and linkage of additional data, ontologies, or graphics to matrix cells
Data characterization	Graphical or quantitative summary of character data, including original authorship, previous usage or modification, temporal accumulation, character structure, type or location, character assessment or critique, or the distribution of missing and/or inapplicable data
Data comparison	Isolation and comparison of *relevant data* in an opposing hypothesis (or opposing hypotheses) that involves comparison of data partitions and calculation *of data similarity indices*
Data-similarity indices	Indices (from 0 to 1.00) that measure the degree of similarity of character data from two or more hypotheses (adjusted for *shared taxonomic scope*) regarding (i) the character states of the *comparable common ancestor* (*ASI*); (ii) the characters used in respective analyses (*CSI*, *aCSI*); or (iii) the character states for *shared data* scored in *identical* or *comparable ingroup taxa* (*CSSI*)

(Table 2). Data characterization, in contrast, involves quantitative assessment and graphical summary of character or character-associated information across a data set. This is a simple yet significant distinction, which can easily be incorporated into current database schemes using ontological terms and other auxiliary character information.

Data comparison involves the more complex issue of comparing character data in one analysis with the data in others. How much data is shared? How much is unique to a single analysis? Are differences in phylogenetic branching pattern the result of unshared data? How similar are the results from shared data? How similar are the character-state scores for shared characters? Data comparison, as outlined below, involves the calculation of simple quantitative indices that measure the degree of similarity between characters and character-state scores between one or more analyses.

Table 3
Basic terms and definitions for data characterization (terms as employed in the character database *CharacterSearch*; Fig. 5)

Term	Definition
Original author	First author of a character (or molecular sequence) in a qualitative or quantitative cladistic analysis
Character status	Assessment of the status (*active*, *inactive*) of a character in a cladistic analysis by the author of that analysis; characters are "active" unless uninformative or excluded on the basis of other rejection criteria
Active character	Character assessed as valid and included in a phylogenetic analysis
Inactive character	Character assessed as invalid and rejected in a phylogenetic analysis
Character-accumulation profile	Cumulative number of characters in a phylogenetic analysis using date of first publication
Character structure	Character order and number of states
Character type	Kind of character (molecular sequence, present/absent, etc.)
Character location	Anatomical (or sequence) location of a character
Missing-data profile	Graphical summary of missing or inapplicable data across characters and ingroup taxa
Data-completeness profile	Graphical summary by location of missing or inapplicable data across ingroup taxa

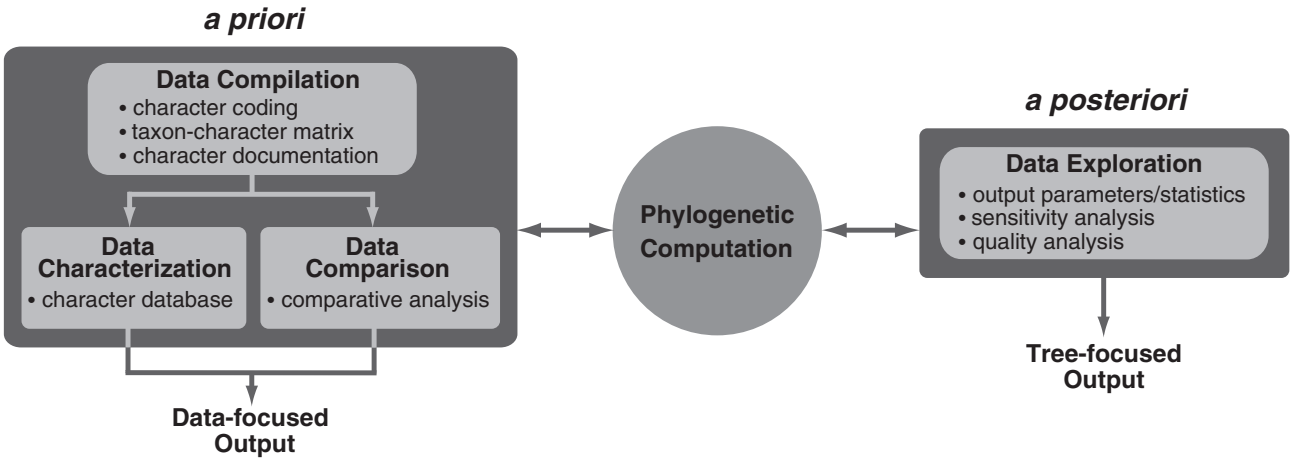

Fig. 4. Proposed scheme for morphology-based phylogenetic analysis. Procedures involved in phylogenetic computation are positioned between *a priori* and *a posteriori* operations. Distinct *a priori* operations include data compilation, data characterization, and data comparison.

Data compilation

Data compilation begins with the conceptualization of character statements followed by character coding and scoring of character states in a taxon–character matrix (Fig. 4). An observation of structural differences in two organisms that otherwise are in close topological correspondence often initiates morphological character conceptualization. An initial proposition, or "primary conjecture of homology", is posited, with the corresponding conditions bound together by a conceptual abstraction, the character. Much ink has been spilt over this process—how characters and their states are initially conceived and formulated, and what criteria are appropriate for testing their validity during this initial phase and prior to the test of congruence, or "secondary homology" (Patterson, 1982; De Pinna, 1991; Poe and Wiens, 2000; Rieppel and Kearney, 2002, 2007; Kearney and Rieppel, 2006).

Completely formulated morphology-based character statements appear to be composed of only four logical components (locators, variable, variable qualifier, character states) brought together in only two specific patterns (neomorphic, transformational) (Sereno, 2007). Character statements, which are open to refutation by way of various comparative, logical, and operational standards or criteria (Rieppel and Kearney, 2002; Kearney and Rieppel, 2006; Sereno, 2007), are then used to score cells in a taxon–character matrix. Character conception, coding, and scoring do not comprise a unidirectional operation, but rather partition an iterative, interlinked process involving refinement of the character and its mutually exclusive conditions (character states). The remainder of data compilation is devoted to character documentation, which can involve explanatory text, images, ontologies, or online information linked to cells in a matrix (Dettai et al., 2004; Mabee et al., 2007; Ramírez et al., 2007).

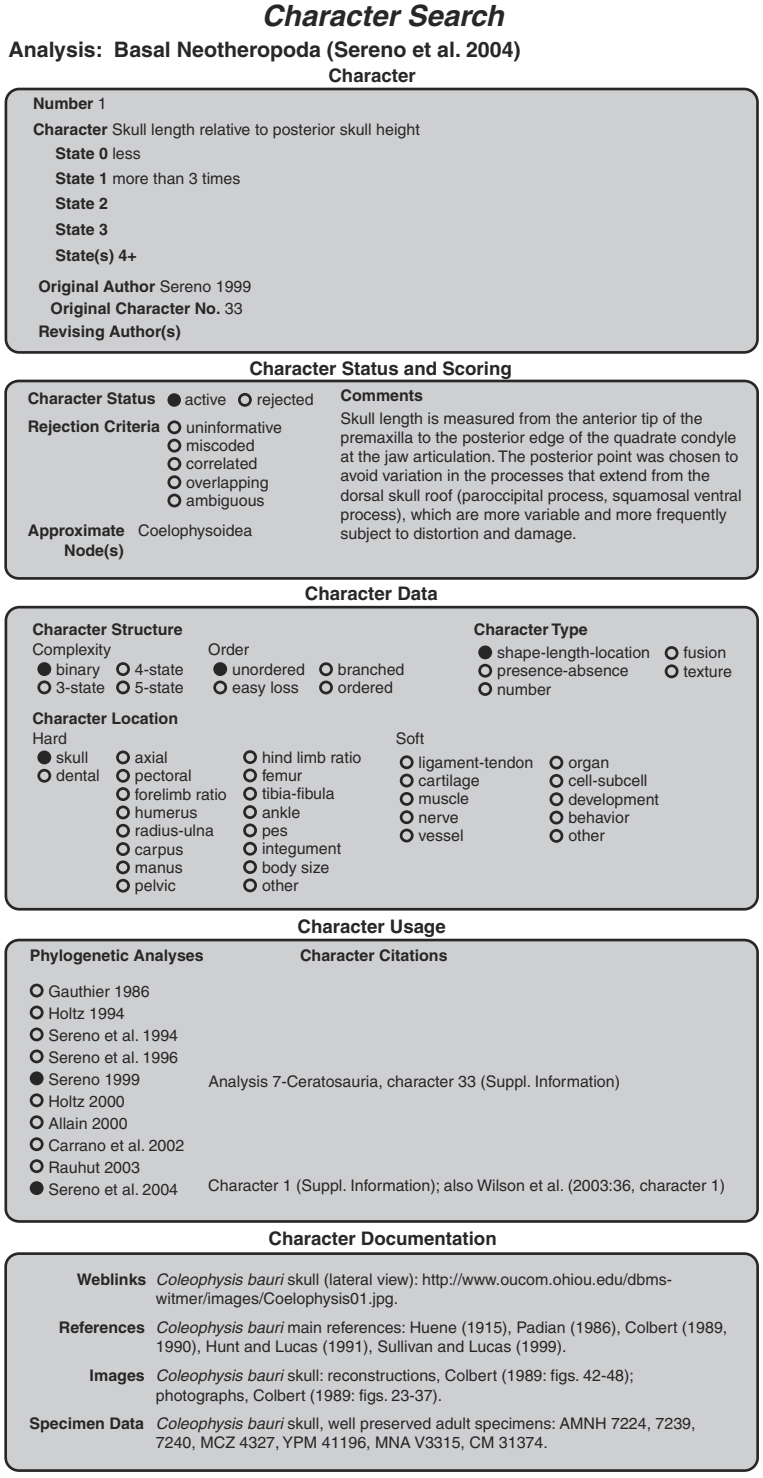

Fig. 5. The first of 169 character records in the file "Basal Neotheropoda" in the database *CharacterSearch* used for data characterization associated with a phylogenetic analysis of theropod dinosaurs (Sereno et al., 2004). This simple database facilitates the logging of relevant character information, generates simple output files and figures (Fig. 6), and can be rendered web-accessible.

A simple database, *CharacterSearch*, allows rapid compiling, searching, and sorting of character data and attendant auxiliary information (Fig. 5). Each character record includes sections for information related to the character, and its character states, status, previous usage, etc., and could be linked to a formalized

ontology. During data compilation, CharacterSearch facilitates logging information related to accepted character data as well as the reasons why other character data are rejected (Sereno, 2007).

Data characterization

Although many databases for characters have been proposed, most are rightly concerned with issues of data compilation, such as character documentation, images, ontologies, voucher specimens and literature citations (e.g. Dettai et al., 2004). As Ramírez et al. (2007, p. 283) stated, "Images are paramount in documentation of morphological data." The main aim of data characterization, in contrast, is better to understand what is here termed "auxiliary character information" for a given data set—the origin, structure, location, temporal accumulation, and completeness of included character data. All but the last utilize auxiliary information linked to the character. Completeness of included character data, on the other hand, is derived from the data matrix. Although rarely explored in current analyses, data characterization (Table 3) could easily be incorporated into current database efforts (Ramírez et al., 2007).

Graphical summaries of auxiliary information about character data and the taxon–character matrix (Figs 6 and 7) can highlight important differences between cladistic hypotheses, as outlined briefly below using as an example a recent analysis of theropod dinosaurs (Sereno, 2004; Sereno et al., 2004). Prior to data characterization, we could not specify the origin of included character data, the identity and amount of rejected character data, the structure and anatomical location of character data, or how these data arose over time in the literature.

Characters

Character status. Character status is the assessment of the validity of a previously-used character statement by the author of a subsequent cladistic analysis (Figs 5 [second box from top] and 6a). Commonly used designations for character status are "accepted" and "rejected". Here, character usage is reframed in terms of activity. A character used in an analysis is regarded as "active", whereas a character that is uninformative or is rejected for whatever reason is labelled "inactive". Character inactivity (rejection) has been justified by a number of criteria (uninformative, variable, high homoplasy, high missing data, doubtful topology, correlated, mixed coding structure, ambiguous; Poe and Wiens, 2000; Rieppel and Kearney, 2002; Sereno, 2007). Recording the justification(s) for character inactivity (Fig. 5 [second box from top]) renders more explicit this *a priori* exercise, which can be summarized graphically (Fig. 6a) and made available online for each character. "Active" character records can be sorted from "inactive" records and summarized graphically (Fig. 6b–f).

Character structure and type. The "structure" of a character refers to how its coding and any special conditions attached to character states. These include the number of character states, weight assignment, and allowable character-state transformations (Fig. 6b). Character "type" refers to the nature of the character statement, which includes presence–absence, quantitative–linear, quantitative–geometric, qualitative-form, topological, etc. (Sereno, 2007; Fig. 6c).

Character authorship. The "original author" of a character is here regarded as the first to use a character in a qualitative or quantitative cladistic analysis (Fig. 6d). In the literature, first authorship can be associated with a much greater range of meanings, such as the first author to describe a character or to coin the current name for a character; the first to incorporate a character into a taxonomic diagnosis; the first to use a character in association with a particular taxon; or the first to recognize the derived state of a transformational character. It has become commonplace, for example, to cite particular authors after characters or synapomorphies in text or in appended character lists. Often the meaning of these citations is unclear. In many cases, multiple citations seem to indicate character usage rather than first authorship. I have separated original authorship from usage, as shown in the example character record (Fig. 5 [first and fourth box from top]).

Marsh (1879), for example, was among the first to recognize the phylogenetic significance of the novel predentary bone in ornithischian dinosaurs, and the first to include it in a taxonomic diagnosis. Sereno (1984) was first to use the predentary bone as a character in cladistic analysis, which is the meaning of "original author" in the character record. The aim here is to track the initial authorship of cladistic character data, in order to track origin and use in subsequent analyses. Recognition of pre-cladistic origin or usage of characters can be noted, such as "based on Marsh (1879)". Restricting "original author" to the initial use of a character in cladistic analysis allows the calculation of a character accumulation profile (Fig. 6f). In the pre-cladistic literature, the phylogenetic meaning of characters and the nature of taxa to which they were linked are often ambiguous and best acknowledged in historical notes.

Two issues will surface regularly in assessing "original author" as described above. First, the use of a character is meant here to apply only to a local region of the tree of life. A predentary bone, for example, appears to have arisen independently in *Hesperornis*, a basal avian. Were this character to prove useful in basal avian cladistics,

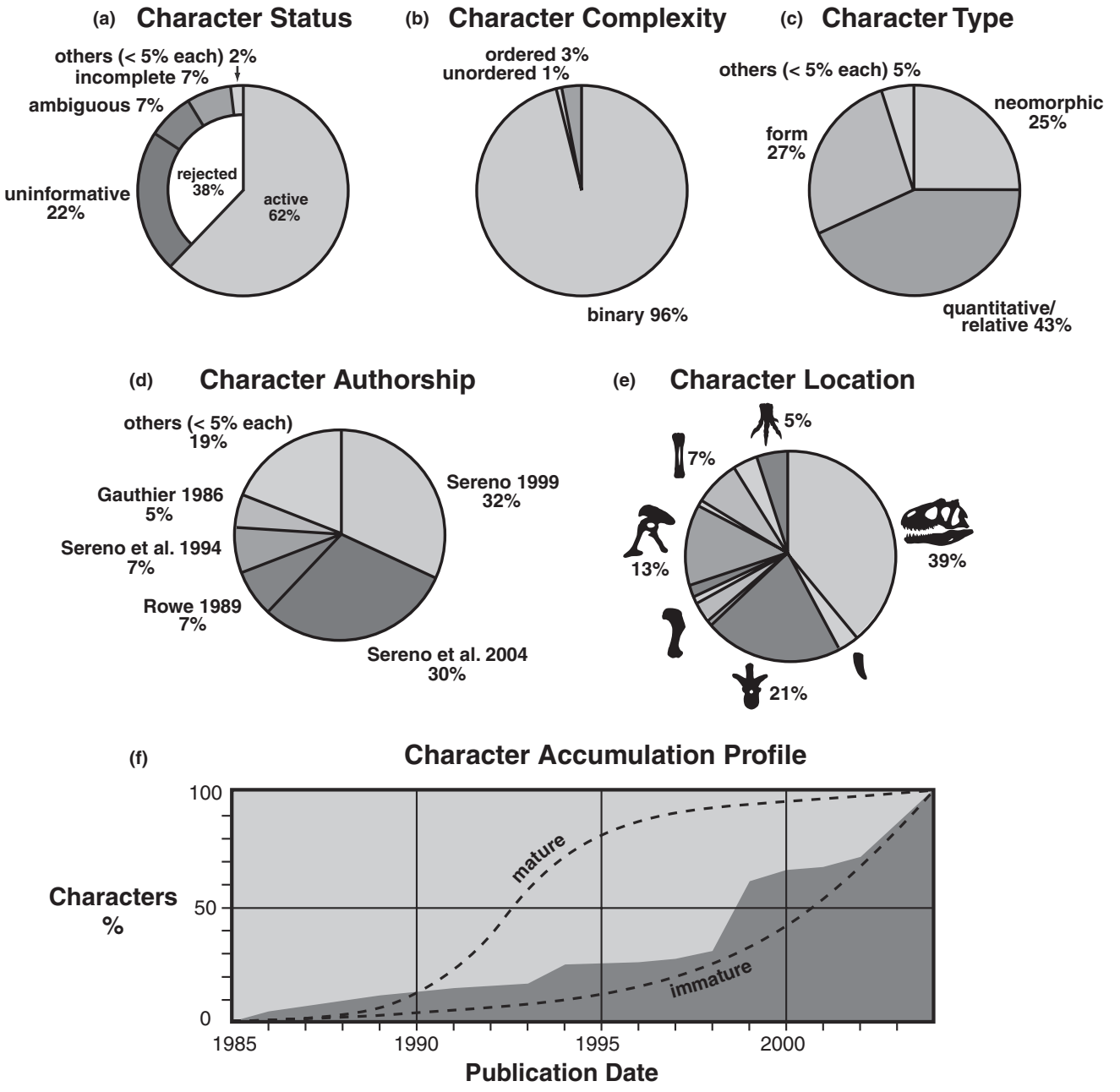

Fig. 6. Data characterization from an analysis of basal theropod dinosaurs (based on Sereno et al., 2004) based on 169 morphological characters. (a) Status of active and rejected character data considered in the course of an analysis. (b) Complexity of active characters. (c) Character type for active characters. (d) Original authorship for active characters. (e) Anatomical location of active character data. (f) Character accumulation profile of active character data based on date of initial publication in a cladistic analysis. Dashed lines show hypothetical curves for ''mature'' versus ''immature'' character data, based on sampling of a finite pool of potential characters.

original authorship would not be Sereno (1984), who used the character in a distant phylogenetic context (ornithischian dinosaurs). Other characters, such as the presence or absence of premaxillary teeth, have arisen many times in the course of tetrapod evolution, and so there may arise some question as to what constitutes the local phylogenetic region. In most cases, however, the original author is reasonably clear, once the general boundaries of a cladistic analysis have been defined.

Second, authors often edit or otherwise modify a character statement used by a previous author. Deciding whether a particular character is ''new'' rather than ''modified from author X'' can be subjective. As a general rule, word order, or other relatively superficial

Table 4
Percentages by anatomical location of characters and missing scores that are depicted graphically in the data-completeness profile (Fig. 7a)

	Number of characters	Character (%)	Character states (missing/total)	Missing data (%)
Cranial	71	42	736/1491	49
Postcranial	98	58	992/2058	48
Skull	66	39	690/1386	50
Teeth	5	3	46/105	44
Axial column	36	21	362/756	48
Pectoral girdle	2	1	25/42	60
Humerus	5	3	46/105	44
Radius–ulna	0	–	–	–
Carpus (wrist)	1	1	15/21	71
Manus	4	2	55/84	66
Pelvis	22	13	217/462	47
Femur	2	1	26/42	62
Tibia–fibula	11	7	86/231	37
Tarsus (ankle)	6	4	59/126	47
Pes	9	5	101/189	53
Whole skeleton	0	–	–	–

changes that do not change the central focus of the character, are best viewed as modification; indeed, this often is necessitated by altering the terminal taxa under consideration. Splitting, merging, or more serious transformations of character statements, however, may not reflect the intention of the original author and are better interpreted as ''new'' characters. Without doubt, the assessment of ''original'' versus ''new'' may be less than obvious. Such interpretive hurdles, nonetheless, will not obscure the fundamental shape of a curve showing character accumulation (Fig. 6f), a novel historical parameter discussed below.

Character location. Character ''location'' is shown here in categories convenient for the hard and soft anatomy of vertebrates (Fig. 5, middle box), although characters could also be linked to an anatomical ontology (Ramírez et al., 2007). Graphical summary (Fig. 6e) presents a better understanding of the location of character data, and character records can be sorted by location. In the example data set for a basal thesopod clade, characters from the axial column can be seen to comprise 21% of all postcranial character data (Fig. 6e, Table 4).

Character-accumulation profile. Morphological character data come from hard and soft tissues and developmental stages, but ultimately comprise a finite universe of potential character data. Characters based on the vertebrate skeleton, for example, are far more limited in number than molecular data, especially when gleaned from imperfect specimens in the fossil record. For deep phylogenetic nodes, in particular, the potential universe of morphological characters is limited. Apomorphies at deep or nested nodes, in addition, should have fewer missing data than terminal taxa based on fossils. It is entirely plausible that intensive research on a particular part of the tree of life could begin to exhaust potential skeletal variation.

A character-accumulation profile uses year of publication by the ''original author'' to plot the cumulative percentage of characters over time for a particular clade (Fig. 6f). The lively debate over turtle origins, for example, is an argument that largely resides at deep nodes within Amniota (Rieppel and Reisz, 1999; Lee, 2001; Hill, 2005; Harris et al., 2007). Are the character data accumulating apace in the latest studies, or have systematists begun to exhaust potential skeletal data at basal nodes within Amniota?

In the example shown here (Fig. 6f), a large influx of character data occurred in the late 1990s, shortly after more complete skeletons were found pertaining to important ingroups. Two-thirds of the character data originated in studies in 1999 and 2004 (Fig. 6d), as also shown by the rising curve in the accumulation profile (Fig. 6f). A hollow curve of this sort suggests that the cladistic problem under consideration may be ''immature'' and open to considerable new character data (Fig. 6f). A ''mature'' profile, on the other hand, would approach a plateau, as new character data become increasingly difficult to discover. Cladistic analyses based on the skeletal characters of extant taxa, in particular, should show maturation of character data over time.

Matrices

Missing information can be an important confounding factor in cladistic analysis, especially when considering extinct or poorly sampled species. Here ''missing information'' or ''missing data'' includes unknown as well as equivocal (ambiguous, polymorphic) character-state scores. Exploring the impact of missing information is usually an *a posteriori* exercise involving the deletion of poorly known taxa (Wilkinson, 1995). *A priori* methods, in contrast, explore the distribution of missing data within the taxon–character matrix. To date, this has been limited to calculating the percentage

of missing data entries per taxon, which is sometimes listed after the last character state. Percentage missing data, nevertheless, can be summarized in more informative ways by using histograms as described below.

Data-completeness profile. A data-completeness profile shows the quantity as well as the completeness of character-state scores by location (Fig. 7a; Table 4). For each data partition, the percentage of character data (above) and missing data (below) are plotted. The profile shows the relative amount of character data in a morphological partition, and how much of that data was scored as missing or equivocal. Data "completeness" is a function of the amount of missing data within a particular data partition. The strongest data partitions contain a significant percentage of the total character data (above) with only a minor percentage of missing data (below).

Analyses involving extant terminal taxa almost always have higher data completeness, because missing data are often a minor component. Extinct terminal taxa, by contrast, can have more missing data than positive character-state scores, as in the example shown here (e.g. pectoral girdle and forelimb; Fig. 7a). The data-completeness profile shows where positive character-state scores are concentrated by partition. In the present example, a greater share of the positive charac-

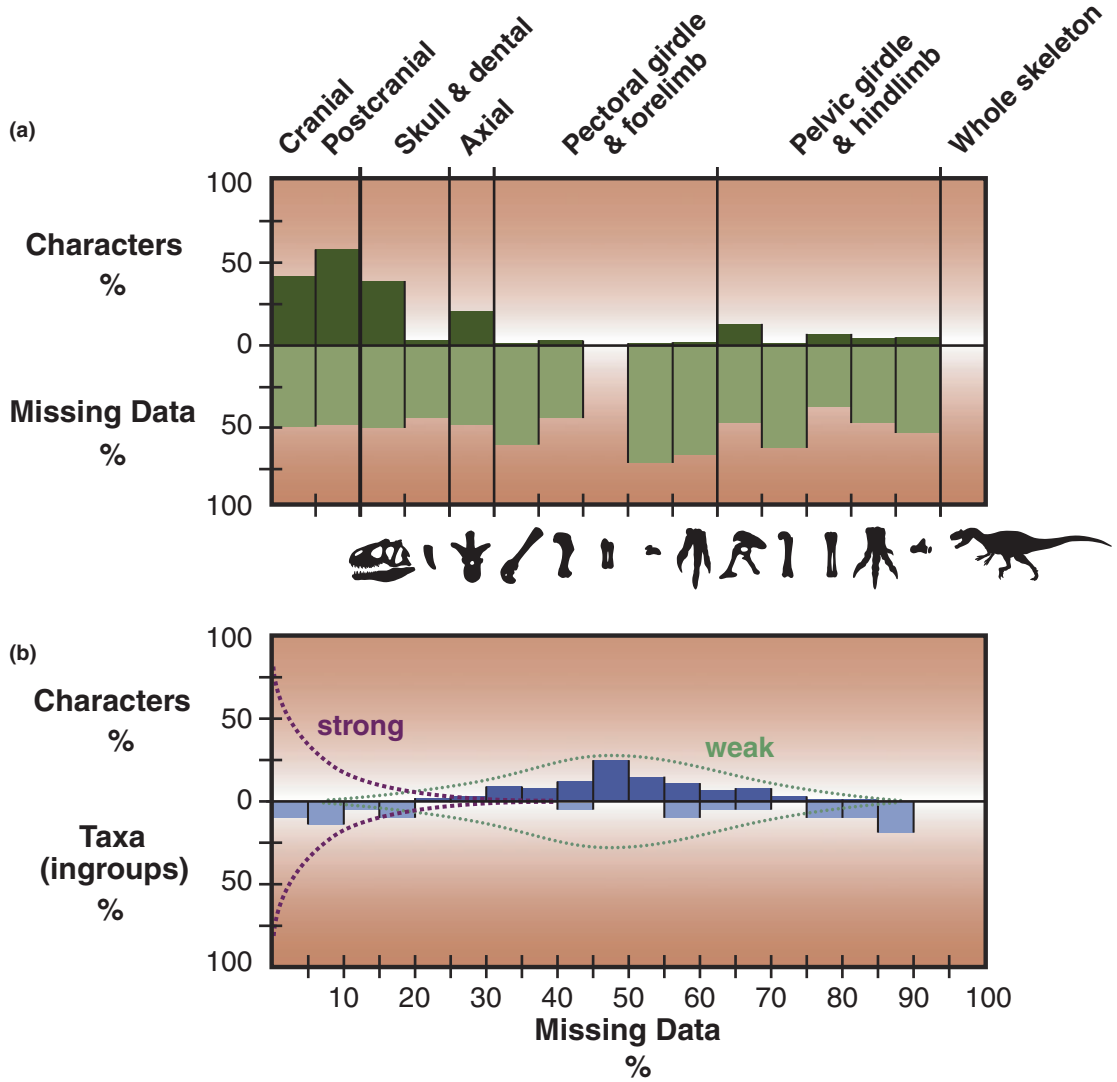

Fig. 7. Data characterization emphasizing missing data with particular application to paleontological data (based on 169 characters in Sereno et al., 2004). (a) Data-completeness profile showing, by anatomical location, both the percentage of character data and missing data (based on Sereno et al., 2004). (b) Missing-data profile showing the percentage of missing character-state scores across characters (above) and taxa (below). Hypothetical "strong" and "weak" missing-data profiles are shown as dotted lines, the former indicating that nearly all characters and taxa have less than 20% missing data.

ter data is located in the postcranium than in the cranium (58%, 42%), and nearly half (48–49%) of both partitions comprise missing character-state scores (Fig. 7a, left pair of columns). Very few positive character data are derived from the pectoral girdle and forelimb, as compared with the pelvic girdle and hind limb. Across the histogram, the percentage of missing data is high, hovering around 50% in all skeletal data partitions.

Missing-data profile. A missing-data profile charts the percentage of missing data (horizontal axis) as a function of characters (above the horizontal) and taxa (below the horizontal) (Fig. 7b; Table 5). Extinct terminal taxa usually have a considerable amount of missing data, which sometimes exceeds positive character-state scores. In the example shown here, approximately 25% of the characters in the matrix are missing as much as 50% of their character-state scores (Fig. 7b). This typically occurs when some terminal taxa are known only from a portion of the skeleton such as the skull. A low bell-shaped curve (dotted line above the horizontal) approximates the distribution of missing data among characters, with very few characters missing less than 20% or more than 90% of character-state scores. This is a ''weak'' missing-data profile common to paleontological analyses.

Plotting missing data across terminal taxa in this example roughly divides these taxa into two subgroups, one with less than 30% missing data and another with more than 75% missing data (Fig. 7b, below horizontal). Based on this plot, it would be interesting to determine if the more complete terminal taxa (with more than 50% positive character-state scores) alone would generate the same relationships as in the initial analysis. Analyses involving extant terminal taxa often have little missing data and almost always generate a ''strong'' missing-data profile that is left-skewed (bold dotted lines to the left; Fig. 7b).

Table 5
Percentages of missing scores among characters and taxa that are depicted graphically in the missing-data profile (Fig. 7b)

Missing data (%)	Character (%)	Taxa (%)
0–4	0	10
5–10	0	14
11–14	0	5
15–29	0	10
20–24	2	0
25–29	3	0
30–34	9	0
35–39	8	0
40–44	12	5
45–49	25	0
50–54	15	0
55–59	11	10
60–64	7	5
65–69	8	5
70–74	3	0
75–79	1	10
80–84	1	10
85–89	0	19
90–94	0	0
95–100	0	0

Data comparison

Systematic *a priori* comparison of character data is rare. Often, no comparisons are made with previous analyses, and character data are simply listed in an appendix. If differences are noted in either character selection or character-state scores, they are often limited to particular characters or nodes. Most comparisons are limited to *a posteriori* cladogram manipulations. Cladogram reconfiguration with constraint trees, for example, assesses the additional length required to reconfigure conflicting results from an opposing analysis. Needless to say, this sheds no light on the character and scoring differences that underlie differing results.

There are two principal impediments to thorough data comparison in morphology-based cladistics: (i) non-overlapping terminal taxa between analyses, and (ii) character data that vary in expression or structure between analyses and are difficult to track and match. This section deals with overcoming the first impediment. The second impediment, which concerns the fairly unstructured manner in which morphology-based characters are formulated, has stimulated recent development of anatomical ontologies (Mabee et al., 2007; Ramírez et al., 2007) and presentation standards for morphological characters (Sereno, 2007).

Taxonomic scope

Unless the authors of two morphology-based analyses have chosen virtually identical terminal taxa, character data that are informative only to one or the other analysis may be intermingled with data that are informative to both. This simple problem—the mixing of shared informative data with data unique to one of two competing hypotheses—is a formidable obstacle to data comparison, especially as data sets grow in size. As a result, little in the way of data comparison occurs, and character lists and taxon–character matrices that only partially overlap continue to accumulate in the literature.

I outline a method below to seek the shared, or common, argument between a pair of such opposing hypotheses to isolate character data in each data set that is informative for that common argument. Particularly meaningful data comparison can be undertaken if we can ''normalize'' competing hypotheses in this way. Character selection and character-state scoring, for

Table 6
Terms and definitions for data comparison (italicized words in definitions are defined elsewhere in the table)

Term	Definition
Taxonomic scope	Unrooted network representing the preferred hypothesis or consensus tree for a set of ingroup taxa
Shared taxonomic scope	*Taxonomic scope* (network) shared by the ingroup taxa of two or more phylogenetic hypotheses
Shared ingroup node	Node identifying the most inclusive node shared by two or more hypotheses as determined by the most proximate *shared outgroup node*
Shared outgroup node	Node identified by the most proximate *comparable outgroup taxon* that lies outside the most inclusive clade of *identical* or *comparable ingroup taxa*
Comparable common ancestor	Ancestral states (or hypothetical common ancestor) at the *shared outgroup node*
Shared or comparable outgroup taxon	Outgroup taxon in two or more hypotheses that is either (i) shared (= same species or supraspecific taxon) or (ii) comparable (= representative of the same ingroup taxon, such as an alternative species exemplar from the same supraspecific clade)
Shared or comparable ingroup taxon	Ingroup taxon in two or more hypotheses that is either (i) shared (= same species or supraspecific taxon) or (ii) comparable (= representative of the same ingroup taxon, such as an alternative species exemplar from the same supraspecific clade)
Unique ingroup taxon	Ingroup taxon with no counterpart in one or more hypotheses under comparison that does not expand taxonomic scope
Explicit exemplar	Species designated as an exemplar for a supraspecific taxon of recognized diversity
Implicit exemplar	Species functioning as an exemplar for a supraspecific taxon of recognized diversity, although not explicitly recognized as such
Relevant data	Character data that remains informative for the *taxonomic scope* under consideration; includes *shared* and *unshared data* partitions in a comparison between hypotheses
Irrelevant data	Character data that are uninformative for the *taxonomic scope* under consideration, including data that may have been informative outside the *shared ingroup node* or within the ingroup for taxa that have been collapsed or removed
Shared data	*Relevant character data* shared by two hypotheses
Unshared data	*Relevant character data* present in only one hypothesis under comparison
Rejected data	User label for *relevant unshared data* in an opposing analysis (or opposing analyses) that are rejected (considered *inactive*) on the basis of rejection criteria (miscoded, correlated, hypervariable, etc.)
Character-coding mismatch	Character-coding variation for the same character between analyses under comparison, which can involve (i) inversion, (ii) differential subdivision, and (iii) non-comparable states
Character-state conflict	Conflictive character-state scores for the same character between analyses involving positive character-state scores (e.g. 0, 1, 2 …)
Character-state disparity	Character-state scores for the same character between analyses involving a positive versus an ambiguous character-state score (e.g. ?, 0/1, –)
Character-state neutralization, swapping, consensus	Conflictive character-state scores for the same character between analyses can be altered to understand their phylogenetic impact by substituting an ambiguous score (neutralization), a score used in an opposing analysis (swapping), or the score used in a majority of analyses (consensus)
Ancestor-similarity index (ASI)	Measure of the proportion of shared character states for the *comparable common ancestor* for *shared data* between two analyses (from 0 to 1.00): $\mathrm{ASI} = (\mathrm{tcs} - (\mathrm{csc} + 0.5(\mathrm{csd})))/\mathrm{tcs}$ tcs = total number of character states compared csc = number of character-state conflicts csd = number of character-state disparities
Character-similarity index (CSI)	Measure of the proportion of shared characters relative to the total number of *relevant* unique characters (total characters) between two analyses (from 0 to 1.00): $\mathrm{CSI} = \mathrm{sc}/\mathrm{tc}$ sc = number of shared characters tc = total number of characters
Available character-similarity index (aCSI)	Measure of the proportion of shared characters relative to the sum of shared and rejected *relevant* unique characters in a previous analysis (from 0 to 1.00): $\mathrm{aCSI} = \mathrm{sc}/(\mathrm{sc} + \mathrm{rc})$ sc = number of shared characters rc = number of rejected characters
Character-state similarity index (CSSI)	Measure of the proportion of shared character states for *relevant shared data* between two analyses (from 0 to 1.00): $\mathrm{CSSI} = (\mathrm{tcs} - (\mathrm{csc} + 0.5(\mathrm{csd})))/\mathrm{tcs}$(abbreviations as in ASI)

example, may be measured effectively as a proportion of total characters and character states, respectively.

Competing hypotheses are "normalized" in two ways: (i) pruning taxa unique to one hypothesis that lie outside the purview of the opposing hypothesis, and (ii) collapsing taxa that are "oversplit" in one hypothesis relative to the opposing hypothesis. Each data set is then reanalysed to isolate and compare character data that remain informative. The goal is to retain as many original terminal taxa as possible in each hypothesis, excluding or collapsing only those that extend beyond the shared phylogenetic problem or network of common nodes—here termed their "shared taxonomic scope" (Table 6).

This "normalization" procedure is not a consensus approach, because there is no attempt to resolve conflicting arrangements for identical or comparable taxa. Nor is this procedure a supertree approach; there is no single output tree, but rather opposing trees composed mainly of identical, or comparable, terminal taxa. The procedure seeks comparable data sets that are informative to the largest set of ingroup taxa originally present in opposing hypotheses.

The simplest "normalization" procedure between two overlapping hypotheses is to select identical terminal taxa, delete all others, reanalyse the data sets, and compare only the residual informative data. This procedure requires no more than a list of terminal taxa for each hypothesis. A case study of tyrannosaurid dinosaurs outlined below closely approximates this simple circumstance.

Retaining only identical, or shared, terminal taxa, however, is often too severe a restriction to allow meaningful comparison between hypotheses. More often opposing hypotheses have chosen closely related, but not identical, terminal taxa for a particular clade. These might include a different species exemplar, one species exemplar versus multiple exemplars, or a species versus a genus or suprageneric taxon. These are here termed "comparable" terminal taxa, which can be retained in each hypothesis with appropriate considerations, while acknowledging at the outset that they are not identical (Table 6). Recognizing comparable terminal taxa that are not identical requires taxonomic or morphological information not present in a simple list of terminal taxa. In the simplest case, one must know that an alternative species exemplar is representative of the same clade. Finally, there can exist terminal taxa present in only one hypothesis that fall between terminal taxa also present in an opposing hypothesis. Retention of these unique nested terminal taxa, to which the character data of an opposing analysis ought to be informative, may affect the relationships of shared or comparable ingroup taxa.

In summary, data comparison between hypotheses with imperfect taxonomic overlap ranges from a simple approach, which limits the comparison to the subset of identical terminal taxa and associated informative data, to a more involved approach, outlined below, which attempts to maximize the comparison to retain as many original taxa and informative characters as possible.

Shared taxonomic scope. The preferred hypothesis for a set of ingroup taxa can be represented by an unrooted network—its "taxonomic scope". The portion of that network that overlaps that of an opposing hypothesis constitutes its "shared taxonomic scope" relative to the opposing hypothesis (Table 6).

Hypothesis 1 has seven terminal taxa divided into two three-taxon clades (taxon A, B) and one singleton (taxon C) (Fig. 8a). For simplicity, we can regard the seven terminal taxa as species. In this case, its

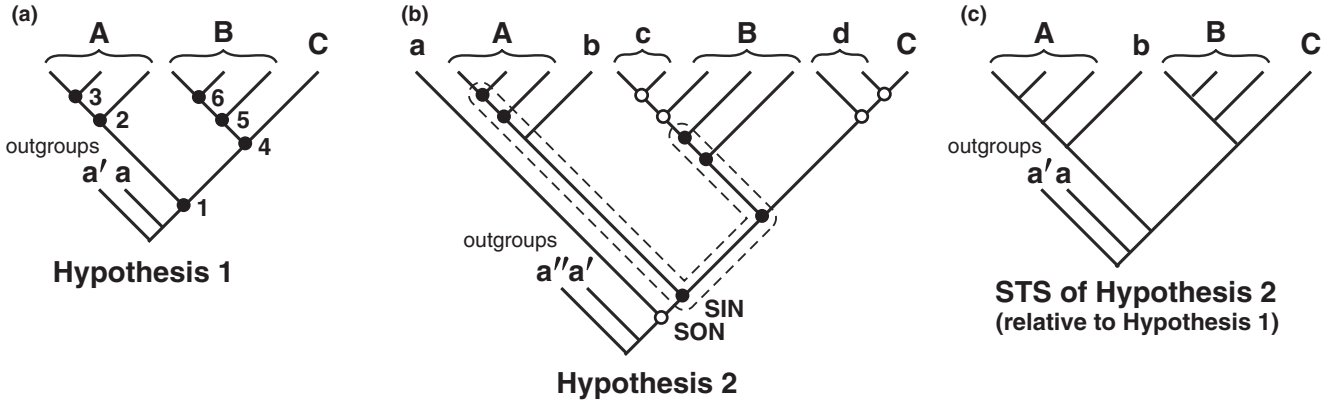

Fig. 8. Taxonomic scope describes the network of a phylogenetic hypothesis. (a) Hypothesis 1 with seven terminal taxa divided into two clades (taxa A, B) and one singleton (taxon C) with successive outgroups (taxa a, a′). (b) Hypothesis 2 with four additional ingroup taxa (taxa a–d), successive outgroups (taxa a′, a″), and shared taxonomic scope with respect to hypothesis 1 (dashed line). (c) Shared taxonomic scope of hypothesis 2 with respect to hypothesis 1 after transfer of taxon a to the outgroup and removal of taxa c and d. Solid dots mark nodes present in hypothesis 1; open dots mark nodes present only in hypothesis 2. Abbreviations: SIN, shared ingroup node; SON, shared outgroup node; STS, shared taxonomic scope.

taxonomic scope is a network of six nodes supported by informative data. The taxonomic scope of hypothesis 1 is thus determined by the immediate outgroup, taxon a, and the degree to which the higher taxa, A–C, are subdivided.

Hypothesis 2 is more expansive in taxonomic scope than hypothesis 1 (Fig. 8b). It includes four additional ingroup taxa (a–d) and several nodes that are not represented in hypothesis 1. It is probable, as a result, that the data set for hypothesis 2 contains characters that would be uninformative to hypothesis 1. To isolate the network of nodes that hypothesis 2 *shares* with hypothesis 1, we need to remove non-overlapping terminal taxa introduced by hypothesis 2. This is determined by (i) the most proximate "shared outgroup node", and (ii) "shared" or "comparable ingroup taxa" (Table 6).

For both hypotheses, the most proximate shared outgroup node is that joining taxon a. Thus taxon a must be removed from the ingroup of hypothesis 2, as otherwise it would expand beyond the lower taxonomic scope (network) of hypothesis 1. Taxa c and d, likewise, have no comparable terminal taxa in hypothesis 1, and thus create additional nodes outside the network of hypothesis 1. Taxa c and d thus must be removed from the ingroup of hypothesis 2. Finally, although taxon b has no comparable taxon in hypothesis 1, it does not add nodes as do taxa a, c, and d. Taxon b lies between nodes 1 and 2 in hypothesis 1, potentially drawing synapomorphies from clade A. Rather than adding network nodes, taxon b subdivides the network of hypothesis 1. Taxon b is here termed a "unique ingroup taxon", or a taxon with no counterpart in an opposing hypothesis, but also one that does not expand the taxonomic scope of that opposing hypothesis (Table 6). The shared taxonomic scope of hypothesis 2 with respect to hypothesis 1, in sum, includes the two three-taxon clades (taxa A, B), the singleton species (taxon C), and taxon b (Fig. 8c). The shared taxonomic scope of hypothesis 1 with respect to hypothesis 2, on the other hand, is hypothesis 1 without modification (Fig. 8a), as its network is entirely within that of hypothesis 2 (Fig. 8b).

The taxonomic scope of hypothesis 1 can be viewed as a network with five internodal locations where added taxa would subdivide, rather than expand, the network (Fig. 8a). In hypothesis 2, taxon b occupies one of those positions, joining the cladogram between two nodes (node 1 and 2) of hypothesis 1; taxon b subdivides, rather than expands, the network (dashed line) shared with hypothesis 1 (Fig. 8b). There are eight possible positions, in contrast, where one or more taxa could be added to hypothesis 1 that would expand its network. Those positions include taxa joining any one of the seven terminal taxa in hypothesis 1, as well as new basal taxa that lie outside the ingroup node (Fig. 8a, node 1). In hypothesis 2, taxa a, c and d join in these positions and expand beyond the nodal network of hypothesis 1 (hollow nodes; Fig. 8b).

For simplicity, the first two hypotheses chosen here to demonstrate the determination of a shared taxonomic scope involve the same ingroup clade and are completely overlapping (Fig. 8a, b). All of hypothesis 1 is thus subsumed within hypothesis 2. The original and shared taxonomic scope are different only for hypothesis 2 (Fig. 8b, c), as the original and shared taxonomic scope of hypothesis 1 are identical with respect to hypothesis 2 (Fig. 8a).

Published cladistic analyses that have overlapping ingroups, in contrast, are usually mutually asymmetrical; the shared taxonomic scope for each hypothesis is more restricted than the taxonomic scope of either original hypothesis (Fig. 9). Three alterations are commonly necessary to achieve shared taxonomic scope: (i) oversplit clades that are represented in both hypotheses need to be collapsed and replaced with a higher-level taxon (marked with an asterisk); (ii) clades that are present in one hypothesis but do not expand taxonomic scope relative to another need to be collapsed and replaced with a higher-level taxon (marked with an asterisk); (iii) unshared exemplars for a particular clade that expand taxonomic scope relative to an opposing hypothesis must be pruned. These operations are shown in a comparison of two hypotheses, which, for simplicity, have the same ingroup node (Fig. 9; outgroups not shown).

Taxon A in hypothesis 1 (Fig. 9a), for example, is oversplit relative to the corresponding taxon in hypothesis 2 (Fig. 9b). Taxon B in hypothesis 1 (Fig. 9a) is also oversplit relative to hypothesis 2, but it uses different exemplars (b, b′) than in hypothesis 2 (b″) (Fig. 9b). Taxon C in hypothesis 1 (Fig. 9a) is not represented in hypothesis 2 (Fig. 9b), but taxon C lies between other taxa (taxon b and taxa d, e) that are present in both hypotheses (or represented by alternative exemplars). Collapsing the exemplars for taxon C is necessary, but maintaining its presence as a singleton terminal taxon does not add network nodes. Taxon d is present in both hypotheses 1 and 2 and thus does not require any modification (Fig. 9a, b). Taxon E in hypothesis 2 (Fig. 9b) is oversplit, but includes among its exemplars the same taxon (taxon e) present in hypothesis 1 (Fig. 9a). Thus the additional exemplar in hypothesis 2 (taxon e′) must be pruned (Fig. 9d).

By collapsing oversplit clades in hypothesis 1 (an asterisk marks a collapsed taxon; Fig. 9c) or removing the unmatched exemplar taxon e′ in hypothesis 2 (Fig. 9d), shared taxonomic scope is achieved for these two hypotheses relative to each another. That is, the network (taxonomic scope) occupied by these hypotheses (dashed line) is comparable (Fig. 9c, d). As in the previous example (Fig. 8a, c), the cladograms depicting

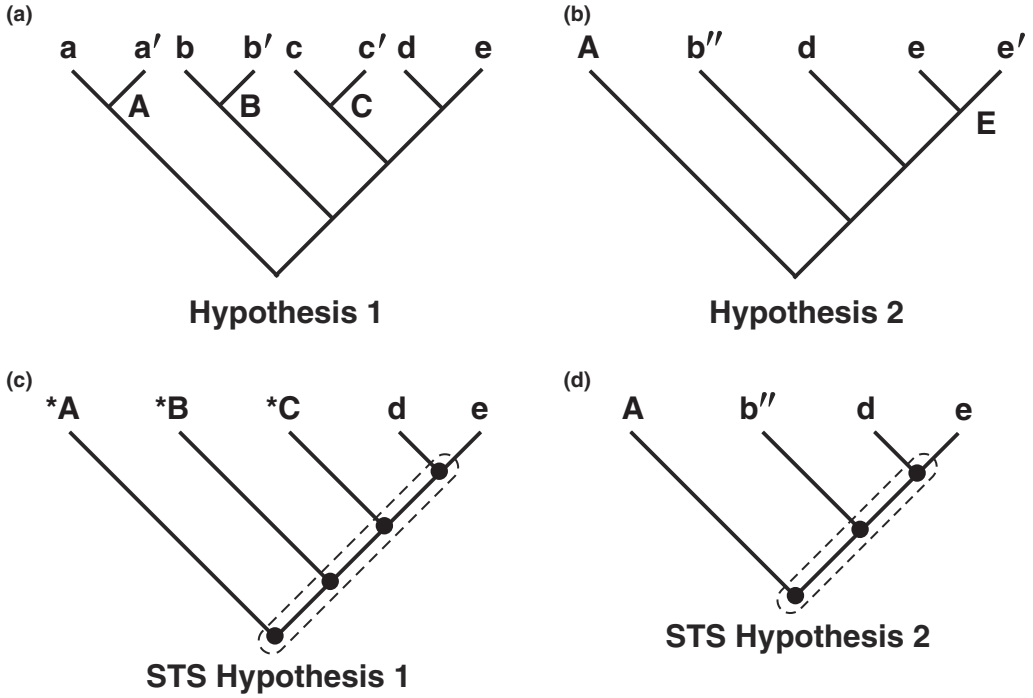

Fig. 9. Determining shared taxonomic scope (STS) of one hypothesis relative to another may involve collapsing an oversplit or unique terminal taxon or removal of a unique sister taxon. (a) Hypothesis 1; (b) Hypothesis 2; (c) STS of hypothesis 1 with respect to hypothesis 2; (d) STS of hypothesis 2 with respect to hypothesis 1. Taxon A in hypothesis 1 is oversplit relative to hypothesis 2. Taxon B in hypothesis 1 is oversplit and uses different exemplars (taxa b, b′) than in hypothesis 2 (taxon b″). Taxon C in hypothesis 1 is oversplit relative to hypothesis 2 but does not add network nodes. Taxon d is present in both hypotheses 1 and 2. Taxon E in hypothesis 2 is oversplit but uses an exemplar (taxon e) present in hypothesis 1, and so the unique sister taxon (taxon e') is removed. Dashed lines indicate the nodal network or its taxonomic scope. An asterisk before taxa A–C indicates a collapsed taxon that was represented by two or more taxa in the original hypothesis.

shared taxonomic scope between these two hypotheses are not identical. Also as in the previous case, the shared taxonomic scope for hypothesis 1 retains a terminal taxon (*C) not represented in the shared taxonomic scope for hypothesis 2, and at least one terminal taxon (e′) was pruned from hypothesis 2. Unlike the previous case, however, differences in taxonomic scope necessitated the collapse of terminal taxa in hypothesis 1 (an asterisk marks a collapsed taxon; Fig. 9c).

Determining shared taxonomic scope is envisioned as a two-step process involving a down-pass to determine the lower shared boundary of the unrooted network (shared ingroup and outgroup nodes) and an up-pass that identifies shared or comparable terminal taxa, collapses oversplit clades, and prunes unique terminal taxa that lie outside the shared network (Fig. 10). Down-pass and up-pass procedures are described in more detail in the following two sections, respectively.

Shared outgroup and ingroup nodes. Outgroup and ingroup nodes exist for any phylogenetic hypothesis with two successive outgroups or a specified ancestral condition (Maddison et al., 1984). Two situations arise, depending on the configuration and number of designated outgroups.

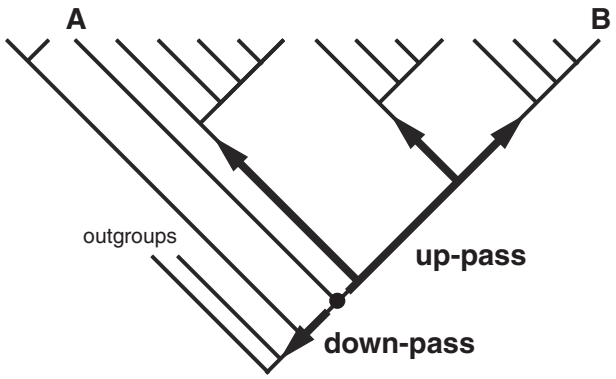

Fig. 10. Diagrammatic depiction of operations needed to determine shared taxonomic scope in a phylogenetic hypothesis that overlaps in taxonomic scope with another. A and B represent comparable terminal taxa from opposing analyses from either side of a basal dichotomy (node = shared ingroup node). The downward arrow depicts the down-pass to identify the shared outgroup node or the lower bound of shared taxonomic space. The upward arrows depict the up-pass that stops the upper bound of shared taxonomic space (terminal taxa or basal nodes of clades that are oversplit in one hypothesis).

In situation 1, with an unpolarized outgroup, a single outgroup (real or all-zero) or clade of multiple outgroups is specified, such that there are no synapomor-

phies at the ingroup node. Monophyly of the ingroup is assumed, as most potential synapomorphies at the ingroup node are more parsimoniously regarded as primitive; the alternative state becomes an autapomorphy of the singleton outgroup, and the character is regarded as uninformative. When comparing two hypotheses under situation 1, identifying the shared outgroup node is a simple matter. Determine the largest clade with identical or comparable terminal taxa clade (the shared ingroup node), and collapse all clades at more inclusive levels into a single outgroup (Fig. 10; taxa A, B). In this case, we are interested only in the interrelationships within the shared ingroup clade, rather than character data that might apply to the shared ingroup node.

Sereno and Brusatte (2009) used situation 1 in their comparison of four analyses of tyrannosaurid dinosaurs. The largest common ingroup clade was identified between these hypotheses, and an outgroup node was established. This limited the problem to the interrelationships of six species. Synapomorphies present at the ingroup node (Tyrannosauridae) in some of the hypotheses were regarded as uninformative when reanalysed for comparison with these more restricted ingroup and outgroup conditions.

Situation 2, with a polarized outgroup, involves either a specified ancestral condition or two or more successive outgroups. In this case, synapomorphies residing at the ingroup node are regarded as informative, because polarity is established by an ancestral condition or pair of successive outgroups (Fig. 10). The down-pass in a comparison between hypotheses must locate the most proximate comparable outgroup taxon (Table 6). One hypothesis may include unique basal taxa between the ingroup and outgroup node of an opposing hypothesis. These unique basal taxa have no counterpart in the opposing hypothesis, yet they lie within its taxonomic scope (network) as explained in more detail below.

In hypothesis 1, taxon a is the immediate outgroup (Fig. 8a). In hypothesis 2, taxon a is within the ingroup, and the most proximate outgroup is taxon a′ (Fig. 8b). The shared outgroup node is identified as the most proximate outgroup taxon between the two hypotheses that lies outside the most inclusive set of identical or comparable ingroup taxa (Table 6). ''Comparable'' here means representative of the same taxon (an alternative exemplar, or a supraspecific taxon including that exemplar) (Table 6). For these two hypotheses, taxon a is the most proximate outgroup, and thus identifies both the shared outgroup and ingroup nodes (Fig. 8a, c).

The configuration of taxa at the base of the cladogram is similar in this two-way comparison. There are no unique ingroup taxa in a basal position between taxon a and clade A. If such a taxon existed in hypothesis 2, the shared ingroup node (SIN) would move toward the shared outgroup node (SON) to include it. The rationale here is uniform—shared taxonomic scope includes all taxa that do not extend the nodal network of another hypothesis. A basal unique ingroup taxon in this position in hypothesis 2 would divide, rather than extend, the network of hypothesis 1. Thus synapomorphies that may have resided at node 1 in hypothesis 1 (Fig. 8a) may now be split in hypothesis 2, with some residing at a more inclusive node that unites this hypothetical unique basal taxon and the remainder of the ingroup. To exclude this unique basal taxon, and instead place the SIN where it is shown at present (Fig. 8b), would be an error, as it would exclude comparable data in hypothesis 2 that may well reside at node 1 in hypothesis 1. Thus unique basal, or stem, taxa that lie inside the SON must be included within the SIN. The SON ultimately determines the location of the SIN, the lower bound of shared taxonomic scope, not the reverse.

Once the SON has been located, the ancestral states at that node represent the comparable starting point in the analyses being compared, or the comparable common ancestor (Table 6). Those states are best determined using any phylogenetic constraints on outgroups or character ordering that was specified by the original author(s).

Collapsing oversplit ingroup taxa. Now the upper bound of shared taxonomic scope can be determined (hypotheses 1 and 2; Fig. 9a, b). Shared (identical) ingroup taxa, of course, are included as part of shared taxonomic scope (taxon d; Fig. 9c, d). The real issue concerns ingroup taxa that are oversplit in one hypothesis compared with another (taxa A–C; Fig. 9a). An oversplit taxon has two or more terminal taxa that are represented by a single terminal taxon in another hypothesis. An oversplit taxon can include (i) two or more exemplars represented by a clade in an opposing hypothesis (taxon A versus a, a′; Fig. 9a, b); (ii) two or more exemplars represented by a different exemplar in an opposing hypothesis (taxa b and b′ versus b″; Fig. 9a, b); or (iii) two or more terminal taxa that belong to a unique ingroup clade that lies within the taxonomic scope of the opposing hypothesis (taxa c and c′ in hypothesis 1, which lies between taxa b″ and d, e in hypothesis 2; Fig. 9a, b).

Any of these three common situations are handled the same way—the oversplit taxon must be collapsed to a single comparable terminal taxon, labelled accordingly with a higher taxon name (marked by an asterisk), and assigned the composite ancestral character states at the collapsed node. This comparable taxon is then inserted as an additional line into the original taxon–character matrix as a replacement for the oversplit terminal taxa. Collapsing terminal taxa as described removes unshared network nodes and limits the hypothesis to the bounds of shared taxonomic space.

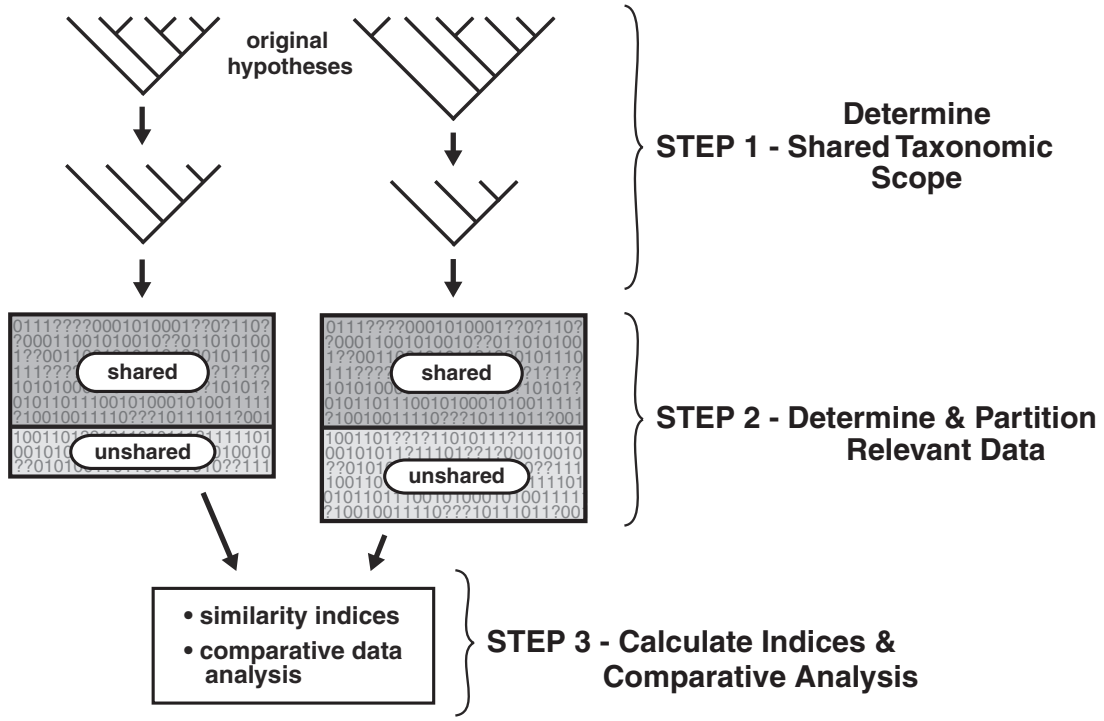

Fig. 11. Graphical summary of three principal steps in data comparison between two hypotheses, which culminated in the calculation of data similarity indices and comparative data analysis.

It must be borne in mind that oversplit taxa and taxa that must be pruned are not the same. Oversplit taxa overlap the nodal network of another hypothesis but must be collapsed to become comparable (taxa a, a′ versus A; Fig. 9a, b). Terminal taxa that do not overlap but rather add nodes to the network shared with another hypothesis must be pruned, as in the case of a novel sister taxon (e′) to a shared (identical) terminal taxon (e; Fig. 9a, b).

Two circumstances are encountered when collapsing a taxon to make a comparable composite terminal taxon with respect to another hypothesis. If the composite terminal taxon was used as a label for a node on the original cladogram, its use as a proxy matches its designation in the original hypothesis (taxa A, B; Fig. 9a, b). If there is no such designation on the original cladogram or in associated text, an appropriate name must be designated for the composite taxon. In both cases, an asterisk can be attached to the composite terminal taxon when used on a cladogram or inserted into a taxon–character matrix (A*–C*; Fig. 9b). Tagged in this way, the collapsed, composite taxon is clearly differentiated from unmodified terminal taxa present in the original hypothesis.

Isolating relevant data. The data matrix is reanalysed to isolate informative character data, once the shared taxonomic scope of a hypothesis has been determined, its terminal taxa adjusted by removal or collapse as needed, and the taxon–character matrix adjusted (deletion of taxa, insertion of composite terminal taxa). This would include all the original informative character data for a hypothesis only if the taxonomic scope and shared taxonomic scope of a hypothesis are one in the same (hypothesis 1; Fig. 8a). Otherwise, some of the original informative data is rendered uninformative by removal or collapse of original terminal taxa. The terms ''relevant'' and ''irrelevant'', rather than ''informative'' and ''uninformative'', are used here to distinguish these data partitions to avoid confusion with their original information content (Table 6). The aim of determining shared taxonomic scope is to allow isolation of relevant character data and to set aside irrelevant character data that are applicable (informative) only in one hypothesis (step 1; Fig. 11).

Data partitions

Shared character data. Once the relevant data partition has been isolated in a comparison of one or more hypotheses, the individual character statements are compared to identify those that are shared (step 2; Fig. 11). Shared character data are composed of equivalent character statements that comprise some proportion of the total relevant character data for two or more hypotheses (Table 6).

Determining which characters are shared is a process that could be facilitated by software that allows scrolling and drag-and-drop linkage between opposing character lists (see below). At present I use information logged during character compilation into the character records in *CharacterSearch*. Overlapping characters can be sorted by selecting authors in the Character Usage box (Fig. 5), and then further sorting these character records to include only relevant character data.

Unshared and rejected character data. Any relevant character data that are not among shared character data are considered "unshared" (Table 6). In most comparisons of opposing hypotheses, unshared character data exist in each hypothesis with no matching character data in the opposing hypothesis.

"Unshared" is a relatively non-judgmental term to identify the unmatched partition of relevant character data between hypotheses. If the comparison of hypotheses is done by an author of a new analysis after review of the character data in a pre-existing hypothesis, then "rejected" may be an appropriate label for unshared data in an opposing analysis that is relevant but purposely excluded. The author in this case has considered all relevant character data in an opposing hypothesis and then has rejected this unshared partition.

Data similarity indices

Terminology. A character's "code" and "character coding" refer to the structure of a character and the process or methodology used to create character statements, respectively (Sereno, 2007). In contrast, a character's "score" and "character scoring" refer to the particular state assigned to the cells of a taxon–character matrix and the process or methodology used to designate character states, respectively. The distinction between character structure and the assignment of particular character states is often obscured in the literature.

The terminology used to describe differences or conflicts in character coding or scoring is even less well established. Here I describe differences in the coding of the same character as "character coding mismatch", as when one analysis divides a particular character into more states than another, or when the same character states are assigned to different values.

The principal focus of this paper, however, is to measure similarity in character selection and scoring, rather than character coding. Character-state differences occur either as conflictive positive character states, here termed "character-state conflict", or as a character state versus a question mark or its equivalent (gap, polymorphism, etc.), here termed "character-state disparity". Character-state conflict can be "real" or "apparent". Real character-state conflict occurs when a particular cell is assigned a different character state, one that refers to a different condition; apparent character-state conflict occurs when a given cell is assigned a different character state that refers to the same condition (due to reversal or shuffling of character-state assignments during character coding). Character-state disparity occurs when a given cell is assigned a positive score in one analysis and an ambiguous score in another. Positive character states are here distinguished from ambiguous character states scores, in which character-state evaluation is not decisive (polymorphism, ambiguity, lack of information, or transformation). Terminological aspects of character coding and scoring are carefully distinguished here (Table 6) because they are critical to a meaningful evaluation of similarity.

When comparing the shared taxonomic scope of two hypotheses, there are three principal reasons for differing phylogenetic results: (i) different character states for shared characters in the hypothetical ancestor; (ii) unshared characters present in only one hypothesis; and (iii) different character states for characters present in both analyses. Four data similarity indices are introduced below to measure the degree of similarity between two hypotheses in these three regards. The ancestor-similarity index (ASI) measures the similarity of the ancestral condition at the shared outgroup node and thus assesses similarity in outgroup assumptions. The character similarity indices (CSI, aCSI) measure the percentage of relevant character data that is shared between hypotheses and thus assess character selection. The character-state similarity index (CSSI) measures the degree to which shared character data have identical character states in shared or comparable terminal taxa and thus assesses character scoring. All range from 0.00 (no similarity) to 1.00 (identity) (Table 6).

Ancestor-similarity index. The ASI is a measure of the of the similarity of the character states at the shared outgroup node (comparable common ancestor) as scored in two opposing analyses:

$$\mathrm{ASI} = (\mathrm{tcs} - (\mathrm{csc} + 0.5(\mathrm{csd})))/\mathrm{tcs}$$

where

tcs = total number of character states
csc = number of character-state conflicts
csd = number of character-state disparities.

An ASI of 1.0 indicates complete agreement of character-state scores at the shared outgroup node between two analyses, which usually is the case only when hypotheses under comparison have selected hypothetical primitive (all-zero) outgroups. The ASI is expected to be significantly less than 1.0 when different outgroups are selected by opposing hypotheses, or when

the same outgroups are selected but are scored differently. Character-state differences between analyses either comprise character-state conflict (e.g. 0 versus 1) or character-state disparity (e.g. 1 versus ?). The penalty in the former case is 1 and in the latter case is 0.5. For binary characters, an unknown character state (?) is operationally equivalent to polymorphism (0/1), which theoretically differs from a single state score (0 or 1) 50% of the time. The same penalties apply to different character-state scores for multistate characters. These penalties are then subtracted from the total number of character states, the index measuring their proportion of total character states.

Character-similarity index. The character-similarity index (CSI) is a measure of the proportion of shared relevant character data between two analyses:

$$\mathrm{CSI} = \mathrm{sc}/\mathrm{tc}$$

where

sc = number of shared characters between two data sets

tc = total number of characters between two data sets.

This index compares characters rather than character states. A CSI of 1.0 indicates complete overlap of characters between two analyses, an improbable circumstance involving morphological data. In this case, the total number of characters (tc) equals the number of shared characters (sc). In a comparison of two analyses, the CSI decreases from 1.0 as the proportion of shared characters decreases relative to the total number of unique characters. The CSI is thus a measure of character selection between comparable hypotheses.

The available character-similarity index (aCSI) is a measure of the proportion of shared relevant character data between two analyses divided by the sum of shared plus rejected relevant character data:

$$\mathrm{aCSI} = \mathrm{sc}/(\mathrm{sc} + \mathrm{rc})$$

where

sc = number of shared characters between two data sets

rc = number of rejected characters between two data sets.

This index focuses the comparison to determine the extent to which available character data are included in a subsequent analysis. A CSI of 1.0 indicates complete incorporation of characters from a previous analysis, an improbable circumstance with morphology-based characters. In a hypothetical comparison between two analyses adjusted for shared taxonomic scope, 10 characters were present in an available analysis, to which 90 additional characters are added (Fig. 12a, left pair of columns). The CSI is 0.1, and the aCSI is 1.0, as indicated (Fig. 12a, under second column). In this example, all available characters in the available analysis are resampled.

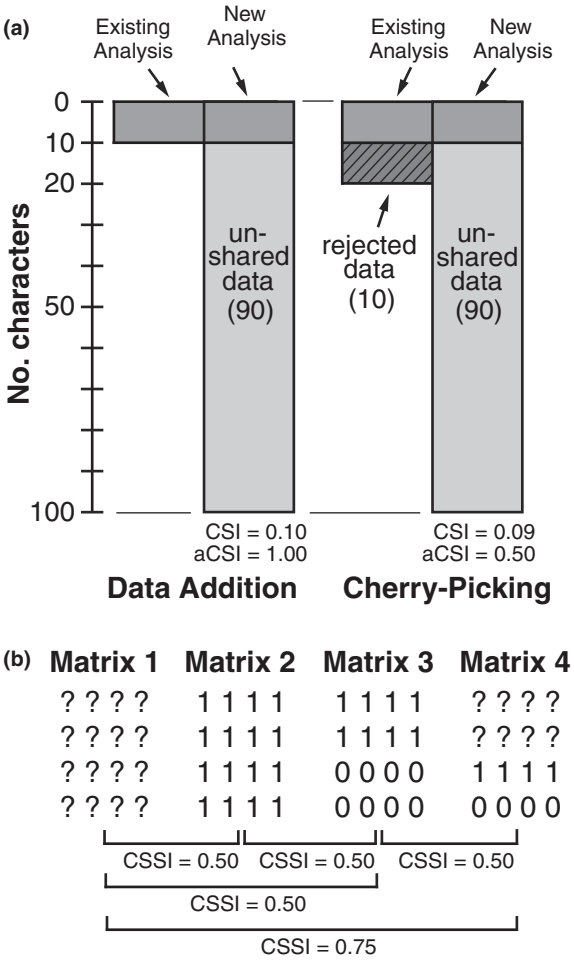

Fig. 12. Calculation of the character and character state similarity indices. (a) Character similarity and available character-similarity indices (CSI, aCSI). "Data addition" scenario in which 90 unshared characters are added to 10 existing characters resulting in a low CSI (0.10) but perfect aCSI (1.00). "Cherry-picking" scenario in which 90 unshared characters are added to 10 out of 20 existing characters (10 are rejected) resulting in a similar CSI (0.09) but much lower aCSI (0.50). (b) Comparison of the character-state similarity index between four hypothetical matrices showing the cumulative effects of character-state disparity and character-state conflict.

If 20 characters are present in the available analysis but only 10 are resampled (10 rejected), only half of the initial character data is resampled. The aCSI is 0.5 rather than 1.0, a value that indicates significant "cherry-picking" of available character data (Fig. 12a, right pair of columns). If 90 additional characters are also added in the subsequent analysis, the CSI is 0.9, just slightly less than 1.0. The aCSI is thus sensitive to available data and is not affected by the addition of significant unshared data in a subsequent analysis. aCSI, in other words, specifically measures the degree of

incorporation of available character data, whereas CSI measures the overall similarity of character data between two analyses.

Character statements vary in morphological data even when describing the same feature (Sereno, 2007). Whether character statements in opposing data sets are similar enough to be regarded as the same is sometimes unclear. Minor differences in wording or swapping of character-state assignments are not cause to register character-state conflict. It is character selection, rather than coding, that is at issue in this index. The same is true if "present" or "absent" character states are sequestered in a neomorphic character in one data set but in another combined with transformational states as a single character. In this case, two characters in one data set would be matched with a single character in another data set. Some of the intricacies of matching characters between data sets are covered elsewhere (Sereno and Brusatte, 2009).

Character-state similarity index. The CSSI is a measure of the similarity between corresponding character states in opposing analyses. It uses the same formula as the ASI, but considers the similarity of character states of comparable characters rather than comparable ancestral conditions:

$$\mathrm{CSSI} = (\mathrm{tcs} - (\mathrm{csc} + 0.5(\mathrm{csd})))/\mathrm{tcs}$$

where

tcs = total number of character states
csc = number of character-state conflicts
csd = number of character-state disparities.

A CSSI of 1.0 indicates identical character-state scores for the same characters in identical or comparable terminal taxa between two analyses, an improbable circumstance given the complexity of morphological data and variation between analyses in specimens, preservation, and interpretation. As with the ASI, the CSSI tracks character-state conflict (e.g. 0 versus 1) and disparity (e.g. 1 versus ?) with similar penalties of 1.0 and 0.5, respectively. If identical or similar character states are assigned different numbers between two data sets, this must be taken into account so that apparent scoring differences are not logged as if they were real.

Comparing cells having missing data with cells having resolved character states generates a CSSI of 0.5 (Fig. 12b, matrix 1 versus matrix 2 or 3). Cells with resolved character states in two matrices that conflict 50% of the time have a similar CSSI (Fig. 12b, matrix 2 versus matrix 3) as do cells with some combination of missing data and character-state conflict (Fig. 12b, matrix 3 versus matrix 4). Because there are two sources for scoring differences (conflict, disparity), comparisons between two analyses will not generally generate an identical CSSI in a comparison with a third analysis (Fig. 12b; CSSI = 0.5 for matrix 1 versus matrix 3 and matrix 3 versus matrix 4; CSSI = 0.75 for matrix 1 versus matrix 4).

Scheme for data comparison

Data comparison, in summary, is presented here as a three-step procedure (Fig. 11). In step 1, the aim is to determine the shared taxonomic scope of opposing hypotheses. That requires locating shared outgroup and ingroup nodes and removing or collapsing terminal taxa as needed (Figs 8 and 9). This procedure involves both a down-pass and up-pass comparison between hypotheses, to ensure complete overlap of what can be viewed as the lower and upper boundaries of the shared ingroup network (Fig. 10).

Step 2 is devoted to partitioning character data (Fig. 11). Once the shared taxonomic scope has been determined, character data for each hypothesis are reanalysed to isolate data informative to both hypotheses—here termed relevant character data. These characters are further divided into unshared and shared partitions, which are present in one or both of the data sets, respectively (Table 6). When a new analysis is compared with a previously published data set, the unshared data partition of the pre-existing analysis may be termed the "rejected" data partition (Table 6).

Step 3 involves measuring the similarity between data sets. Four data-similarity indices can be calculated based on the character states of the comparable common ancestor for shared data (ASI), the shared versus unshared data partitions (CSI, aCSI), and the character states of shared character data in comparable terminal taxa (CSSI) (Table 6).

Exploring the differences between opposing cladistic hypotheses can occur at three levels (Table 7). The first, or taxon, level considers shared taxonomic scope and clarifies the degree to which opposing hypotheses overlap. The second, or character, level considers all aspects of characters, including the number that are shared and the phylogenetic signal coming from different data partitions. The third, or character-state, level is focused on character states in shared character data that highlight discrepancies and determine how these affect phylogenetic signal.

Discussion

Morphology-based cladistics must confront four challenges to successfully combat the recent and pointed critiques that its procedures are hopelessly flawed: (i) reduction of unnecessary variation in the presentation and structure of character data between hypotheses; (ii) effective compilation and evaluation of pre-existing character data; (iii) isolation and quantification of the

Table 7
Strategies for data characterization and comparison using shared taxonomic scope (STS), character-similarity indices (character-similarity index, CSI; available character-state similarity index, aCSI; character-state similarity index, CSSI), and character-state swapping or consensus. For more meaningful (normalized) results, all approaches start with level I, with a consideration of the terminal taxa chosen and the calculation of shared taxonomic scope. Level II is concerned with character statements, not how they are scored. Level III is the most detailed and involves the examination of character-state scores in shared character data

Level	Focus	Specific questions	Approach
I	Taxonomic scope	1. How much overlap is there between opposing phylogenetic studies?	Determine STS
		2. To what degree are they addressing the same problem?	
II	Character statements	3. Is character selection responsible for conflicting results?	Determine CSI and aCSI
		4. How much character data is shared between key studies?	Explore the impact of character selection using data partitions
		5. How much character data exists for a particular phylogenetic problem?	Compile relevant character data and auxiliary character information;
		6. Who created available character data and how did it accumulate over time	chart character authorship, etc
III	Character-state scores	7. Are differing character-state scores responsible for conflicting results?	Determine shared character partition and calculate CSSI
		8. How similar are character-state scores for shared characters?	Explore impact of character-state assignments by neutralization, swapping, or consensus
		9. How would character-state neutralization, swapping or consensus affect results?	

root causes for differing phylogenetic results such as character selection and scoring; and (iv) management of all the above in the face of analyses that differ substantially in taxonomic scope. Each of these challenges is addressed below.

Reducing variation in character statements

Variation in character coding and formulation is a much more significant hindrance to comparison of morphological than molecular data sets. In large analyses that differ significantly in taxonomic scope, locating overlapping character data is often a formidable task that must be done by hand. The primary reason for these difficulties is that little consensus exists over character coding or even what minimally constitutes a cladistic character (Forey and Kitching, 2000; Hawkins, 2000; Poe and Wiens, 2000).

Logical structure of character data. A morphological character has sometimes been viewed as a logical proposition (Woodger, 1952; Kluge, 2003). I have argued elsewhere that a morphological character is better understood as a highly patterned linguistic statement—a character statement—that functions in cladistic analysis as a codified variable (Sereno, 2007). All character statements are composed of a character, which locates or describes something, and a statement, which describes its manifestations (Fig. 13a). Character statements appear to be composed of only four discrete functional components: locator (L_n), variable (V), variable qualifier (q), and character state (v_n). A character-statement tree, in analogy to a phase-structure tree in generative grammar (Chomsky, 1965), outlines the structural and functional components of a morphological character (Fig. 13b, c).

Two fundamental patterns exist for character statements: neomorphic and transformational (Fig. 13b, c). Neomorphic character statements, often termed ''presence–absence'' characters, identify a feature that is either present or absent in a terminal taxon. Although some have considered existence and non-existence as transformational states, these states can be understood as an assessment of ''state of being'', as opposed to transformational states of a variable such as ''length''. Transformational character statements, in contrast, identify a variable that is expressed as mutually exclusive, transformational (''changed form'') conditions. Some disparity in morphological character data between analyses could be eliminated if neomorphic and transformational character statements were not intermixed during character coding (Sereno, 2007).

Standards for morphology-based characters. Substantial variation in the formulation and presentation of morphological characters is a major hindrance in data comparison. As suggested above, there are only a few discrete functional components and fundamental patterns for morphological character data. A greater appreciation of the common structure and ''syntax'' of morphological characters (Sereno, 2007) and adoption of anatomical ontologies (Mabee et al., 2007; Ramírez et al., 2007) may substantially reduce variation in morphological character data.

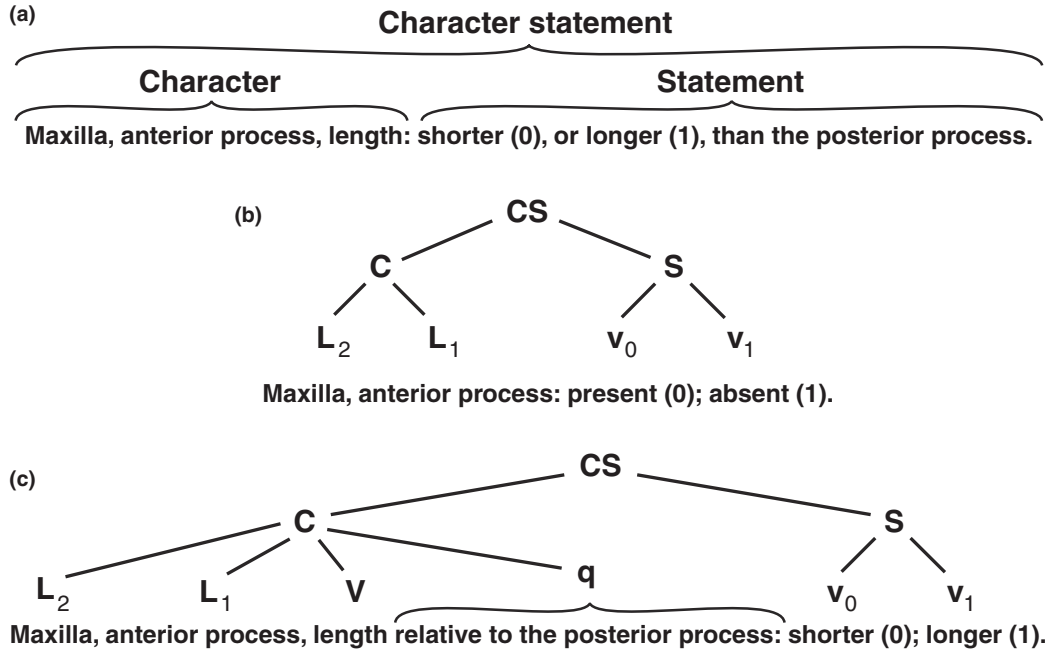

Fig. 13. Logical basis of morphological character data (after Sereno, 2007). (a) Two-part structure (character, statement) of a complete character statement. (b) Neomorphic character statement with character-statement tree. (c) Transformational character statement with character-statement tree. Abbreviations: C, character; CS, character statement; L_1, primary locator; L_2, secondary locator; q, variable qualifier; S, statement; V, variable; $v_{0,1}$, character states.

Evaluating pre-existing character data. Character selection is defined here as the partial resampling of pre-existing character data that are potentially informative for a particular cladistic hypothesis. Whether intentional or accidental, character selection plays a major role in generating different phylogenetic results between opposing analyses. Character selection can be reduced only by more complete data characterization and comparison (Figs 6, 16). For a given cladistic problem, as much pre-existing relevant character data as possible needs to be located and evaluated to be able to answer basic questions in the first two levels of comparison (Table 7). Ideally, cladistic analysis is envisioned as an endless research cycle that fully reassesses available character data (Kluge, 1991, 1998). A ''research spiral'', however, may more aptly characterize morphology-based cladistics as currently practiced; data sets invariably increase in size over time for problems of similar taxonomic scope, but successive analyses usually fall far short of fully digesting or incorporating pre-existing character data.

Case study. A relatively simple case study for data characterization and comparison involves the interrelationships among tyrannosaurids, which have been considered recently in four cladistic studies (Figs 14–16). The phylogenetic hypotheses generated differ in taxonomic scope, outgroup assumptions, number of ingroups, characters, character states, and character documentation (Sereno and Brusatte, 2009). One of the analyses is much broader in taxonomic scope, incorporating as many as 75 ingroups and 638 characters (Holtz, 2004; Holtz et al., 2004; Table 8). Another included only seven ingroups and 35 cranial characters (Currie et al., 2003). None included a comparative analysis of characters or character-state scores. Thus little *a priori* comparative information exists for these analyses, a common situation in cladistic literature. Without further analysis, we are only able to compare the cladograms generated by the hypotheses, rather than *a priori* factors, such as character choice and character-state scores, which are responsible for differing results. These cladograms differ most noticeably in the position of the tyrannosaurids *Alioramus* and *Tarbosaurus* (Fig. 15).

To obtain that comparative information, the three-step procedure outlined above was applied (Fig. 11). Terminal taxa were limited to six monospecific genera of tyrannosaurids that are present in most or all of the analyses (*Albertosaurus*, *Daspletosaurus*, *Albertosaurus*, *Gorgosaurus*, *Tarbosaurus*, *Tyrannosaurus*). We used situation 1 to locate the proximate comparable outgroup, thus limiting the comparative analysis to interrelationships within Tyrannosauridae (see Sereno and Brusatte, 2009). The analyses were re-run with ingroups limited to the aforementioned genera, isolating relevant character data (which remained informative) from other character data that are uninformative within Tyrannosauridae. Relevant characters varied in number among the analyses from 19

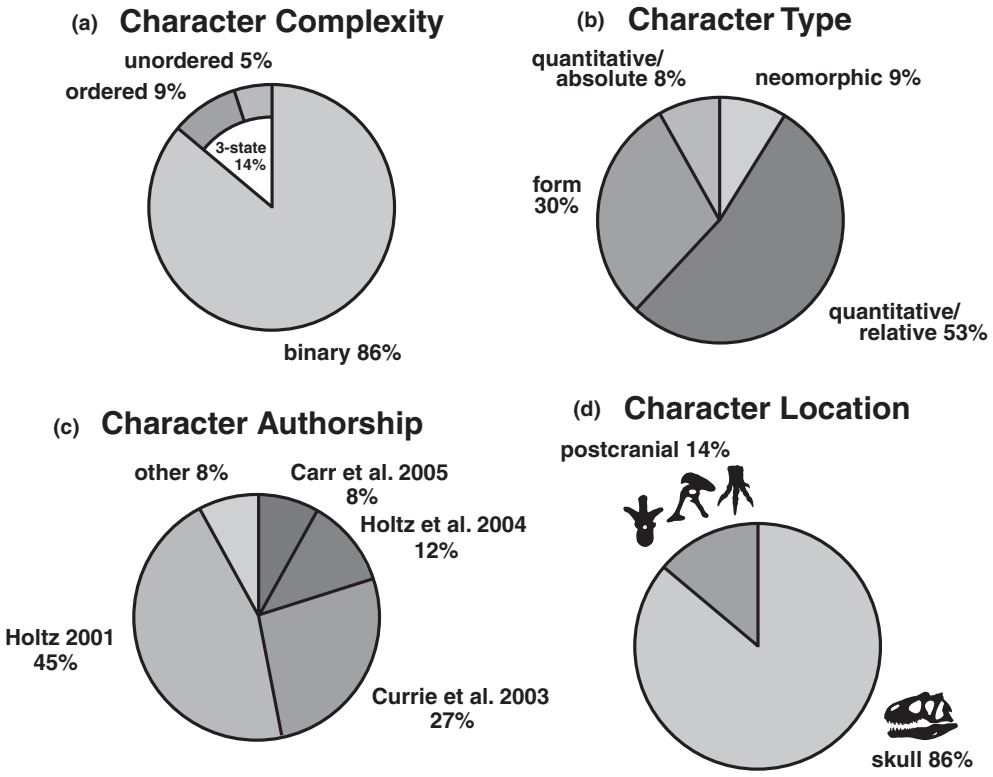

Fig. 14. Data characterization using *CharacterSearch* for 85 morphological characters used in analysing relationships among tyrannosaurid dinosaurs (Sereno and Brusatte, 2009). (a) Character complexity; (b) Character type; (c) Character authorship; (d) Character location.

to 48 (Table 8) with a pooled accumulation of 85 unique characters, which were logged and characterized (Fig. 14). Holtz (2001) generated nearly half the character data in the first analysis (Fig. 14c). More than 80% of the relevant characters are binary and located in the cranium (Fig. 14a, d). Relevant characters were compared to identify shared and unshared data partitions between pairs of analyses. Similarity indices for characters and character-state scores were calculated between pairs of analyses (Fig. 16) in order to address level II questions (Table 7).

The remarkable outcome of this data comparison is that only four relevant characters are shared by all four analyses. No analysis has more than about half of the 85 unique characters present across all of the analyses (Sereno and Brusatte, 2009). Less than half the characters are shared by at least three of the four analyses, and many characters are used in only a single analysis. Character selection was measured using the CSI, or the number of shared relevant characters divided by the total number of relevant characters in a comparison between two analyses. Six pairwise comparisons between the four analyses show remarkable disparity in character selection between analyses (Fig. 16, upper right cells). The greatest similarity in character selection involves the two analyses with the same first author (Holtz, 2001; Holtz et al., 2004). Yet even here, the CSI is 0.58, or an overlap of only 58% of the relevant character data. The CSI for other pairwise comparisons is extremely low, ranging from 0.14 to 0.19, or an overlap of less than 20% of the relevant character data. In this light, the overall similarity between these hypotheses is remarkable, given that most share less than 20% of relevant character data. The overwhelming amount of character dissimilarity between analyses was hidden. The hypotheses are based largely on disparate character data.

Evaluating conflicting character-state scores

Morphology-based systematists often have little idea how comparable character data are scored in an opposing analysis, or how any scoring differences affect phylogenetic results. We need to locate and quantify differences in character-state scores for shared character data in outgroups and comparable ingroup terminal taxa. The ASI and CSSI measure character-state variation between competing data sets. Character-state disparity or conflict, in addition, can be analysed by neutralizing or swapping conflicting character states (Table 6). In this way, level III questions regarding character-state scores can be answered (Table 7).

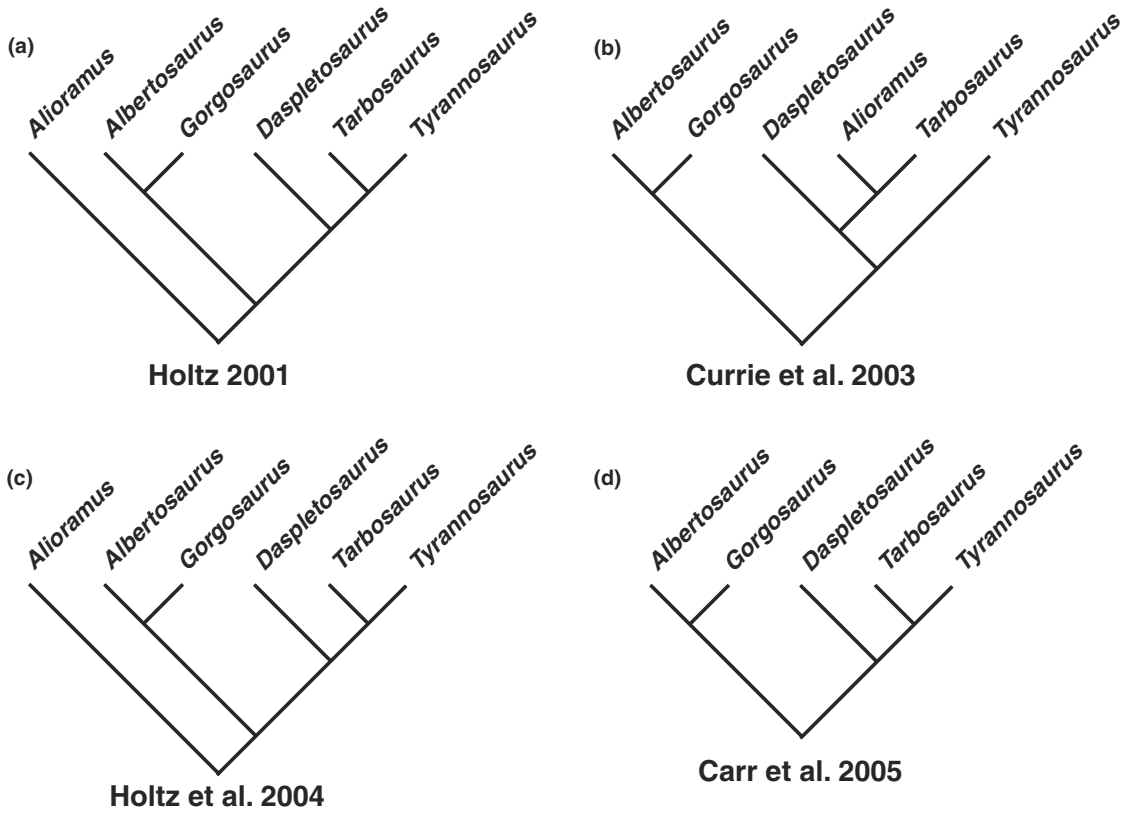

Fig. 15. Cladograms (strict consensus) from analyses of Tyrannosauroidea after reduction in terminal taxa. (a) Holtz (2001) showing genera rather than species; (b) Currie et al. (2003); (c) Holtz et al. (2004) with *Alectrosaurus* removed; (d) Carr et al. (2005) with *Tyrannosaurus bataar* and *Albertosaurus libratus* referred to the genera *Tarbosaurus* and *Gorgosaurus*, respectively.

Analyses	Holtz, 2001	Currie et al., 2003	Holtz et al., 2004	Carr et al., 2005
Holtz, 2001		0.15	0.58	0.15
Currie et al., 2003	0.83		0.19	0.10
Holtz et al., 2004	0.92	0.81		0.14
Carr et al., 2005	0.96	0.93	0.95	

Character Similarity Index (CSI)

Character State Similarity Index (CSSI)

Fig. 16. Pairwise comparison of character selection (CSI) and character-state scoring (CSSI) between four analyses. CSI and CSSI measure character selection and scoring differences, respectively, between two analyses (after Sereno and Brusatte, 2009).

Case study. The interrelationships of tyrannosaurids based on four recent analyses (Fig. 15) are compared in order to determine how much character-state disparity or conflict is present in shared character data between hypotheses. A surprising number of discrepancies surface when a detailed comparison of character-state scoring is undertaken. Thirty character statements show a total of 44 conflicting character-state scores from one

Table 8
Four phylogenetic analyses considered tyrannosaurid relationships but differ considerably in taxonomic scope and included character data. The original number of ingroup taxa and characters is reduced, sometimes severely, when the problem is limited to shared taxonomic scope focusing on tyrannosaurid relationships

Authors		Ingroups		Characters	
Number	Analysis	Original	Reduced	Original	Reduced
1	Holtz, 2001	14	6	111	42
2	Currie et al., 2003	7	6	77	34
3	Holtz et al., 2004	75	6	638	48
4	Carr et al., 2005	7	5	31	19

analysis to another. Most of these discrepancies (84%) involve character-state mismatch (e.g. 1 versus 0) rather than character-state disparity (16%, e.g. 1 versus ?), as tabulated elsewhere (Sereno and Brusatte, 2009).

The CSSI ranges from 0.81 to 0.99 for the six possible pairwise comparisons between the four analyses (Fig. 16, bottom left cells). Analyses that share the greatest number of character states (and thus are compared more effectively) include 198 character states for Holtz (2001) versus Holtz et al., (2004) and 78 character states for Currie et al. (2003) versus Holtz (2004). In the case of the former two analyses with the same lead author, nearly one in 10 character states (CSSI = 0.92) for the same characters in the same taxa are scored with a different state. This discrepancy in character-state scoring, which is hidden in respective data matrices, is very significant and may well give rise to results that differ by as little as a couple of steps. In the latter comparison involving Currie et al. (2003) and Holtz (2004), CSSI is lower (0.81), indicating that nearly 20% of character-state scores for the same characters in the same taxa differ in some significant manner.

The phylogenetic significance of conflictive character-state scores can be explored by neutralizing them with an ambiguous state, swapping them with scores from an opposing hypothesis, or using states based on consensus (level III; Table 7). If the analysis of Currie et al. (2003) is reanalysed with character-state scores from Holtz et al., (2004) where they conflict, the closer relationship of *Tarbosaurus* to *Daspletosaurus* rather than *Tyrannosaurus* is reversed, matching that in the other hypotheses (Fig. 15). When character swapping is applied in reverse to the analysis of Holtz (2004), the relationships remain the same. Although many additional comparisons and reanalyses are possible (Sereno and Brusatte, 2009), one of the principal aims of isolating conflictive character-state scores is to resolve them by re-examination of original materials.

Tangible indices and software

Consistency and retention indices are simple measures of the self-consistency of character data, and were widely adopted by morphology-based cladists once personal computers and appropriate software became available. Character selection and variation in character-state scores, likewise, have long been acknowledged as major factors underlying different phylogenetic results in morphology-based analyses. Although the similarity indices proposed above (ASI, CSI, aCSI, CSSI) measure these factors tangibly, their broader use will surely depend on the availability of facilitating software.

Software for data characterization similar to that shown here is easily achieved (Figs 5–7 and 14). Data comparison between two or more hypotheses is more involved (Fig. 17). Terminal taxa, characters, and matrices must be loaded and compared with those in another analysis. Adjusting for shared taxonomic scope involves linkage of shared or comparable terminal taxa between analyses and the collapse, renaming and/or pruning of others. The resultant modified data sets must be reanalysed. Some characters may require alternative character-state assignments to overlap properly. Opposing data sets and cladograms must be available simultaneously for comparison with highlighted disparate or conflicting character-state scores. To explore underlying differences between hypotheses, an application should facilitate running particular data partitions (e.g. shared characters) or assigning alternative character states to conflictive scores (e.g. neutralizing).

Historical sketch and future promise

This final section attempts to put comparative cladistics, as outlined above, in historical context by considering how morphology-based cladistics originated and evolved in methodological complexity. Two distinctive processes, here termed ‘‘atomization’’ and ‘‘quantification’’, appear to have played critical, synergistic roles in converting nineteenth-century evolutionary narrative into cladistic methodology (Fig. 18). Narrative terms for transformation, such as ‘‘homology’’ and ‘‘ancestral stock’’, were eventually replaced by a more specific, atomized description of transformation (e.g. ‘‘synapomorphy’’, ‘‘clade’’), which ultimately led to its quantification. Characters came to be understood as quantitative variables for simultaneous evaluation.

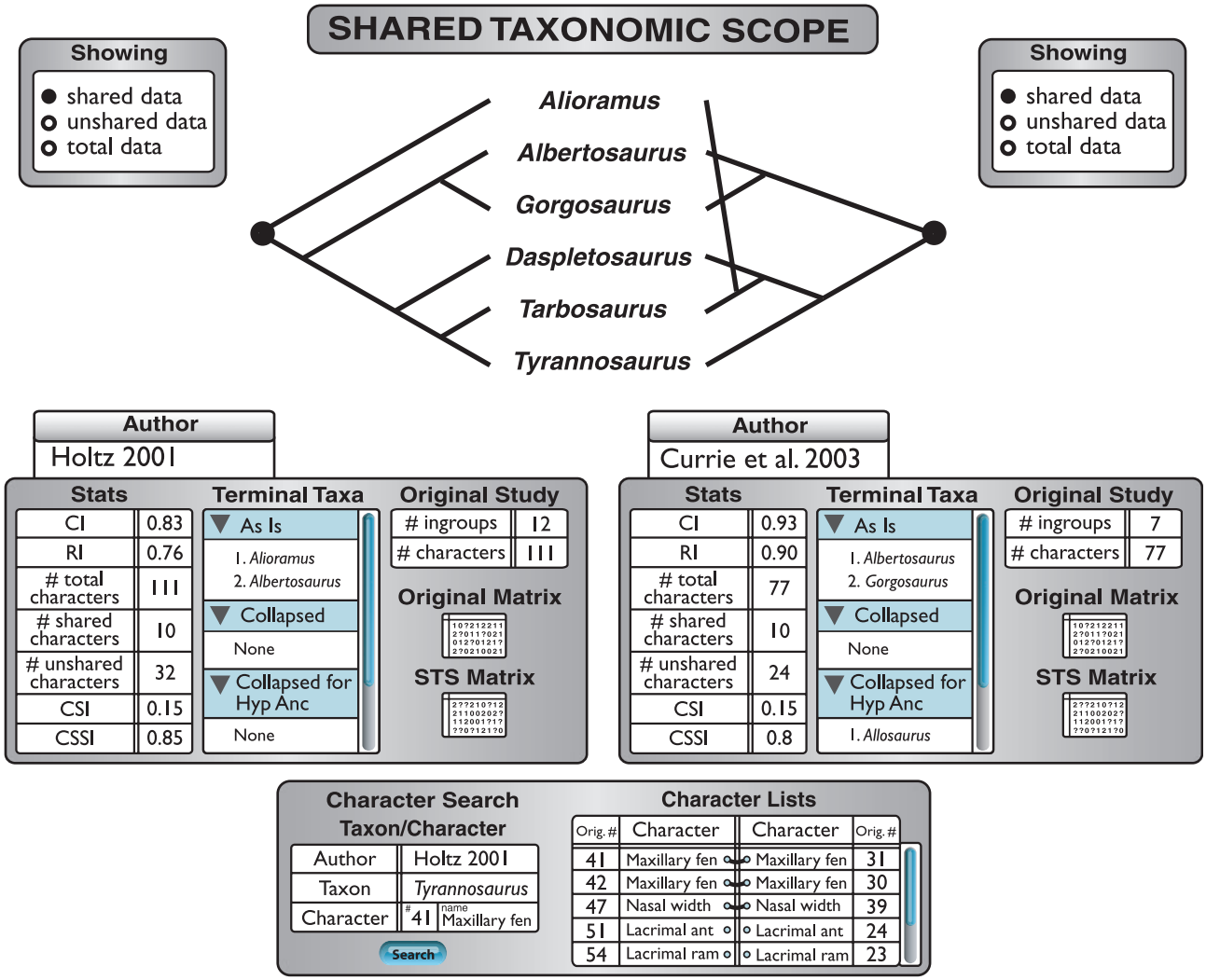

Fig. 17. Draft home screen for software to facilitate data comparison. Data matrices, associated character lists, and cladograms from opposing hypotheses could be opened simultaneously. Taxon collapse and removal would allow the determination of shared taxonomic scope. Drag-and-drop between comparable terminal taxa and character lists would facilitate alignment of cladograms and the determination of shared character data, respectively. Phylogenetic analysis of data partitions would allow the isolation of conflicting phylogenetic signal. Cell-to-cell comparison of shared data between comparable terminal taxa would highlight character-state conflict and disparity, and allow automatic calculation of similarity indices.

Subsequently, evaluation of phylogenetic trees was quantified, resulting in *a posteriori* tree-based measures of length, consistency, robustness, decisiveness, tree space, and others. This paper proposes quantitative evaluation of character data, outlining *a priori* character-based indices to measure similarity in character selection and character-state scoring.

Darwinian phylogenetics. In 1859, Darwin epitomized a more contemporary conceptualization of phylogeny as descent with modification as the best explanation for hierarchical patterns of morphological traits among extant species and a fossil record documenting extinct species that implied transformation over time (stage I; Fig. 18; Table 9). The "great and universal feature in the affinity of all organic beings, namely, their subordination in group under group" is the "hidden bond of connection which naturalists have sought … genealogical in its arrangement, with the grades of difference expressed by the terms genera, families, orders, etc." (Darwin, 1859, p. 333).

Darwinian phylogenetics has been characterized aptly as a form of historical narrative, the phylogenetic tree emerging as its graphical emblem, its most potent narrative device (O'Hara, 1992). Morphological transformation was captured as evolutionary chronicle, adjustable in scale and replete with a "canon of events that are taken to have been key innovations" (O'Hara, 1992, p. 153). Although sometimes demoted by cladists as playing a "superficial" role for its lack of impact on

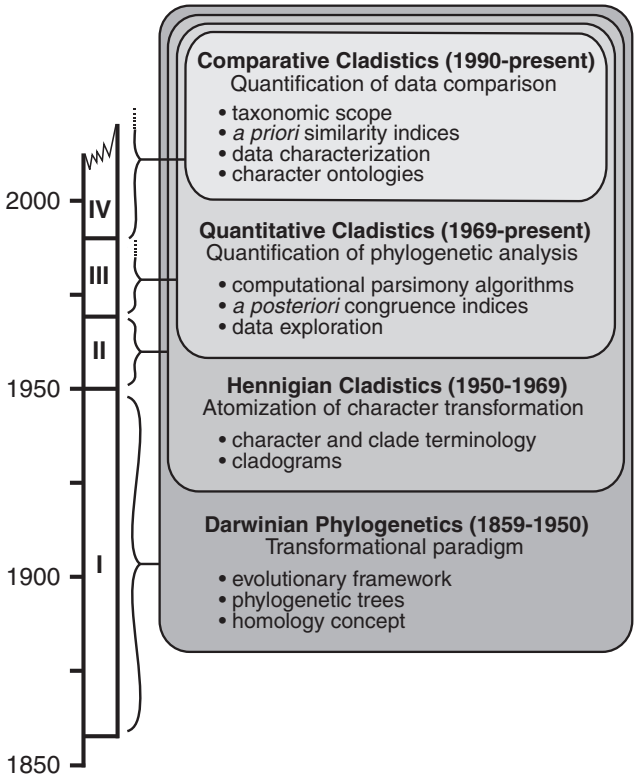

Fig. 18. Major conceptual stages in the practice of morphological phylogenetic analysis from Darwin to present-day cladistics. Darwin established the transformational paradigm we use in phylogenetic analysis. Hennigian cladistics atomized morphological transformation into its component parts. Farris and Kluge initiated an era of quantitative cladistics using parsimony and *a posteriori* congruence indices. Comparative cladistics focuses on *a priori* character issues and quantifies data comparison with *a priori* similarity indices.

taxonomy (Mishler, 2000, p. 661), Darwin's evolutionary worldview had much greater impact on the practice of systematics than classificatory convention (Griffiths, 1974). Nonetheless, the narrative lexicon of Darwinian phylogenetics (character, homology, analogy, convergence, primitive, advanced, etc.) did not distil morphological transformation into component parts in a way that would allow quantification. For much of the early twentieth century, as a result, morphological transformation was viewed as little more than untestable phylogenetic narrative.

Hennigian cladistics. Nearly a century after publication of Darwin's *Origin of Species*, Hennig (1950, 1952, 1953, 1965, 1966, 1969) initiated the atomization of morphological transformation (stage II; Fig. 18; Table 9), with a few notable precursors (e.g. Donoghue and Kadereit, 1992). Hennig introduced specific terms for the kinds of group that can be identified on a phylogenetic tree and the components of character change (Dupuis, 1984; Richter and Meier, 1994). For groups, he underscored the importance of complete (or monophyletic) groups over incomplete (or paraphyletic) groups, and for characters he outlined their subdivision into plesiomorphic (primitive) and apomorphic (derived) states along with some methods to differentiate these conditions. ''Synapomorphy'', or shared apomorphy, he eventually regarded as the basis for recognizing monophyletic groups in contrast to groups based on overall similarity. Over a period of 20 years, a distinctive ''Hennigian cladistics'' was born (Dupuis, 1984; Richter and Meier, 1994).

Hennig's methodology, however, did not bring a quantitative method or criterion, such as parsimony, to the problem of character conflict. Hennig (1965, 1966), furthermore, mistakenly equated characters and character states, two distinctive functional components in modern morphology-based character statements (Sereno, 2007).

Quantitative cladistics. Shortly after 1966, when Hennigian cladistics gained a wider audience, a quantitative solution to the problem of character conflict was proposed (Kluge and Farris, 1969; Farris, 1970). Under a criterion of maximum parsimony, with the twin assumptions of character independence and mutual exclusivity of character states, Kluge and Farris (1969) promoted the search for minimum-length trees (or cladograms), in the process creating the first index to

Table 9
Terms and definitions for important stages in the historical development of morphology-based phylogenetic analysis

Term	Definition
Darwinian phylogenetics	Descriptive transformational analysis of morphology depicted in temporally calibrated (or uncalibrated) trees employing a range of terms such as homology, analogy, parallelism, and convergence
Hennigian cladistics	Descriptive transformational analysis of morphology employing terms that (i) atomize characters and character change (plesiomorphy, apomorphy, synapomorphy, homoplasy), and (ii) more precisely specify/identify groups (monophyly, paraphyly, polyphyly) and branching patterns (cladogram)
Quantitative cladistics	Quantitative cladistic analysis of morphological and/or molecular data (i) coded as variables (characters, character states) for computer-assisted computation, and (ii) analysed and compared with previous results by a range of *a posteriori* indices and measures (consistency, retention, homoplasy and decay indices, bootstrap, jackknife, consensus techniques, etc.)
Comparative cladistics	*A priori* analysis of character data that involve (i) data compilation, (ii) data characterization, or (iii) data comparison (ASI, CSI, aCSI, CSSI) to refine and document character data and determine the root causes underlying differing phylogenetic results

measure character consistency (stage III; Fig. 18; Table 9). During the ensuing 20 years, however, the analysis of morphological data remained a hand operation, with the shortest tree and its "robustness" largely the result of qualitative assessment. "Quantitative cladistics" was rapidly adopted in the late 1980s, once software for personal computers became available that facilitated data entry, parsimony analysis, and calculation of congruence indices and consensus trees. Standards and expectations were raised for *a posteriori* analysis and assessment of results. Various measures and methodologies have emerged since that time to explore and compare *a posteriori* results aimed at evaluating branch support, tree space, sensitivity, saturation, decisiveness, and others (Grant and Kluge, 2003; Hillis et al., 2005).

Comparative cladistics. Rigorous assessment of character data, in contrast, is some 30 years behind. A new awareness was raised in 1990 when *a priori* operations that create character data were labelled the *bête noire* of cladistic analysis (Pogue and Mickevich, 1990), culminating in recent critiques and calls for greater "explicitness" (stage IV; Fig. 18; Table 9). Character documentation often has not gone beyond statements such as "character data are derived in part from studies X, Y, and Z." Character critiques, furthermore, are usually buried in text and lack quantitative comparison with character data in pre-existing analyses.

Most systematists would agree that the innumerable assumptions and decisions regarding character selection, delineation, coding, and scoring can, and often do, overshadow subtle differences and options regarding particular assumptions during phylogenetic computation. The delineation and documentation of character data (Mabee et al., 2007; Ramírez et al., 2007; Sereno, 2007) and the quantification of data characterization and comparison outlined in this paper, are efforts to focus attention on *a priori* operations in morphology-based cladistics.

Understanding how differences in character selection and scoring affect cladistic results is central to comparative cladistics (stage IV; Fig. 18; Tables 7 and 9). At present, data comparison is laborious, especially with large matrices or comparisons involving multiple analyses. Software-facilitated data comparison, on the other hand, may promise a future when the presentation of novel phylogenetic results is accompanied by an exploration of the underlying causes.

Conclusions

To maintain relevancy and rigour in systematics, morphology-based cladistic analysis must enter a comparative era that overcomes long-recognized limitations in data compilation, comparison and synthesis. Those limitations involve character delineation, selection, coding, and scoring of character data—*a priori* operations divided here into data compilation, data characterization, and data comparison.

For data compilation:

1. Complete character statements are divisible into character and statement, which are composed collectively of four components (locator, variable, variable qualifier, character state) arranged in two fundamental patterns (neomorphic and transformational). A minimal set of presentation standards and linkage to character ontologies will help to minimize unnecessary variations in character data that hinder data comparison.

2. Besides image capture and recall, online databases can facilitate a more systematic search for character data relevant to a particular phylogenetic problem (including rejected data) that would also capture auxiliary information useful to data characterization.

For data characterization:

3. Auxiliary information about morphological character statements includes valuable comparative information about character authorship, status, structure, type, anatomical location, and temporal accumulation.

4. Missing information within the data matrix can be summarized effectively in graphical form as a function of terminal taxa as well as character location.

For data comparison:

5. Data comparison is a three-step procedure that establishes shared taxonomic scope (step 1), isolates and then partitions relevant character data (step 2), and measures data similarity and explores the differences between data sets (step 3).

6. Shared taxonomic scope, a fundamental concept for data comparison between two or more phylogenetic analyses, can be specified by location of the most proximate shared (or comparable) outgroup and by adjusting the inclusiveness of ingroup terminal taxa.

7. Character-state scores differ in two ways, here termed character-state disparity and character-state conflict.

8. Similarity indices quantify key data comparisons on a scale from 0.00 (no similarity) to 1.00 (identity). The ASI measures the similarity of the ancestral condition at the shared outgroup node and thus assesses similarity in outgroup assumptions; character similarity indices (CSI, aCSI) measure the percentage of relevant character data that is shared between hypotheses and thus assesses character selection; the CSSI measures the degree of similarity in the character states of shared character data and thus assesses character scoring.

9. Data comparison explores similarity at three levels. Taxon-level comparison clarifies the degree to which opposing hypotheses overlap; character-level

comparison considers all aspects of characters and the phylogenetic signal of shared and unshared data partitions; character state-level comparison highlights scoring differences and how these affect phylogenetic signal.

10. Atomization and quantification are two fundamental historical trends that can be traced from the historical narrative of morphological transformation in Darwinian phylogenetics, through the atomization of character change in Hennigian cladistics, to the treatment of characters as independent variables in quantitative cladistics. Quantification of data characterization and comparison is the frontier for a comparative cladistics that seeks to understand better the reasons underlying differing phylogenetic results.

Acknowledgements

I thank Carol Abraczinskas for execution of all of the figures and D. Blackburn, S. Brusatte, J. Hopson, P. Mabee, D. Pol, J. Wilson and an anonymous reviewer for their helpful critique of the content of this paper.

References

Blanco, W., Gaitros, C., Gaitros, D., Jammigumpula, N., Maneva-Jakimoska, K., Paul, D., Ronquist, F., Seltmann, K., Winner, S., 2006. MorphBank, version 2.2. http://www.morphbank.net

Brower, A.V.Z., 2000. Homology and the inference of systematic relationships: some historical and philosophical perspectives. In: Scotland, R., Pennington, R.T. (Eds.), Homology and Systematics. Systematics Association, Special Volume Series 58, London, pp. 10–21.

Brower, A.V.Z., Schawaroch, V., 1996. Three steps of homology assessment. Cladistics 12, 265–272.

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., 2003. LAGAN and multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 13, 721–731.

Bryant, H.N., 1989. An evaluation of cladistic and character analyses as hypothetico-deductive procedures, and the consequences for character weighting. Syst. Zool. 38, 214–227.

Carr, T.D., Williamson, T.E., Schwimmer, D.R., 2005. A new genus and species of tyrannosauroid from the Late Cretaceous (Middle Campanian) Demopolis Formation of Alabama. J. Vert. Paleontol. 25, 119–143.

Chomsky, N., 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, MA.

Clark, A.G., Whittam, T.S., 1992. Sequencing errors and molecular evolutionary analysis. Mol. Biol. Evol. 9, 744–752.

Coates, M., Ruta, M., 2000. Nice snake, shame about the legs. Trends Ecol. Evol. 15, 503–507.

Currie, P.J., Hurum, J.H., Sabath, K., 2003. Skull structure and evolution in tyrannosaurid dinosaurs. Acta Palaeontol. Pol. 48, 227–234.

Darwin, C., 1859. The Origins of Species by Natural Selection. John Murray, London.

Datta, S., van Engelen, R., Gaitros, D., Jammigumpula, N., 2007. Experiences with tracking the effects of changing requirements on Morphbank: a web-based bioinformatics application. In: John, D., Kerr, S. (Eds.), Proceedings of the 45th Annual Southeast Regional Conference, 23–24 March, Winston-Salem. Association for Computing Machinery, New York, pp. 413–418.

Day, W.H., McMorris, F.R., 1992. Critical comparison of consensus methods for molecular sequences. Nucleic Acids Res. 20, 1093–1099.

De Laet, J., 2005. Parsimony and the problem of inapplicables in sequence data. In: Albert, V.A. (Ed.), Parsimony, Phylogeny, and Genomics. Oxford University Press, Oxford, pp. 81–116.

De Pinna, M.C.C., 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7, 367–394.

Dettai, A., Bailly, N., Vignes-Lebbe, R., Lecointre, G., 2004. Metacanthomorpha: essay on a phylogeny-oriented database for morphology—the acanthomorph (Teleostei) example. Syst. Biol. 53, 822–834.

Donoghue, M.J., Kadereit, J.W., 1992. Walter Zimmermann and the growth of phylogenetic theory. Syst. Biol. 41, 74–85.

Dupuis, C., 1984. Willi Hennig's impact on taxonomic thought. Annu. Rev. Ecol. Syst. 15, 1–25.

Farris, J.S., 1969. A successive approximations approach to character weighting. Syst. Zool. 18, 374–385.

Farris, J.S., 1970. Methods for computing Wagner trees. Syst. Zool. 19, 83–92.

Forey, P.L., Kitching, I.J., 2000. Experiments in coding multistate characters. In: Scotland, R., Pennington, R.T. (Eds.), Homology and Systematics. Systematics Association, Special Volume Series 58, London, pp. 54–80.

Grant, T., Kluge, A.G., 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. Cladistics 19, 379–418.

Griffiths, G.C.D., 1974. On the foundations of biological systematics. Acta Biotheor. 23, 85–131.

Harris, S.R., Pisani, D., Gower, D.J., Wilkinson, M., 2007. Investigating stagnation in morphological phylogenetics using consensus data. Syst. Biol. 56, 125–129.

Hawkins, J.A., 2000. A survey of primary homology assessment: different botanists perceive and define characters in different ways. In: Scotland, R., Pennington, R.T. (Eds.), Homology and Systematics. Systematics Association, Special Volume Series 58, London, pp. 22–53.

Hecht, M.K., Edwards, J.L., 1976. The determination of parallel or monophyletic relationships: the proteid salamanders—a test case. Am. Nat. 110, 653–677.

Hedges, S.B., Maxson, L.R., 1996. Molecules and morphology in amniote phylogeny. Mol. Phylogenet. Evol. 6, 312–314.

Hedges, S.B., Sibley, C.G., 1994. Molecules vs. morphology in avian evolution: the case of the "pelecaniform" birds. PNAS 91, 9861–9865.

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

Hennig, W., 1952. Die Larvenformen der Dipteren: Eine Übersicht über die bisher bekannten Jugendstadien der zweiflügeligen Insekten, 3rd edn. Akademie Verlag, Berlin.

Hennig, W., 1953. Kritische Bemerkungen zum phylogenetischen System der Insekten. Beitr. Entomol. Sonderh. 3, 1–85.

Hennig, W., 1965. Phylogenetic systematics. Annu. Rev. Entomol. 10, 97–116.

Hennig, W., 1966. Phylogenetic Systematics. University of Illinois Press, Urbana.

Hennig, W., 1969. Die Stammesgeschichte der Insekten. Kramer, Frankfurt.

Hill, R.V., 2005. Integration of morphological data sets for phylogenetic analysis of Amniota: the importance of integumentary characters and increased taxonomic sampling. Syst. Biol. 54, 530–547.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47, 3–8.

Hillis, D.M., Heath, T.A., John, K.S., 2005. Analysis and visualization of tree space. Syst. Biol. 54, 471–482.
Holtz, T.R. Jr, 2001. The phylogeny and taxonomy of the Tyrannosauridae. In: Carpenter, K., Tanke, D. (Eds.), Mesozoic Vertebrate Life. Indiana University Press, Bloomington, pp. 64–83.
Holtz, T.R. Jr, 2004. Tyrannosauroidea. In: Weishampel, D.B., Dodson, P., Osmólska, H. (Eds.), The Dinosauria. University of California Press, Berkeley, pp. 111–136.
Holtz, T.R. Jr., Molnar, R.E., Currie, P.J., 2004. Basal Tetanurae. In: Weishampel, D.B., Dodson, P., Osmólska, H. (Eds.), The Dinosauria. University of California Press, Berkeley, CA, pp. 71–110.
Jenner, R.A., 2004. The scientific status of metazoan cladistics: why current practice must change. Zool. Scr. 33, 293–310.
Kearney, M., Rieppel, O., 2006. Rejecting the given in systematics. Cladistics 22, 369–377.
Kirsch, J.A.W., 1982. The builder and the bricks: notes toward a philosophy of characters. In: Archer, M. (Ed.), Carnivorous Marsupials. Royal Zoological Society of New South Wales, Sydney, pp. 587–594.
Kitching, I., Williams, D., Forey, P.L., Humphries, C.J., 1998. Cladistics: the theory and practice of parsimony analysis. Oxford University Press, Oxford.
Kjer, K.M., Gillespie, J.J., Ober, K.A., 2007. Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. Syst. Biol. 56, 133–146.
Kluge, A.G., 1991. Boine snake phylogeny and research cycles. Misc. Publ., Mus. Zool., Univ. Mich. 178, 1–58.
Kluge, A.G., 1998. Total evidence or taxonomic congruence: cladistics or consensus classification. Cladistics 14, 151–158.
Kluge, A.G., 2003. The repugnant and the mature in phylogenetic inference: atemporal similarity and historical identity. Cladistics 19, 356–368.
Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of the anurans. Syst. Zool. 18, 1–32.
Knight, A., Mindell, D.P., 1993. Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of Fea's viper. Syst. Biol. 42, 18–31.
Lee, M.S.Y., 2001. Molecules, morphology, and the monophyly of diapsid reptiles. Contrib. Zool. 70, 1–22.
Li, C., Wu, X.C., Rieppel, O., Wang, L.T., Zhao, L.J., 2008. An ancestral turtle from the Late Triassic of southwestern China. Nature 456, 497–501.
Mabee, P.M., Ashburner, M., Cronk, Q., Gkoutos, G.V., Haendel, M., Segerdell, E., Mungall, C., Westerfield, M., 2007. Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol. Evol. 22, 345–350.
Maddison, D.R., Maddison, W.P., 2003. MacClade 4: Analysis of Phylogeny and Character Evolution, ver. 4.06.
Maddison, W.P., Maddison, D.R., 2005. Mesquite: A Modular System for Evolutionary Analysis, ver. 1.06. http://mesquiteproject.org
Maddison, W.P., Donoghue, M.J., Maddison, P.R., 1984. Outgroup analysis and parsimony. Syst. Zool. 33, 83–103.
Maddison, D.R., Swofford, D.L., Maddison, W.P., 1997. NEXUS: an extensible file format for systematic information. Syst. Biol. 46, 590–621.
Marsh, O.C., 1879. Notice of new Jurassic reptiles. Am. J. Sci. (Ser. 3) 18, 501–505.
Mickevich, M.F., 1982. Transformation series analysis. Syst. Zool. 31, 461–478.
Mishler, B.D., 1994. Cladistic analysis of molecular and morphological data. Am. J. Phys. Anthropol. 94, 143–156.
Mishler, B.D., 2000. Deep phylogenetic relationships among "plants" and their implications for classification. Taxon 49, 661–683.
Mishler, B.D., 2005. The logic of the data matrix in phylogenetic analysis. In: Albert, V.A. (Ed.), Parsimony, Phylogeny, and Genomics. Oxford University Press, Oxford, pp. 57–70.
Miyamoto, M.M., Fitch, W.M., 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44, 64–76.
Neff, N.A., 1986. A rational basis for *a priori* character weighting. Syst. Zool. 35, 110–123.
O'Hara, R.J., 1992. Telling the tree: narrative representation and the study of evolutionary history. Biol. Phil. 7, 135–160.
Patterson, C., 1982. Morphological characters and homology. In: Joysey, A., Friday, A.E. (Eds.), Problems of Phylogenetic Reconstruction. Academic Press, London, pp. 21–74.
Patterson, C., Johnson, G.D., 1997. The data, the matrix, and the message: comments on Begle's "Relationships of the osmeroid fishes". Syst. Biol. 46, 358–365.
Pennisi, E., 2003. Modernizing the tree of life. Science 300, 1692–1697.
Poe, S., Wiens, J.J., 2000. Character selection and the methodology of morphological phylogenetics. In: Wiens, J.J. (Ed.), Phylogenetic Analysis of Morphological Data. Smithsonian Institution Press, Washington, DC, pp. 20–36.
Pogue, M.G., Mickevich, M.F., 1990. Character definitions and character state delineation: the bête noir of phylogenetic inference. Cladistics 6, 319–361.
Pol, D., 2004. Phylogenetic relationships of basal Sauropodomorpha. PhD thesis, Department of Geosciences, Columbia University, New York.
Ramírez, M.J., Coddington, J.A., Maddison, W.P., Midford, P.E., Prendini, L., Miller, J., Griswold, C.E., Hormiga, G., Sierwald, P., Scharff, N., Benjamin, S.P., Wheeler, W.C., 2007. Linking of digital images to phylogenetic data matrices using a morphological ontology. Syst. Biol. 56, 283–294.
Richter, S., Meier, R., 1994. The development of phylogenetic concepts in Hennig's early theoretical publications (1947–1966). Syst. Biol. 43, 212–221.
Rieppel, O., 2004. What happens when the language of science threatens to break down in systematics: a Popperian perspective. In: Williams, D.M., Forey, P.L. (Eds.), Milestones in Systematics. CRC Press, Boca Raton, pp. 57–100.
Rieppel, O., 2006. The merits of similarity reconsidered. Syst. Biodiv. 4, 137–147.
Rieppel, O., Kearney, M., 2002. Similarity. Biol. J. Linn. Soc. 75, 59–82.
Rieppel, O., Kearney, M., 2007. The poverty of taxonomic characters. Biol. Phil. 22, 95–113.
Rieppel, O., Reisz, R.R., 1999. The origin and early evolution of turtles. Annu. Rev. Ecol. Syst. 30, 1–22.
Rieppel, O., Zaher, H., 2000. The intramandibular joint in squamates, and the phylogenetic relationships of the fossil snake *Pachyrhachis problematicus* Haas. Fieldiana Geol. 43, 1–69.
Scotland, R., Pennington, R.T. (Eds.), 2000. Homology and Systematics. Systematics Association, Special Volume Series 58, London.
Scotland, R.W., Olmstead, R.G., Bennett, J.R., 2003. Phylogeny reconstruction: the role of morphology. Syst. Biol. 52, 539–548.
Sereno, P.C., 1984. The phylogeny of Ornithischia: a reappraisal. In: Reif, W.E., Westphal, F. (Eds.), Third Symposium on Mesozoic Terrestrial Ecosystems, Short Papers. Attempto Verlag, Tübingen, pp. 219–226.
Sereno, P.C., 2004. *CharacterSearch*: online database for characters. J. Vert. Paleontol. 24, 112A.
Sereno, P.C., 2007. Logical basis for morphological characters in phylogenetics. Cladistics 23, 565–587.
Sereno, P.C., Brusatte, S.L., 2009. Comparative assessment of tyrannosaurid interrelationships. J. Syst. Palaeontol. 7, in press.
Sereno, P.C., Wilson, J.A., Conrad, J.L., 2004. New dinosaurs link southern landmasses in the Mid-Cretaceous. Proc. R. Soc. Lond. B 271, 1325–1330.
Smythe, A.B., Sanderson, M.J., Nadler, S.A., 2006. Nematode small subunit phylogeny correlates with alignment parameters. Syst. Biol. 55, 972–992.
Stevens, P.F., 1991. Character states, morphological variation, and phylogenetic analysis: a review. Syst. Bot. 16, 553–583.

Stevens, P.F., 2000. On characters and character states: do overlapping and non-overlapping variation, morphology and molecules all yield data of the same value? In: Scotland, R., Pennington, R.T. (Eds.), Homology and Systematics. Systematics Association, Special Volume Series 58, London, pp. 81–105.

Thacker, P.D., 2003. Morphology: the shape of things to come. Bioscience 53, 544–549.

Thompson, J.D., 1999. BAliBASE: a benchmark alignment database for the evaluation of multiple alignment programs. Bioinformatics 15, 87–88.

Wägele, J.W., 2004. Hennig's phylogenetic systematics brought up to date. In: Scotland, R., Pennington, R.T. (Eds.), Homology and Systematics. Systematics Association, Special Volume Series 58, London, pp. 101–125.

Wagner, G.P. (Ed.), 2001. The Character Concept in Evolutionary Biology. Academic Press, San Diego.

Watanabe, M., 2002. Describing the "tree of life": attainable goal or stuff of dreams? Bioscience 52, 875–880.

Wheeler, Q.D., 1986. Character weighting and cladistic analysis. Syst. Zool. 35, 102–109.

Wheeler, W.C., 2001. Homology and the optimization of DNA sequence data. Cladistics 17, 3–11.

Wheeler, W.C., 2003. Implied alignment: a synapomorphy-based multiple-sequence alignment method and its use in cladogram search. Cladistics 19, 261–268.

Wiens, J.J., 2001. Character analysis in morphological phylogenetics: problems and solutions. Syst. Biol. 50, 689–699.

Wilkinson, M., 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. Syst. Biol. 44, 501–514.

Woodger, J.H., 1952. Science without properties. Brit. J. Phil. Sci. 2, 193–216.