

## Logical basis for morphological characters in phylogenetics

Paul C. Sereno\*

*Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637, USA*

Accepted 1 March 2007

---

### Abstract

Systematists have questioned the distinction between characters and character states and their alignment with the traditional concept of homology. Previous definitions for character and character state show surprising variation. Here it is concluded that *characters* are simply features expressed as independent variables and *character states* the mutually exclusive conditions of a character. Together, characters and character states compose what are here termed *character statements*. Character statements are composed of only four fundamental functional components here identified as locator, variable, variable qualifier, and character state, and these components exist in only two patterns, neomorphic and transformational. Several controversies in character coding and the use of “absent” as a character state are understood here as a consequence of incomplete character statements and the inappropriate mixing of neomorphic and transformational character statements. Only a few logically complete patterns for morphological character data exist; their adoption promises to greatly reduce current variability in character data between analyses.

© The Willi Hennig Society 2007.

---

“Character” as a cladistic concept was first explored in detail in an influential paper by Patterson (1982) entitled “Morphological characters and homology.” Surprisingly, the term “character” was never defined. Patterson used “character” interchangeably with “homolog”, “anatomical singular”, “feature”, and “characteristic” (Patterson, 1982, pp. 23, 25). He identified “utilitarian” systematists (e.g., Blackwelder, Crowson, Ross) who equated “character” and “homology”, and Patterson likened this to his view that “homologies are the characters of monophyletic, or natural, taxa” (Patterson, 1982, p. 62). For Patterson, characters alone were sufficient to capture morphological transformation; following Hennig (1966, p. 89) and Bock (1974, p. 387), he suggested that “character” and “character state” are operationally one in the same (Patterson, 1982, p. 25). Many authors, perhaps unintentionally, adopt this view when speaking of derived “characters”, rather than derived “character states”—wording that has long been recognized as ambivalent (Michener and Sokal, 1957; Colless, 1985; Rodrigues, 1986).

In contrast to “character”, Patterson (1982) presented several definitive statements about “homology.” He developed an idea first forwarded by other cladists that “homology” and “synapomorphy” are best understood as synonyms (e.g., Wiley, 1975; Bonde, 1977; Cracraft, 1978; Nelson, 1978; Nelson and Platnick, 1981). Patterson defined “homology” variously as “a hypothesis of monophyletic grouping”, “similarity characterizing monophyletic groups”, a “relation characterizing natural groups”, or simply “discovery” (Patterson, 1982, pp. 21, 61, 65), and he differentiated “taxic” from “transformational” homology (de Pinna, 1991; Rieppel, 1994). Others have distinguished “primary homology” (the initial proposition of similarity) from “secondary homology” (shared similarity based on congruence; de Pinna, 1991; Brower and Schawaroch, 1996). “Character” and “primary homology”, according to these authors, are synonyms (Table 1, definitions 7, 10, 12).

### Scope of the present paper

So what constitutes a phylogenetic “character” or, if they are one in the same, a “primary homolog”? Are

---

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

Table 1

Previous definitions given for “character” by a traditional systematist (definition 1), Hennig (definition 2), and quantitative cladists (definitions 3–15) that do not explicitly incorporate character independence or the mutual exclusivity of character states

No.	Definition	Reference
1	“Any attribute of an organism or a group of organisms by which it differs from an organism belonging to a different category or resembles an organism of the same category”	Mayr et al. (1953, p. 315)
2	“those peculiarities that distinguish a semaphoront (or a group of semaphoronts) from other semaphoronts ‘characters’”	Hennig (1966, p. 7)
3	“a theory that two attributes which appear different in some way are nevertheless the same (or homologous)”	Platnick (1979, p. 542)
4	“a part or attribute of an organism that may be described, figured, measured, weighed, counted, scored or otherwise communicated by one biologist to other biologists”	Wiley (1981, p. 8)
5	“we consider a multistate character to be any set of more than two organic or inorganic states that have, through some process, transformed from one into the other”	O’Grady and Deets (1987, p. 268)
6	“attributes of organisms that have undergone evolutionary change...a gene, a morphological trait, an ontogenetic sequence, a behavioral attribute, or any other heritable feature”	Mabee (1989, p. 151)
7	“A primary homology statement is conjectural, based on similarity, and reflects the expectation that there is a correspondence of parts [of organisms] that can be detected by an observed match of similarities”	de Pinna (1991, p. 373)
8	“Any feature that is an observable part of an organism”	Grande and Rieppel (1994, p. 261); Liem et al. (2001, G-6)
9	“a particular feature interpreted as transformationally homologous to another feature”	Zelditch et al. (1995, p. 180)
10	“(1)...an [sic] hypothesis of <i>primary homology</i> in two or more <i>terminal taxa</i> based on original observations of organisms. (2) A [sic] observable feature of an organism used to distinguish it from another”	Kitching et al. (1998, p. 201)
11	“The terms ‘character’ and ‘primary homology statement’ become one in the same”	Williams and Seibert (2000, p. 185)
12	“any feature shared among organisms that we think will provide information to use in phylogenetic analysis...the sum of features showing particular similarities...topographical homologies...topographical identities...or relationships of primary homology...with each other that we observe in different organisms”	Stevens (2000, p. 82)
13	“hypotheses...subject to the cladistic test of congruence in a parsimony analysis”	Forey and Kitching (2000, p. 55)
14	“an observation that captures distinguishing peculiarities among organisms...a logical relation established between intrinsic attributes of two or more organisms that is rooted in observation and that, if corroborated by congruence, is hypothetically explained as an historical relation”	Rieppel and Kearney (2002, p. 61)
15	“a series of singular historical events”	Grant and Kluge (2004, p. 24)

“characters” and “character states” logically distinct? And why is there such latitude in the format, composition and selection of characters employed by morphologists (Poe and Wiens, 2000; Wiens, 2000)? These are epistemological questions that focus on how these terms have been used and ought to be defined, and do not reflect on character discovery, testing, ordering, coding, weighting, partitioning, polymorphism or variation (Miyamoto and Cracraft, 1991; Stevens, 1991; Mabee and Humphries, 1993; Wheeler, 1995; Mabee, 2000; Wagner, 2001a; Rieppel and Kearney, 2002).

First, I discuss the relevance of the term “homology” and its derivatives to a discussion about characters and character states; I will argue that they are unnecessary. Second, I compile previous definitions for “character”, including Hennig’s, to elucidate how “character” and related terms have been used and variously interpreted by systematists; I conclude that key concepts, such as character independence, are often overlooked and that little terminological consensus exists to describe the basic components of character data. Third, I outline the

logical basis underlying character and character states; I suggest that there exist only two fundamental patterns, neomorphic and transformational, and that these are composed of four functional components. Fourth, I review common controversies in morphology-based character data, such as use of the character state “absent”; I conclude that these are largely the result of incomplete character statements and unjustified mixing of neomorphic and transformational patterns.

## Historical background

### *Characters versus homologs*

Patterson (1982, p. 22) observed cynically that “it is possible to write a good book on phylogenetics without mentioning homology”, citing well-known work. This may be a blessing in disguise, as the concept of homology has gained no singular meaning from its inception over a century ago (de Beer, 1971; Sattler,

1984; Wagner, 1989; de Pinna, 1991; Hall, 1994a,b; Lauder, 1994; Bolker and Raff, 1996). The multifarious definitions of “homology”, “homolog(ue)” and “homologizing” now often function as impediments to clear thinking in comparative biology. According to Stevens (2000, p. 83), the concept of “homology” is “so ambiguous that it is best replaced by the synonyms of its particular usages.” It is telling, for example, that the term “homology” and its etymological derivatives were never once cited in recent reviews of character coding (Swofford and Maddison, 1992) and exploratory methods in phylogenetic analysis (Grant and Kluge, 2003). “Homology” and its derivatives which do not appear to have added clarity to phylogenetic character concepts, are not used in the remainder of this paper.

“Character”, on the other hand, is a term that is commonly encountered in both theoretical and descriptive phylogenetics. Character description and analysis and exploration of the structure of character data are central activities. The precise meaning of “character”, nevertheless, is often neglected. The omission of a clear definition of “character”, then, is what ought to warrant concern in book-length treatments of “comparative method” in evolutionary biology (Harvey and Pagel, 1991), the nature of morphological data (Wiens, 2000), or the relations between homology and systematics (Scotland and Pennington, 2000).

#### *“Character concept” in phylogenetics*

As with homology, a diversity of “character concepts” have emerged in comparative biology (Fistrup, 2001; Wagner, 2001a), some with little applicability to phylogenetics (e.g., Wagner, 2001b, p. 3): “A biological character can be thought of as part of an organism that exhibits causal coherence to have a well defined identity and that plays a (causal) role in some biological processes.”

Traditional taxonomists often cite “diagnostic characters” as taxonomic differentia (Richards, 2003) irrespective of their phylogenetic status (Mayr et al., 1953, p. 315): “Any attribute of an organism or a group of organisms by which it differs from an organism belonging to a different category or resembles an organism of the same category.”

Actual usage of “character” by systematists is more varied, as Colless (1985, p. 230) attempted to canvass: “an attribute, a set of attributes, a feature, a characteristic, a property, a part, a morphocline, a differentia, an homology, a truth, a theory, an aspect (of an organism), a basis for comparison, a similarity, and (mathematically) a variable, a function, a mapping, and equivalence relations, and a set of probability distributions.”

Fistrup (1992, pp. 45–46) observed that “character” could denote “a prescription for observation, an inde-

pendent unit of information...or the hypothetical units/events that play a causal role in the process being studied.” Even more recently, Rieppel and Kearney (2002, p. 59) admitted “elusiveness of the concept of *character* in morphological studies may also lie at the root of criticisms of morphology-based phylogenetic analyses.”

Although Thiele (1993) and Richards (2003) devoted entire papers to evaluating the “perfect cladistic character” and “character individuation”, respectively, in phylogenetics, neither provided a definition for “character.” Others have offered definitions that bear a striking resemblance to Owen’s (1843, p. 379) original definition of “homolog”: “The same organ in different animals under every variety of form and function.”

The seeming nonsensical statement that characters are composed of entities that are “different yet the same” comprises the heart of a large number of published definitions for “character”, including that proposed by Hennig (Table 1). Definitions that state that characters are differentia alone imply such a comparison with comparable, but different, conditions in other organisms. The focus is on characters as differentia—differences in the conditional state of one organism as compared with, or in transformation to/from, another. Authors that proffer this kind of definition for character not surprisingly suggest that there is no difference between “character” and “character state”, both functioning as differentia and thus interchangeable for all intents and purposes (Hennig, 1966; Platnick, 1979; Patterson, 1982).

Grant and Kluge (2004, p. 24) take Richards (2003) and others to task for misinterpreting Hennig’s “transformation series” concept of “character” as “property-based.” As Grant and Kluge and others have long noted (Farris et al., 1970), modern use of “character” in cladistics actually corresponds with Hennig’s “transformation series”, and his use of “character”, “special character” and “character condition” corresponds with modern notions of “character state.” Be that as it may, Hennig (1966, pp. 89–90) was very clear in his equation of what we now call “character” and “character state”: “we must always be aware of the fact that ‘characters’ that can be compared are basically only character conditions...a, a”, a’, etc. will sometimes denote different characters, and sometimes different conditions of one and the same character...”

Hennig (1966) circumscribed a “transformation series” (i.e., a character) on the basis of homology but offered little in the way of an operative definition for such a series. How is such a series defined or delimited? Grant and Kluge (2004), likewise, offered little themselves in the way of a definition for “character” beyond “a series of singular historical events”, although Kluge (2003) clearly viewed characters and character states as logically distinct.

### Characters versus character states

The distinction between character and character state, as will be seen below, is critical and was not generally appreciated until the middle of the twentieth century with the rise of numerical approaches to taxonomy (Freudenstein, 2005). Michener and Sokal (1957) provided one of the earliest discussions that specifically identified the ambiguous use of “character” for both “character” and “character state” (which they termed “state code”). They understood characters in phylogenetic analysis as variables that must be *independent* of one another. Mathematical variables, of course, may or may not be independent. Character independence was not initially explicitly cited by Farris et al. (1970) nor in closely related papers (Kluge and Farris, 1969; Farris, 1970) that initiated quantitative cladistics. Character independence, nonetheless, is requisite for characters as variables under parsimony, even if that status is largely assumed (Felsenstein, 1983; Farris, 1983; Wilkinson, 1995; Farris et al., 1995; Rieppel and Kearney, 2002). Such independence has sometimes been incorporated into the definition of “character” (Table 2, definitions 4, 6).

Before 1970, the concept of *mutually exclusive* conditions was limited largely to numerical taxonomists with more ancient roots in logical philosophy (Michener and Sokal, 1957; Cain and Harrison, 1958; Colless, 1967, 1985; Crovello, 1970; Fistrup, 1992). When Farris et al. (1970) ushered in the era of quantitative cladistics, they argued that a character is composed of *mutually exclusive* conditions, or character states. Characters thus were coded *variables*, which not surprisingly, share much in common with a typical mathematical understanding of that term—“A quantity or force which, throughout a mathematical calculation or investigation, is assumed to vary or be capable of varying in value” (Little et al., 1955, p. 2335). Mutual exclusivity has

sometimes characterized alternative conditions in definitions of character (Table 2, definitions 1, 5).

### Characters as statements or propositions

Following Woodger (1952), Kluge (2003, p. 358) defined a character as “an observed property of an organism... instantiated by an organism.” This character definition, Kluge stated, “is an ordinary categorical proposition” that “concerns only the nature of the subject and the predicate, which in the case of a phylogenetic character refer to the organism and the particular property.” In a typical cladistic character, we encounter what some might call the “character”, followed by a colon and two or more character states. How does this very typical phylogenetic construct correspond to propositional “subject” and “predicate” or to the “part” and “property” referred to above by Kluge?

As an example, Kluge offered “this snake has a caudal vertebra with a hemapophysis”, which he symbolized as “organism A has part X with property 1.” He described the character as “part X” and its states as the “property.” So is part X (subject) the “caudal vertebra” and the property (predicate) the “hemapophysis”? Kluge later suggests that size, form or other features of the hemapophysis, such as length constitute possible properties, so one can infer that “caudal vertebra with a hemapophysis” must be part X (subject). In phylogenetic data, however, properties such as “length” are also commonly listed before the colon, so is “length” part of the subject and only its instantiations (character states) the predicate?

To explore this further, Kluge’s character is reconstructed as it might appear in a phylogenetic analysis:

Caudal vertebrae, hemapophyseal length: shorter (0), or longer (1), than the neural spine.

Table 2

Previous definitions of “character” that explicitly incorporate either character independence or the mutual exclusivity of character states

No.	Definition	Reference
1	“a collection of mutually exclusive states...derived directly from just one other state”	Farris et al. (1970, p. 172)
2	“two or more states, which within the study at hand cannot be subdivided logically, except for subdivision brought about by the method of coding”	Sneath and Sokal (1973, p. 74)
3	“a feature of organisms that can be evaluated as a variable...with two or more states. This stipulation, of course, excludes ‘absent, present’ evaluations of defining characters...absence can be included in a cladistic variable, but not as the plesiomorphic state”	Pimentel and Riggins (1987, pp. 201, 206)
4	“characters must be independent and have a transformation series”	Pogue and Mickevich (1990, p. 619)
5	“a set of alternative conditions, called character states, that that are considered able to evolve one to another”	Maddison and Maddison (1992, p. 45)
6	“an exhaustive partition of $m$ taxa (including all outgroup or hypothetical ancestral taxa) into $n$ non-overlapping subsets corresponding to the character states, together with a set of ‘rules’ that specify the number of units (‘steps’) in the transformation between each ordered pair of character states”	Slowinski (1993, p. 157)
7	“Characters used in phylogenetic analysis are generally assumed to be independent variables”	Pleijel (1995, p. 310)

What constitutes the character proposition, the character, the subject or predicate? Many systematists would use “character” to describe the entire line or, alternatively, to refer only the portion preceding the colon. This anterior portion, however, also clearly includes a property (length). Possibly the “subject”, then, would be limited to the variable “part”, or the caudal hemapophysis. The “predicate” ought to include the property and its conditional states, although some might restrict it to include only the states that follow the colon. The same character may be expressed equivalently as follows:

Caudal vertebrae, hemapophyseal length relative to the neural spine: shorter (0), longer (1).

A portion of what was in the predicate and seemingly linked to conditional states now has shifted forward into the subject in advance of the colon. Some cladists also might use fewer words:

Caudal hemapophysis: short (0), long (1).

Although proponents of propositional logic may fare better in dissecting the above examples, I follow an alternative route below that identifies characters as variables, which is more closely aligned with the principal analytical proposition used to interpret character data (parsimony).

### Logical structure

The two concepts discussed previously are here referred to as *precept I* (characters as independent variables), and *precept II* (character states as mutually exclusive conditions of a character), because they are logical assumptions underlying maximum parsimony as well as many other analytic approaches to character data. I argue below that independence and mutual exclusivity are the only necessary and sufficient definitional properties of characters and character states, respectively. Although this paper focuses on morphological characters, these points apply equally to molecular characters.

#### *Characters as independent variables*

Characters function as independent variables in parsimony analysis (Sokal and Sneath, 1963; Farris, 1983; Felsenstein, 1983). Character independence is a fundamental property (Pogue and Mickevich, 1990; de Pinna, 1991; Farris et al., 1995; Pleijel, 1995; Wilkinson, 1995; Hawkins et al., 1997; Brower, 2000; Hawkins, 2000). Characters “must be independent”, which is “instantly recognized when differing hypothetical characters delimit differing subsets of taxa” (Pogue and Mickevich, 1990, p. 319). As Rieppel and Kearney (2002, p. 61) noted, “cladistic analysis based on parsimony assumes the independence of characters that are subjected to the test of congruence.

The primary role of independence was also cited by Freudenstein (2005) in his consideration of molecular characters and character states. Instead of referring directly to character independence, however, he defined a character as follows (Freudenstein, 2005, p. 968): “individualized assemblages of features (states) among taxa that are the result of duplications, fusions, or foreign acquisitions (‘novelties’) and whose elements exhibit paralogous or equivalent non-orthologous relationships to other assemblages.”

Although he never explicitly defined paralogy, he described paralogous proteins as those “found in a single individual as a result of gene duplication” as well as duplicate genes present in separate species (Freudenstein, 2005, p. 966). Clearly the operative principle here is independence, as realized by gene duplication. Elsewhere he stated this explicitly (p. 967): “If a particular DNA segment is known not to be independent of another (such as member of a tandem repeat array that undergo concerted evolution), it should not be called a distinct character, because independence is a basic requirement for systematic characters.”

There seems little justification to swap ill-defined processes that generate independence for independence itself in a definition of a phylogenetic character.

Character independence is considered in more detail here using tail color as an example (modified from Maddison, 1993):

1. Tail, color: red (0); blue (1).

Another phylogeneticist might prefer to divide the tail into dorsal and ventral margins and a central area, coding each for its color. Taken to a further extreme, individual tail scales/feathers could be coded separately for color, especially if some had unusual shape, structure or location. In this way, tail color could eventually dominate results from a taxon-character matrix. Obviously, if subregions of the tail or individual scales/feathers do not show independent color variation, their interpretation as independent characters is unwarranted. Thus, even if the genetic-developmental foundation governing the generation or transformation of a morphological character is partially or completely unknown (as is usually the case), apparent lack of independence is sufficient to eliminate many potential characters from consideration. Character dependence (correlation) is a leading issue of concern for both morphological and molecular characters (as determined prior to and after analysis; Werdelin and Tullberg, 1995), underscoring the primary role played by character independence in the delineation of character statements. Character independence is a critical a priori assumption for parsimony analysis.

Character independence is at the root of another well-known controversy—how to code and score taxa that lack the structure under consideration (Maddison, 1993; Pleijel, 1995; Hawkins et al., 1997; Forey & Kitching,

2000; Hawkins, 2000). Consider the following pair of character statements:

2. Tail, color: red (0); blue (1); absent (2).
3. Tail: present (0); absent (1).

The character state “absent” for character 2 (meaning the tail is “absent”) is not independent of the character states for character 3, the presence/absence of the tail itself, as noted by Maddison (1993), Pimentel and Riggins (1987) and Hawkins et al. (1997). The character state “absent” in character 2, thus, is a redundant observation. That is the primary reason this coding scheme fails; separate characters are not fully independent because of redundant coding. As outlined in more detail below, this redundancy arises because of a logical coding error: “absent” is not a reasonable character state for the variable “color.” For the coding scheme for character 2 to make sense, “absent” must be a color, a mutually exclusive condition comparable with “red” or “blue”. “Absent” is not a color but rather an observation that something does not exist, an observation already covered by character 3.

In summary, in its most basic sense a phylogenetic *character* is here defined as a heritable, organismal feature (i.e., an observable condition) expressed as an independent variable (Table 3). This is the core of a universal definition that may be applied to either morphological or molecular phylogenetics; there is no need to include paralogy or non-orthology (Freudenstein, 2005), terms that owe their relevance to their role in generating character independence.

#### *Character states as mutually exclusive conditions*

Character states are widely appreciated as mutually exclusive conditions (Freudenstein, 2005). As a purely logical construct, something understood as “present” cannot at the same time be “absent.” Likewise, transformation within a character from one state to another is contradicted by the coexistence of alternative states; if character states do represent alternative conditions of the “same thing”, they ought not to coexist (test of conjunction; Patterson, 1982; Kluge, 2003). For the vast

majority of morphological and molecular characters, coexistence of alternative states suggests that such features actually pertain to separate characters and, thus, are independent.

A particular terminal taxon, of course, may not uniformly express a particular character state. Polymorphism is a well-known exception in both its typical manifestations (as coexisting alternative conditions at the specific level) or as uncertainty in the basal (groundplan) condition for supraspecific taxa with ingroup variation (Mabee and Humphries, 1993). Serial structures and/or character duplication present other opportunities for the coexistence of character states that were once interpreted as mutually exclusive. Nonetheless, the presence of such variation or complexity need not complicate the logical primacy of mutual exclusivity in the definition of a character state.

Grant and Kluge (2004, p. 26) defined a character state as “the least inclusive historical individuals that result from heritable transformation events.” Freudenstein (2005, p. 968), in contrast, defined the same as “mutually exclusive features among taxa of a single paralog-equivalent assemblage that exhibit orthologous relationships to each other.” Although orthology was not defined, Freudenstein (2005) reported that orthologous proteins are “variants of a protein in different species.” A *character state* is defined here as simply a mutually exclusive condition of a character (Table 4). There seems to be little justification to insert concepts of orthology and species into the definition and no reason to exclude its subordinate relationship to “character”.

#### *Character statements*

In traditional taxonomic as well as cladistic literature, the term “character” has been variously used as noted above; it can refer to a variable feature, its variable conditions, or both. A “character list”, for example, is usually taken to refer to both characters and character states. If we accept characters as variables and character states as variable conditions, however, what do we call a character *plus* its character states? I outline below the

Table 3  
Other basic terms and definitions

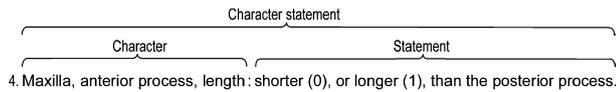
Terms	Definitions
Feature	Observable condition of an organism
Character	Heritable, organismal feature expressed as an independent variable and composed of a locator (or locators) and, optionally, a variable and variable qualifier(s)
Statement	Alternative character states and (optionally) a variable qualifier(s)
Character statement	A character and its associated statement
Primary locator (L <sub>1</sub> )	Single structural feature cited for specific character location
Secondary locator (L <sub>2</sub> )	Structural feature(s) cited for general location of the primary locator
Morphological character	Morphological feature that varies either by its presence or absence or by two or more mutually exclusive transformational states
Qualitative character	Character of form that is difficult to express in terms of absolute or relative magnitude
Quantitative character	Character of magnitude that is either absolute or relative (relational, proportional)

Table 4

Morphological character statements are composed of character and statement, which are here divided into their four logical components (with respective symbolic abbreviation and definition)

Character statement part	Component	Symbol	Definition
Character	Locator	$L_n$	Morphological structure
	Variable	V	Aspect that varies
	Variable qualifier	q	Variable modifier
Statement	Character state	$v_n$	Mutually exclusive condition of a character

terminology employed in this paper using the hypothetical example outlined below in Scheme 1.



The phrase preceding the colon is regarded as the *character*. Two words and their numeric codes—“shorter (0)” and “longer (1)” —are regarded as *character states*. That leaves unnamed the entire line and the entire phrase after the colon. I will call these the *character statement* and *statement*, respectively (Table 3). In sum, then, a character statement is composed of two parts, a character and a statement, the latter of which includes character states. That fundamental parts of a typical line of morphological data are variously identified or remain unnamed is symptomatic of the present lack of definitional rigor in morphology-based phylogenetics.

As mentioned above, some phylogeneticists deny there exists a logical distinction between character and character states (Pleijel, 1995), a position here regarded as untenable. Most authors use a colon to separate these components, as in character statement 4. Some authors blend the two. Thiele (1993, table 1), for example, listed the following:

5. Dorsal fin anterior; posterior.
6. Pollen boat-shaped; globose.

Is “Dorsal fin” the character in example 5? In the logical framework presented here, these character statements are incomplete because they omit important components, such as the particular feature that varies (i.e., the variable). They can be rewritten as complete character statements with appropriate variables as follows:

7. Dorsal fin, location: anterior (0); posterior (1).
8. Pollen, form: boat-shaped (0); globose (1).

In formal logic, a “statement” is generally taken to be a construct with propositional content that proposes that some fact is, or is not, the case or presents a proposition for falsification. Characters statements outlining transformation, as in character statement 7 above, seem to fit this concept; if the dorsal fin is not in either an anterior or posterior position, the character statement is not true or must be modified.

What if the dorsal fin is absent or potentially so modified as to be unrecognizable? It might then be

necessary to consider another character statement involving only the “presence” or “absence” of a dorsal fin. Is such a presence–absence construct also a “statement” that can be falsified? That seems to depend on the interpretation of “absence.” If by “absence” we include the inability to recognize a feature that might actually exist in some altered condition, then it is a component of a statement that can be falsified by recognition of the transformed condition. If the dorsal fin became attached as a small anterior lobe of the caudal fin, for example, its absence is falsified and the character requires reshaping to accommodate the free versus the attached existence of the dorsal fin.

The term “statement” therefore would seem to apply to all character constructs created and used by practicing systematists, the primary audience of the present contribution. The important point here is to underscore the urgent need for explicit terminology that recognizes *characters* as independent variables, *character states* as mutually exclusive conditions, and *character statements* as the construct they compose in combination.

#### *Four components of character statements*

There are four logically distinct base-level components that compose all character statements; thus far only *character states* have been named. Characters are composed of *locators*, *variables* and *variable qualifiers*, as symbolized and defined in Table 4. These four components, some of which are required and others optional, are used in only a few stereotypical patterns. All character statements, be they morphological or molecular, can be atomized into these four components and rendered in symbolic form (Table 5).

A *locator* is a term that points to a particular feature, be it a structure or a location in a molecular sequence. *Primary* ( $L_1$ ) and *secondary* locators ( $L_2$ ) are functionally distinct, but their separation is somewhat arbitrary. The primary locator points to a specific location, i.e., the structure or feature of interest. A primary locator, however, is often a structural feature without a singular name that cannot, by itself, unambiguously identify the feature of interest. One or more secondary locators, thus, are cited to narrow the location/identity of the primary locator. Secondary locators are optional and inclusive; the secondary locator (“maxilla”) includes the

Table 5  
Logical structure for morphologic characters. Two fundamental types are recognized

Character type	Example	Symbolic notation
Neomorphic	Maxilla, anterior process: absent (0); present (1)	$L_2, L_1; v_0; v_1$
Transformational	Maxilla, anterior process, length relative to the posterior process: shorter (0); longer (1)	$L_2, L_1, Vq; v_0; v_1$
	Maxilla, anterior process, length: shorter (0), or longer (1), than the posterior process	$L_2, L_1, V: v_0, v_1, q$

primary locator (“anterior process”). In the example shown below in Scheme 2, there is one primary and one secondary locator.

$L_2 \quad L_1 \quad V \quad v_0 \quad v_1 \quad \overbrace{\quad q \quad}$   
 4. Maxilla, anterior process, length: shorter (0), or longer (1), than the posterior process.

“Anterior process” and “maxilla” are primary and secondary locators, respectively. “Maxilla”, the secondary locator, points to the bone of interest so the particular structure, “anterior process”, may be identified unambiguously. In a vertebrate skeleton, more than a single structure is identified as an “anterior process” (e.g., anterior process of the iliac blade, etc.). The secondary locator “maxilla” thus narrows the focus to a particular bone, on which there is a single “anterior process”, the primary locator or feature of interest. There is only one primary locator, the specific feature of interest.

In the example character, “Maxilla, anterior process”, one could take a more extreme position and argue that “process” is the primary locator, and “anterior” the first, least general, secondary locator that points to the specific process on the maxilla. In this case, I would argue that “anterior process” is the commonly cited lowest-level term, a binomen that would be translated into Latin in formal nomenclature. Rarely would a morphologist, after introducing this structure, refer again to this feature simply as the “process” as opposed to the “anterior process.” I would argue therefore that “anterior process” is best viewed as the primary locator, the feature of interest, although this designation is not without ambiguity.

There is no ambiguity in a character such as “Frontal, shape.” In this case, “frontal” is the unique, one-word primary locator. A secondary locator could be added as a prefix, such as “Skull roof, frontal, shape.” In this case, “skull roof” is clearly secondary and used to point to the general region where the bone of interest is located. Unlike “maxilla”, it cannot be transformed into a single-word adjective (“maxillary”) and more closely affixed to the primary locator, and as a result is more easily understood as a secondary locator.

Secondary locators, thus, narrow the focus to the general location of the primary locator. The exact form or position of a secondary locator is irrelevant; its function remains the same. In the example above, the secondary locator, “Maxilla”, is a noun in lead position in the character. Other functionally equivalent variants include “Maxillary anterior process”, “Anterior process of the

maxilla”, and “Anterior maxillary process”, all of which cite a cranial bone, the “maxilla”, as a secondary locator.

There may be more than a single secondary locator, but there is only one primary locator. “Anterior process” is used as a primary locator in the example above but as a second secondary locator (along with “maxilla”) in the following example:

9. Maxilla, anterior process, fluted articular ridges, number: two (0); three (1).

Here the specific feature of concern, “fluted articular ridges”, is located on the anterior process of the maxilla. This location is specified by two nested secondary locators, “maxilla” and “anterior process”, which provide general and more specific locations, respectively, for the feature of interest. Again, the character may be expressed with primary and secondary locators in other positions, such as “Fluted articular ridges of the maxillary anterior process.”

One could list a more lengthy hierarchy of nested secondary locators that are more general than “maxilla”. They could compose a more elaborate character as follows:

10. Skull, cranium, dermal skull roof, jaws, maxilla, anterior process, fluted articular ridges, number: two (0); three (1).

Most cladists employ only one or two secondary locators, which is sufficient for specialists to properly locate the specific feature of concern. Headers are typically used to subdivide a serial list of character statements. In this way, a general location for character statements is indicated without adding to the complexity of individual characters. “Skull” is used as a header in this manner in the following example:

#### Skull

9. Maxilla, anterior process, fluted articular ridges, number: two (0); three (1).

The primary locator points to the specific feature of interest. This feature may be present, absent or variable in taxa under consideration. In the above examples, “anterior process” and “fluted articular ridges” are the primary locators in character statements 4 and 9, respectively. Typically, there is only a single structure that is present/absent or variable and thus only a single primary locator.

A *variable* is the aspect of the character that varies. In the above examples, “length” and “number” are the variables in character statements 4 and 9, respectively. If no variable is given, then the character itself functions as

the variable with character states “present” and “absent.” These two distinct character types are commonly confused by eliminating the variable and listing only the variable conditions (character statements 5, 6, 11):

4. Maxilla, anterior process, length: shorter (0), or longer (1), than the posterior process.

11. Maxilla, anterior process: shorter (0), or longer (1), than the posterior process.

12. Maxilla, anterior process: present (0); absent (1).

Character statement 4 is complete, whereas character statement 11 omits the variable, “length”, which must be inferred from the character states. The character in character statement 11 (“Maxilla, anterior process”), thus, is identical to that in character statement 12, which, to the contrary, is a presence–absence character statement. Problems stemming from omission of the variable are discussed further below.

A *variable qualifier* is a phrase that modifies the variable. In the following example, the variable qualifier “than the posterior process” modifies the variable “length.” It renders “length” a relative, rather than absolute measure, and identifies the measurement of comparison:

4. Maxilla, anterior process, length: shorter (0), or longer (1), than the posterior process.

In the above example, the variable qualifier is located in the statement after the character states. Variable qualifiers, however, are often listed redundantly within each character state, as if they are a part of the character state:

13. Maxilla, anterior process, length: shorter than the posterior process (0); longer than the posterior process (1).

When placed within a character state (i.e., within the phrase prior to the numeric code), the variable qualifier must be repeated redundantly within each state because, in actuality, they do not belong there and can be removed. To prove this point, a single copy of redundant variable qualifiers located within character states can be relocated adjacent to the variable it qualifies. The variable qualifier, now expressed as “relative to the posterior process”, is best understood as an unusually mobile part of the character:

14. Maxilla, anterior process, length relative to the posterior process: shorter (0); longer (1).

Although the variable qualifier may reside in either the character or statement, there is no logical or practical reason for listing it redundantly within each character state.

Finally, the best known basal component of character statements is a *character state*, a mutually exclusive condition of a variable. As argued below, there are two kinds of character states, those that record the presence or absence of a character and those that partition observed or hypothetical states as transformations, one to another.

In bioinformatics there has recently been an attempt to deconstruct and code phylogenetic character data (Diederich, 1997) in a manner that allows more effective

linkage with controlled anatomical vocabularies (ontologies) and genetic databases (Paterson et al., 2004; Bard, 2005; Li et al., 2005; Mabee et al., in press). One model for descriptive character data is tripartite with “structure-property-state/value” fields (Diederich, 1997). Alternatively “entity-attribute-value” (EAV) fields have been proposed (Li et al., 2005; Mabee et al., in press). Although EAV generally corresponds with locator, variable, and character state as outlined above, real character data is difficult to shoehorn into this format without many of the considerations discussed in this paper. Does a presence-absence character, for example, have a “property” or “attribute” other than its “values” (present, absent)? Diederich et al. (1997) noted, in addition, that “property” (or attribute) is sometimes omitted in descriptive character data, which lists only the “structure” and its associated “states.” Paterson et al. (2004) went further to suggest that “property” might best be eliminated altogether for qualitative morphological characters. The opposite approach is taken here. True transformational character statements, be they quantitative or qualitative (see below), require a variable (=property, attribute), and this important component should be an explicit part of a character statement.

#### *Two kinds of characters: neomorphic and transformational*

There are two fundamental patterns for evolutionary change of an observable condition in an organism: (1) *de novo* appearance or loss without trace, and (2) transformation from one state to another comparable state. These compose the two fundamental character patterns in quantitative cladistics, here termed *neomorphic* and *transformational*, respectively.

Neomorphic characters are composed only of locators, and their character states are limited to “present” and “absent” (Table 5). There is no variable or variable qualifier as these terms are defined above. Because there is no variable, the locator itself functions as if it were a variable, which then is either present or absent. Neomorphic characters are regarded as “variables” only in the general sense that all character statements include alternative character states. Neomorphic character statements, nonetheless, differ logically from transformational character statements. A new bone, such as the predentary of ornithischian dinosaurs, is a neomorphic character that lacks any comparable transformational state among outgroup taxa. The character simply identifies the bone. There is no hypothetical or proposed transformation between the predentary and another bone; the bone is either present or absent.

These characters have been mixed together with others in reductive coding schemes that collectively have been termed “nominal variable” or “presence–absence” characters (Pimentel and Riggins, 1987; Pleijel,

1995). Here the character statement pattern outlined above is termed neomorphic, because they record the presence/absence of “new” structures that do not have comparable, recognizable transformational states. The absence of comparable, recognizable transformational states does not invalidate or weaken a character; the shared origin or loss of a novel structure, which has no precedence or comparable condition in other taxa, often presents convincing, verifiable, character evidence for monophyly. Whether or not such a shared character state is a synapomorphy for a particular clade, of course, is a secondary question evaluated by phylogenetic analysis (congruence).

Neomorphic characters as described above are equivalent to “evolutionary novelties” when these are defined as follows (Müller and Wagner, 1991, p. 243): “A morphological novelty is a structure that is neither homologous to any structure in the ancestral species nor homonomous to any other structure of the same organism.”

These authors objected to equating “evolutionary novelty” and “apomorphy”, choosing to exclude variation in form (size, shape) and differentiation of repetitive parts. They restricted “novelties” to apomorphies that constitute “new elements”, avoiding in this connection any “assumptions about the mechanistic bases of novelties.” Although they prefaced their remarks by claiming that they plan to ignore a “purely phenomenological treatment of novelty” as in “taxonomic characters” (Müller and Wagner, 1991, p. 230), that is exactly what their definition achieves. As defined above, an “evolutionary novelty” differs from a neomorphic character only by their decision to exclude “absence” or “loss” of a structure as a “novelty.” Neomorphic characters are here defined to include both the appearance and disappearance of features that lack comparable transformational states in other organisms.

Transformational characters, unlike neomorphic characters, include a variable. Transformational character states are mutually exclusive conditions of that variable (Table 5). Transformational characters, when completely formulated with an appropriate variable, “imply their respective conditions” (Pogue and Mickevich, 1990, p. 320). The variable in a transformational character restricts the realm of possible character states. In character statement 4, the variable “length” is compatible with a limited range of absolute or relative character states that describe linear dimension. Unlike neomorphic character statements, transformational character statements link together multiple observations as alternative conditions of the same attribute, a hypothetical construct that functions as a variable.

Unfortunately, these two character types, neomorphic and transformational, are often (1) intermixed, or (2) incompletely formulated in the cladistic literature, blurring their phenomenological distinction. In the first case,

the neomorphic character state “absent” is often mixed with transformational states, creating a chimerical character statement that is not completely independent of other character statements. As explored in further detail below, this violates precept I regarding character independence, as was noted long ago (Colless, 1985, p. 232). In the second case, transformational characters are often expressed without a variable, which then is implied by—or must be inferred from—the character states. An incomplete transformational character of this sort superficially resembles a neomorphic character, which does not have a variable.

I argue here that the two character types are logically distinct, even when apparently describing the “same character.” Consider the following examples:

**15.** Manual digits, number: five (0); four (1).

**16.** Manual digit I: present (0); absent (1).

**17.** Manual digits, formula: I–V (0); II–V (1).

If a five-digit hand is primitive and four-digit hands are observed, a transformation has occurred (character statement 15). If one can identify a specific digit as the one that has been lost (e.g., digit I), the character is neomorphic (character statement 16). The two are logically distinct and arise from different character and character state conceptions. In the former, transformational conditions are observed in digit number that do not identify specific digits. This may be due to developmental evidence regarding the nature of the transformation or simply to ignorance or ambiguity in digital identity. In character statement 16, manual digit I is specifically identified as present in some taxa and absent in others. Those with four digits have no comparable transformational feature.

In this case, “gain” or “loss” of a feature is *not* regarded as transformation, or literally “changed form.” The feature (digit I) has no “form” when “absent”; “absence” is a conditional “state of being” (in this case interpreted as “loss”). Of course, both transformational and neomorphic character statements involve “change” in observed taxa, otherwise they would be discarded as invariant. Nevertheless, if digit I in the above example is determined to be the focal point of change, it is logically incorrect to use all manual digits as a covering transformational character as in character statement 17; digits II–V do not change and represent superfluous redundancy in the character states.

When change is restricted to the presence or absence of digit I (properly formulated in character statement 16), the locus of ignorance is refocused. In the first case, a digit was lost, although its particular identity was unknown (character statement 15); the observed transformational states (five, four) are all that was known. In the second case with digit I identified as the locus of change, we do not know what became of this digit in four-fingered hands and refer to that condition as “absent.” In other words, in the first case we

observe a transformation from one state to another, the underlying nature of which is uncertain. In the second case, we observe a feature, which has no observable transformational condition in some individuals. Some would claim that such a feature (digit I) is truly absent in these four-fingered individuals; others would claim that such a feature is really there but unrecognizable; still others would claim that we simply cannot know. All I am claiming in this connection is that this kind of character is not the same as those linked to observable, comparable transformational states. I refer to the former as neomorphic (“new form”) characters. Neomorphic character statements do not involve postulated transformation beyond the appearance or disappearance of the character. Transformational character statements involve “changed form” not simply the *de novo* appearance or disappearance (presence, absence) of a particular feature that otherwise has no alternative, variant condition.

This distinction will doubtless raise objections from those who contend that *all* characters must come from, or vanish into, something. Brower and Schawaroch (1996, p. 269) claimed that *all* characters comprise “transformations of more general characters.” Platnick (1979, p. 543) wrote: “A character is thus a theory, a theory that two attributes which appear different in some way are nonetheless the same (or homologous). As such, a character is not empirically observable; hence any (misguided) hope to reduce taxonomy to mere empirical observation seems futile.”

These are examples of what I have termed transformational character statements, which attempt to link alternative conditions as mutually exclusive states of a variable attribute. This is not true of neomorphic characters, which simply point to a feature that is either present or absent. This fundamental distinction remains, even when “absence” is taken to mean only the “inability to recognize.” I have highlighted this distinction among character statements not merely for its epistemological content, but because it plays a major role in current controversies regarding methods for character coding as seen below.

Systematists have overlooked this simple logical division among character statements (Ghiselin, 1984; Colless, 1985; Rodrigues, 1986). Ghiselin (1984, pp. 104–5), for example, did not distinguish neomorphic and transformational characters and, further, questioned the distinction between character and character state: “Another problem is a curious tendency to treat characters as if they had to be substances. Substance...in its most basic sense means the concrete, individual thing...Characters, can fall under other categories than substance; for example, quality (a color)...The problem is evident in discussions of what are called ‘character states’—in which the character is generally identified with a part and the state with an attribute.”

A neomorphic character points to a structure and in this sense is a “substance”, which is either present or absent. A transformational character, in contrast, includes a variable (something that varies), which exists as mutually exclusive conditions or “attributes.” Both neomorphic and transformational characters, thus, are organismal features expressed as independent mathematical variables in quantitative cladistics; character states constitute the variable conditions (Tables 3 and 4).

Colless’ (1985, p. 230) approach recognized three distinct uses of the term “character”, the first as “attribute”, the second as “part”, and the third as “variable”: “Taking a simple example of a taxonomic datum, that A has brown wings, then: character<sub>1</sub> (= attribute) is ‘has brown wings’...; character<sub>2</sub> (= part) is ‘wings’; character<sub>3</sub> (= variable) is ‘wing color’.”

Colless (1985, p. 232) then distinguished character variables as either “intension” (feature plus variable) or “extension” (feature plus list of alternative attributes) and further confused his scheme with comments on “presence” and “absence” as states. In the present paper, “wing” is a locator, “color” is a variable, and “brown” is a character state; all compose easily identified components of a transformational character statement, symbolized as  $L_1, V: v_0; v_1$  and written as:

Wing, color: brown (0); white (1).

Despite some confusion of terms, Colless’ understanding of the role of character statements in quantitative phylogenetics was clearly stated: “‘character’ is now widely employed...for a set of mutually exclusive attributes (Colless, 1967), constituting a logical or mathematical *variable*—or in scholastic logic, a fundamentum divisionis (Ghiselin, 1984).” Colless’ “character-attribute” is nothing more or less than a *character state* from a transformational character (Thiele, 1993). Colless’ “character-part” is nothing more or less than a *locator* in either a neomorphic or transformational character. He was correct in identifying the ambivalent use of “character” to refer to both a character state and part, or all, of a character, which continues unabated today (e.g., “shared derived characters” versus the more accurate “shared derived character states”).

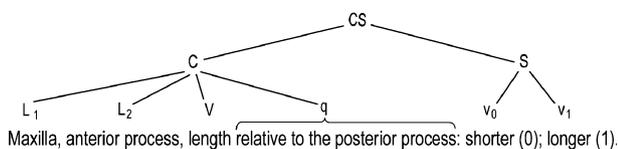
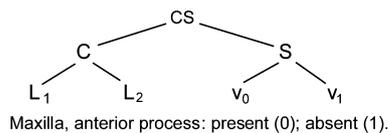
Other authors have subdivided character statements by their coding peculiarities or complexities. Forey and Kitching (2000) focused on character coding, expanding Pleijel’s (1995) fourfold division of characters to six categories. Hawkins (2000) divided character statements into “conventional” and “non-conventional” categories, the latter subdivided into nominal variable, unspecified homolog, composite, ratio, logically related, conjunction, unifying, inapplicable data, positional, and mixed. Both “conventional” and “non-conventional” categories include neomorphic and transformational character statements, several non-conventional categories violate either precept I (character independence) or precept II (mutual exclusivity of character states), and several of

the categories are descriptive rather than logical (e.g., ratio characters).

Later I present a descriptive division of transformational characters as well (Table 7); what is of concern here, however, is a more fundamental division of character type. If we respect the two operative precepts cited above for character data under maximum parsimony and search for fundamental patterns for characters and character statements, only two fundamental patterns exist: neomorphic and transformational.

#### A “generative grammar” for character statements

The approach outlined here recognizes phylogenetic character data as a language among systematists. The more precise terminology above recognizes character statements (CS) as composed of characters (C) and statements (S), the former composed of locators ( $L_1$ ,  $L_2$ ) and (optionally) a variable (V) and variable qualifier (q) and the latter composed of character states ( $v_0$ ,  $v_1$ , etc.) and (optionally) a variable qualifier (q). All character statements are composed of these four logical components (locators, variable, variable qualifier, character states; Table 4), which are combined in only two specific patterns, neomorphic and transformational (Table 5). The two fundamental patterns for character statements are depicted below for two hypothetical characters (Scheme 3).



The tree that links their logical components, a *character-statement tree*, outlines the structure of the character statement with a function analogous to “phrase structure trees” in generative grammar (Chomsky, 1965). The root node is the character statement (CS), which is always divisible into character (C) and statement (S). Terminal, or leaf, nodes include the four basic components that compose character statements (locators =  $L_1$ ,  $L_2$ ; variable = V; variable qualifier, q; character state =  $v_0$ ,  $v_1$ ).

Variation observed in real character data for each of these patterns points recursively to a small set of “production rules” that compose a “generative grammar” for morphological character statements. Neomorphic character statements show less variation than

transformational character statements. In both neomorphic and transformational characters, there may be one or more secondary locators. In neomorphic characters the only other regular variant that I have observed is the addition of the word “rudimentary” or an equivalent (“poorly developed”) in combination with either “absent” or “present.” In these cases, a phylogeneticist is lumping within a single character state an incipient or remnant condition with one in which the structure is absent or indisputably present (discussed in more detail below). Transformational character statements vary in the position of the variable qualifier, if one is present, and in the number of character states. Completely formulated morphological character statements, thus, are quite limited in structure and logical composition and are guided by a small set of five “production rules”: (1) character statements are composed of characters and statements, the former always including a single primary locator and the latter always including at least two character states; (2) characters may include one or more secondary locators; (3) neomorphic character statements utilize only two character states (absent, present) with optional inclusion of near-absent or near-present conditions; (4) transformational character statements include one qualifier and, optionally, more than two character states; and (5) transformational character statements may include a variable qualifier positioned either in the character or in the statement.

#### Presentation order

Using character statement 4, six alternative versions and their symbolic notation are presented in Table 6. All six are complete transformational character statements; all include the same logical components; all would be regarded as acceptable by cladists. The first two are preferable, as explained below. Reducing variation in the manner in which we express transformational character statements will enhance data comparison, which is an increasingly significant hurdle in morphology-based phylogenetics.

#### Secondary locators first

In character statements 1 and 2 in Table 6, the secondary locator “Maxilla” is positioned before the primary locator “anterior process.” The more general anatomical term thus is listed before the more specific. Hierarchical ordering within a character, from general to more specific, mimics the use of headers to subdivide lists of character statements (e.g., character statement 9). It is preferable, then, to position the most general locator in the lead position in a character statement.

Is there any reason to list the more specific primary locator first (Table 6, character statement 3), a locator

Table 6  
Character statement variation, symbolic notation, and comments

No.	Character statement variants	Symbolic notation	Comments
1	Maxilla, anterior process, length relative to the posterior process: shorter (0); longer (1)	$L_2, L_1, Vq: v_0; v_1$	Preferred form
2	Maxilla, anterior process, length: shorter (0), or longer (1), than the posterior process	$L_2, L_1, V: v_0, v_1, q$	Preferred form
3	Anterior process of the maxilla, length relative to the posterior process: shorter (0); longer (1)	$L_1L_2, Vq: v_0; v_1$	Inverted locators ( $L_1$ before $L_2$ )
4	Maxilla, length of the anterior process relative to the posterior process: shorter (0); longer (1)	$L_2, VL_1q: v_0; v_1$	Split locators (variable in between)
5	Length of the maxillary anterior process relative to the posterior process: shorter (0); longer (1)	$VL_2L_1q: v_0; v_1$	Variable before locators
6	Length of the anterior process of the maxilla relative to the posterior process: shorter (0); longer (1)	$VL_1L_2q: v_0; v_1$	Variable before inverted locators ( $L_1$ before $L_2$ )

Table 7  
Kinds of transformational morphological characters

No.	Category	Hypothetical example	Description of character states
Quantitative			
1	Quantitative-absolute	Calyx, circumference: 1–2 cm (0); 3–5 cm (1)	Quantitative measures or dimensionless numbers specified in absolute terms (e.g., “5” or “1.5 cm” or “60°”)
2	Quantitative-relative/linear	Maxilla, anterior process, length relative to the posterior process: shorter (0); longer (1)	Dimensionless comparisons or ratios involving linear measurement; quantification involves relative terms (e.g., “shorter than X” or “30% of” X)
3	Quantitative-relative/geometric	Maxilla, maxillary fenestra, shape: circular (0); oval (1)	Dimensionless comparisons or ratios involving geometric shape; quantification involves relative terms (e.g., “larger than X” or “most acute angle”)
Qualitative			
4	Form	Brow ridges, form: rounded (0); prominent, protruding (1)	Any qualitative aspect of size/shape
5	Appearance	Petal, color: yellow (0); orange (1)	Any aspect of color or surface appearance (translucency, reflectivity, texture, etc.)
6	Topology	Antorbital fossa, configuration of dorsal margin: nasal participates, separates maxilla and lacrimal (0); nasal excluded, maxilla-lacrimal contact (1)	Topologically expressed variation
7	Composition	Medial distal carpal composition: distal carpal 1 (0); distal carpals 1 + 2 (co-ossified) (1)	Compositionally expressed variation
8	Ontogeny	Otic region, timing of exoccipital–opisthotic fusion: post-hatching (0); embryonic (1)	Ontogenetically expressed variation

(“anterior process”) that by itself may not be singular? By listing the most general secondary locator at the front of a character, one can group together characters that involve the same general secondary locator, such as characters of the “maxilla.” Organizing characters in this manner often brings in proximity characters of related interest and thus is more useful for those examining or searching character data.

*Variable and variable qualifier last*

In character statements 1 and 2 in Table 6, the variable “length” is positioned after all locators and just in advance of the variable qualifier (“relative to the posterior process”) and characters states (“shorter”,

“longer”). The variable then is positioned near the entity that qualifies the variable as well as the variant conditions. As these components (variable, variable qualifier, character states) are functionally the most closely related, the end of the character is viewed as the preferable position for the variable and variable qualifier.

The variable, nonetheless, can be positioned anywhere within the character, such as at the front end or between locators (Table 6, character statements 4–6). Yet, is there a plausible reason to place the variable before locators (Table 6, character statement 5)? Variables, such as “length”, are rarely unique aspects of a character. To my knowledge, no one has ever grouped character statements according to their variables. As we

will see below, omitting variables altogether is a common problem in character data. Variables are preferably positioned at the end of the character adjacent to the components to which they are most closely allied (variable qualifier, character states).

### Transformational character statements

Character statements are first divided, as argued above, into neomorphic and transformational based on two discrete logical patterns. Neomorphic characters point to a structure that is typically scored as absent or present. The existence, or “state of being” of the structure is the point of focus, not its form or any other variable attribute.

Morphology-based transformational character statements, in contrast, come in a great variety of forms. Many descriptive criteria could be used to subdivide such character statements, such as anatomical region or even original author. The most general and widely cited subdivisions, however, seek to identify the inherent nature of the character statement. “Discrete” versus “continuous” and “qualitative” versus “quantitative” have been proposed, the former focusing on the nature of character states and the latter on the nature of the character. I consider the latter, “qualitative” versus “quantitative”, to be the more fundamental of the two, as it allows further subdivision based on the nature of the character. Characters are arguably the most substantive issue in character data (Table 7). Only morphology-based characters are discussed here; behavioral, molecular and other kinds of character statements are not considered.

#### *All morphological characters could be “quantitative”*

Wiley (1981) argued that “qualitative” and “quantitative” refer to modes of expression rather than intrinsic qualities of character data. Baum (1988) and Stuessy, (1990) argued further that all morphological characters are quantitative, or could be expressed in quantitative terms. Both of these propositions seem to be true. Most so-called qualitative characters could be redefined in quantitative terms. The shape “oval”, the texture “rugose”, the color “red”, and even an ontogenetic fusion could be described in quantitative terms, although doing so may not enhance communication or understanding among phylogeneticists.

Qualitative characters persist because they are an efficient means to describe features, such as “variegated” or “spiral”, that are easily recognized by qualitative perception but complex or problematic when defined quantitatively. A more detailed qualitative description for “spiral”, for example, is a “curve, which turns around some central point or axis, getting progressively

closer to or farther from it.” A true mathematical description would be even more intuitively abstract although more precise—and possibly necessary for a gastropod taxonomist. The modern field of theoretical morphology has increasing reach across a wide spectrum of organisms and is predicated on the mathematical description of form, both simulated and actual (McGhee, 1999; Eble, 2000).

Although qualitative character statements may only exist by preference and convenience rather than logical necessity, they will surely continue to claim a large part of the character data analyzed by morphology-based phylogeneticists. A qualitative character, thus, may be defined as one that is difficult to express in terms of absolute or relative magnitude (Table 3). I adopt a primary division of transformational character statements into “quantitative” versus “qualitative” and further subdivide each based on character patterns common in the cladistic data (Table 7). This classificatory scheme is presented as a heuristic tool rather than a rigid or complete system; some character statements can be allocated to more than a single category.

#### *Quantitative character statements*

Quantitative character statements are subdivided on the basis of mensural states that are either absolute or relative. Character states composed of absolute values (e.g., “3”) or an absolute range are comparatively rare, due to size variation related to growth, sexual dimorphism or population variation. To overcome this variation, quantitative character states are often based on relative measures, either linear or geometric. Character states describing linear measures, such as “less than 50%” or “longer than bone X” are relative linear measures. Relative geometric character states often specify particular shapes or angles, such as “less than 90°”.

#### *Qualitative character statements*

Qualitative characters are subdivided into five categories based on the nature of the variable: form, appearance, topology, composition and ontogeny (Table 7). Form-characters include any aspect involving size, shape or any other structural feature, such as texture. Appearance-characters include color, odor or any other sensual attribute of an organism (excluding behavior/function) that does not involve the assessment of form. Topological characters involve positional information, such as the relation between structural elements and do not specifically address form. Compositional characters involve the identification of component parts, and do not specifically address form or position. Ontogenetic characters specifically address transformational, topological or compositional variation during development.

## Common logical shortcomings

### Missing variable

The variable is often omitted from transformational characters (Table 8, character statement 1). The variable, however, is not an optional component of a transformational character; it is a logical necessity to state as clearly as possible the variable attribute rather than site only the location of variation (locator). When the variable is omitted, it exists as an unstated inference (Table 8, example 1). A scholar must infer the variable from the character states. Sometimes this is little more than a nuisance, as the missing variable, say “length” or “form”, may be obvious. In many cases, to the contrary, the variable is not obvious. Its precise specification may be critical to the structure and scoring of one or more character statements. Specifying the variable in precise terms and making sure that the variable and its

associated character states are logically consistent is an essential ingredient of well-considered character data.

Thiele (1993, p. 298), for example, listed the following three character statements:

Adult leaf margins: always strongly toothed (0); tending entire (1); entire (2).

Adult leaf margins: flat or recurved (0); revolute (1).

Adult leaf margins: undulate (0); flat (1).

A significant proportion of his character statements include redundant characters followed by differing sets of character states, a pattern that is occurring more frequently as cladistic data sets increase in size. Are these duplicate and triplicate characters really the same? If we accept them at face value as a single character, “adult leaf margin” (with variables omitted), their states should be combined into a single seven-state statement. There is no logical recourse to avoid seven states, if indeed they are the same character with mutually exclusive conditions. These characters, however, were

Table 8  
Typical (hypothetical) transformational characters of poor construction (corrected versions in italics)

No.	Character statement	Problem	Critique
1	Sepal: 5 (0); 6 (1); 8 (2) <i>Sepal, number: 5 (0); 6 (1); 8 (2)</i>	Variable missing	This “quantitative-absolute” character does not specify the variable
2	Maxilla, anterior process, length: short (0); long (1) <i>Maxilla, anterior process, length relative to the posterior process: shorter (0); longer (1)</i>	Variable qualifier missing	This “quantitative-relative/linear” character (Table 7, example 2) uses relative terms (short, long) for states but does not specify the comparative linear dimension
3	Maxilla, anterior process, length: shorter relative to the posterior process (0); longer relative to the posterior process (1)( <i>see character 2</i> )	Variable qualifier misplaced	The variable qualifier does not vary and thus is not part of a character state
4	Maxilla, anterior and posterior processes: present (0); absent (1) <i>Maxilla, anterior process: present (0); absent (1)</i> <i>Maxilla, posterior process: present (0); absent (1)</i>	Twinned locators	If anterior and posterior processes are not always correlated or preserved, they are better as separate characters
5	Sepal, number: 5 (0); 6 (1); 8 (2); absent (3) <i>Sepal, number: 5 (0); 6 (1); 8 (2)</i> <i>Sepals: present (0); absent (1)</i>	“Absent” as zero	“Absent” is not a number
6	Maxilla, anterior process, length: shorter relative to the posterior process (0); longer relative to the posterior process (1); absent (2) <i>Maxilla, anterior process, length: shorter relative to the posterior process (0); longer relative to the posterior process (1)</i> <i>Maxilla, anterior process: present (0); absent (1)</i>	“Absent” as length	“Absent” is not a length nor magnitude of any kind
7	Maxilla, anterior process, shape: tab-shaped (0); prong-shaped (1); absent (2) <i>Maxilla, anterior process, shape: tab-shaped (0); prong-shaped (1)</i> <i>Maxilla, anterior process: present (0); absent (1)</i>	“Absent” as shape	“Absent” is not a shape

never intended to be the same, but rather owe their identical appearance to missing variables. The first character state in the first character statement refers to the strength of tothing, whereas the second and third states refer to the distribution of tothing. The unstated variable of the first character therefore might be “tothing” or perhaps “extent of tothing”, which would complete the character as “Leaf margin (adult), tothing” or “Leaf margin (adult), extent of tothing”, respectively. The second and third characters may be independent, but it is not immediately obvious how their variables would differ, especially given that both are associated with the state “flat.” The key point here is that precise delineation of the variable is a critical, required component of transformational characters, but one that is often omitted.

A second and widespread means to omit variables from characters is to seamlessly blend character and character states, as Thiele (1993; Table 1) does elsewhere:

Dorsal fin anterior; posterior

Canalis caroticus internus formed by pterygoid to distal length; not

These two character statements lack variables that in both cases help to mask common character shortcomings. In the first case, the missing variable, “position”, is intuitive but helps to mask a second problem—a missing variable qualifier.

Dorsal fin, position: anterior (0); posterior (1).

Adding the variable “position” properly frames the associated relative character states, “anterior” and “posterior”, but underscores the absence of a variable qualifier critical to such relative measures (discussed below). “Anterior” or “posterior” to what, one may well ask? Presented with a random fish, one would not be able to score this character with confidence. Systematists who have coined such incomplete, untestable character statements—myself included—have a false sense of security that other scholars will come to the same conclusions given similar comparative materials. Below I add a hypothetical variable qualifier:

Dorsal fin, position relative to the pelvic fin: anterior (0); posterior (1).

In the second of Thiele’s characters listed above, the omission of the variable helps to obscure logical inconsistencies in the character states. The missing variable might be “composition of canal” or perhaps “pterygoid contribution.” The first character state, “formed by pterygoid”, suggests the former, as if the character concerns the bony contribution to the wall of the canal. The second character state, “not”, suggests the latter, as if the character concerns only the degree to which the pterygoid participates in the canal. The character states appear to be logically inconsistent or, at the very least, incomplete. Both options are completed (hypothetically) below:

Canalis caroticus internus, bony composition: pterygoid (0); pterygoid and palatine (1).

Canalis caroticus internus, degree of pterygoid contribution: pterygoid forms more (0), or less (1), than 50% of canal length.

#### *Missing variable qualifier*

The variable qualifier, like the variable, is often subject to omission from transformational characters, which similarly complicates their evaluation in taxa. The most common missing variable qualifier involves “quantitative relative/linear” characters (Table 7, example 2). The quantitative measure in such characters is relative; the variable qualifier supplies the comparative measure. When omitted, the character states must be taken at face value, as they are impossible to formally evaluate.

The character “Maxilla, anterior process, length” is a quantitative relative/linear character when followed by relative character states such as “short” and “long” (Table 8, example 2). A phylogeneticist composing such a character has evaluated the length of the process in taxa and decided that some are “short” and others “long” relative to another measure or set of measures. The relative comparisons, when left unspecified, are difficult to evaluate. The phylogeneticist presumably anticipates that the proposed relative length difference will show up in so many comparative measures it will be relatively obvious to other scholars. This is not always, or even typically, the case.

Thiele (1993, Table 1) correctly identifies these as “ratio characters” while failing to realize that the ratio is incomplete, that a character statement so constructed is tantamount to being given the numerator but not the denominator of a ratio. The two characters cited above involving the dorsal fin and relative contribution of the pterygoid to the internal carotid canal are examples of “relative-geometric” and “relative-linear” quantitative characters, respectively (Table 7, examples 2, 3). Although ratios have a contentious history in systematics (Atchley et al., 1976; Dodson, 1978), they are eminently testable, and testability is the sine qua non of a phylogenetic character statement.

#### *Misplaced variable qualifiers*

Variable qualifiers are sometimes incorrectly inserted into character states (Table 8, example 3). This is regarded here as logically inconsistent. The variable qualifier qualifies the variable; it is not a part of the variable condition (character state). Because the variable qualifier is not part of a character state, it must be inserted redundantly within each character state in the statement. Misplaced, redundant variable qualifiers are easily rectified by repositioning a single copy after the variable in the character or after the character states in the statement (Table 6, character statements 1, 2, respectively).

### *Entangled characters*

“Entangled” may be the best way to describe some character constructions such as the following (Lee, 1998):

Stapedial footplate not surrounded (0), tightly surrounded (1) by bony ridges projecting from lateral surface of braincase.

Lee and Caldwell (2000) subsequently rejected this character statement altogether because they regarded the state “not surrounded” as too restrictive. In their critique, Rieppel and Kearney (2002) discussed the above character statement in particular, arguing that it “fails the test of similarity because of the lack of topological correspondence of the bony flanges” in two of the taxa under consideration (mosasaurs versus snakes). Such lack of correspondence, to the contrary, does not invalidate the character statement but rather changes the way taxa are scored. The real impediments to interpretation and testing in this case lie in the entangled structure of the character statement. What is the character—“stapedial footplate” or “bony ridges”? What is the variable—“tightness”? Lee’s transformational character could be reformulated (hypothetically) much more clearly in the model format outlined above (Table 6, example 2):

Braincase, lateral bony ridges, degree of encirclement of stapedial footplate: less than 15% (0), or greater than 50% (1).

The original formulation by Lee (1998): (1) comingled components of the character and statement; (2) positioned the most general locator (“braincase”) at the end of the character statement rather than at the front of the character; (3) omitted the variable (“degree of encirclement of stapedial footplate”); and (4) failed to give a complete relative measure or variable qualifier (“footplate circumference”). The main point here is that character statements are much easier to evaluate and test when properly and completely formulated.

### *Twinned character components*

Sometimes two primary locators, variables or variable qualifiers are combined in a single character statement (Table 8, example 4). These instances are relatively rare compared with the problems cited above. Primary locators and qualifiers are twinned because a phylogeneticist regards the pair as correlated (Table 8, example 4). Character correlation is an important issue in character data. The view taken here is that character correlation should be evaluated prior to the formulation of character statements, which should only offer for evaluation a single character. If twinned characters are not completely correlated or if only one or the other is known, then scoring ambiguity may occur.

Twinned variable qualifiers can create the same ambiguity for the same reasons, as in the following examples:

18. Manual digit I, length relative to manual digits II or III: shorter (0); longer (1).

19. Manual digit I, length relative to either manual digits II or III: shorter (0); longer (1).

20. Manual digit I, length relative to both manual digits II and III: shorter (0); longer (1).

21. Manual digit I, length relative to manual digit II: shorter (0); longer (1).

22. Manual digit I, length relative to manual digit III: shorter (0); longer (1).

The first two versions allow ambiguity in scoring, because if either digit II or III is shorter (or longer) than digit I, the remaining digit will be longer (or shorter) and create an ambiguous scoring situation. Version 20 encounters similar problems, as there exists no character state to accommodate the mixed condition (with either II or III shorter or longer than I). This character statement, thus, presumes that both relative measures for digit I are correlated and will never be disassociated; it thus effectively down-weights the character (if all are given equal weight). Versions 21 and 22 simply split the pair of variable qualifiers into two separate character statements, which remains the best solution if character correlation is not securely established.

### “Absent” as a character state

Controversy over the use of the character state “absent” in character coding has become a cause célèbre in large measure because precepts I and II have not been uniformly adopted and because incomplete transformational characters take the form of neomorphic characters. As a result, “absent” has been erroneously inserted into transformational character statements. I suggest below that the character state “absent” has a proper home only within neomorphic character statements.

Before this issue is addressed, however, the use of “absent” in any context must be defended, because both the origin and loss of neomorphic structures have been relegated to the unknowable by some cladists. Pimentel and Riggins (1987, p. 206) posited that primitive “absent” is unknowable: “For example, carpels are present in all angiosperms and vertebrae in all vertebrates but are unknown for any other group of organisms. ‘Absence’ of carpels or vertebrae therefore is neither a character nor a character state (cf. Nelson, 1978).”

Nelson (1978, p. 344), contrary to the citation in the above quotation, argued that derived “absent” (i.e., loss) may be unknowable: “One may doubt, for example, that any characters are truly lost, rather than transformed. Apparent loss may be an indication that the characters and transformations are merely poorly understood and, consequently, wrongly defined.”

This follows a time-honored tradition within phylogenetics that has viewed “loss” (derived “absent”) with

skepticism, allegedly because the particular pathway toward the elimination of a structure may not be known in any detail (Hecht and Edwards, 1976, 1977). The view taken here is that “present” and “absent” conditions of a structure are among the clearest of morphology-based character states. How it could be viewed otherwise is baffling. The interpretation of “present” or “absent” states as either primitive or derived within the ingroup, of course, is determined a posteriori by phylogenetic analysis. This allows interpretation of derived states as either independent “origin” or “loss.” Ambiguity in definition, logical structure, and testability in scoring are typically greater among transformational character statements.

The other central controversy surrounding “absent” is how to code inapplicable features, i.e., features pertaining to a structure that that is “absent” in one or more taxa (Maddison, 1993). Should we combine “absent” with other character states, or use a “?” for inapplicable data? I suggest below that the inapplicable character state scored with a “?” is a logically consistent and requisite component of character data involving neomorphic and transformational characters.

#### *Neomorphic “absent”*

A typical neomorphic character statement includes a locator, or locators, followed by the character states “present” and “absent.” “Present” and “absent” record the “state of being” of a locator. The main ambiguity involved in neomorphic characters occurs when the locator is “almost present” or “almost absent”, in which case “rudimentary”, “poorly developed”, or less desirable polarized equivalents (incipient, vestigial) are attached to either “present” or “absent” states. Phylogeneticists inevitably must accommodate the ambiguous margins of “present” and “absent”, given the continuous nature of biological variation and the gradual origin and loss of most structures.

“Rudimentary” and its equivalents, nevertheless, are transformational character states involving the minimal size of a structure. If no absolute measure is given, “rudimentary” is understood as a relative linear or geometric dimension. “Present” and “absent”, on the other hand, are neomorphic character states that record “state of being” not an assessment of linear or volumetric dimensions. Technically speaking, “rudimentary” should be included as the minimal condition of a character concerned with the size or magnitude of a structure and not its presence or absence. Indeed, as argued below, when such a size character exists, it is inappropriate to include “absent” as a character state. However, when no such size character exists, appending “rudimentary” or its equivalents to either “present” or “absent” ranks as a common, and perhaps unavoidable, exception to the clean separation between neomorphic and transformational character statements.

It should be noted here that “absence” and “loss” are not the same. “Absence” is an observable condition. “Loss”, in contrast, is a relational concept that infers former presence; it is a secondary interpretation of absence in the light of a phylogenetic hypothesis.

#### *Transformational “absent”*

Two minus two equals “zero” not “absent.” “Zero” is a number; “absent” is a “state of being.” Surprising it is, then, that “absent” routinely supplants zero as a character state in quantitative-absolute transformational character statements (Table 8, example 5). Likewise, when the jaws of a digital calipers close, the dial reads “0.00” not “absent.” “Absent” is not a length measurement. Of course, neither is “0.00.” The caliper dial reads “0.00” when you have nothing to measure. It is nonsensical to have a character state “absent” or “zero” for a character involving the length of a structure. To measure length requires the presence of a measurable feature. “Absent”, likewise, is not a measure of shape or color. Yet, “absent” is routinely included in character statements regarding the shape or color of a structure (Table 8, example 7).

The primary reason I believe “absent” has so frequently been mixed with transformational character states is due to missing variables. As pointed out above, when the variable of a transformational character is omitted—say “length” or “color”—the character is composed only of one or more locators and is neomorphic in form (Table 5). “Absent” then becomes an intuitively reasonable, albeit logically inappropriate, transformational character state. These seemingly innocuous logical shortcomings involving “absent” affect subsequent phylogenetic analysis and are at the center of opposing character coding schemes termed “multistate” and “contingent”.

In “multistate” coding, “absent” is included as a character state alongside transformational character states. The fairly obvious shortcoming here involves precept I, character independence, as has been noted by several authors (Maddison, 1993; Pleijel, 1995; Forey and Kitching, 2000; Hawkins, 2000). If “absent” is included as a character state alongside others that record variation in the length of a structure, how do we code other variation in that structure, say its color or shape? If we again create an “absent” character state, it will redundantly record the absence of the structure, violating character independence (Maddison, 1993). One solution, which mercifully seems never to have gained many adherents, was proposed by Maddison (1993, p. 580):

perhaps the best general solution to this problem is to fuse characters that include the same state “tail absent” into a single character. In the example given, the two characters of tail color and tail covering would become a single five-state

character with states tail absent, tail blue and scaled, tail blue and hairy, tail red and scaled, and tail red and hairy, and an elaborate step matrix would be used to indicate the cost of various transformations of state.

Wilkinson (1995, p. 298) called this “composite” coding, the assembly of trait combinations into individual character states.

A simpler solution, termed “contingent” coding by Forey and Kitching (2000, p. 56), separates the presence–absence of a structure from character statements concerning its variation. This effectively separates neomorphic and transformational components. For taxa that lack a particular structure, transformational character states are coded with a “?” for condition unknown, or inapplicable. Hawkins (2000) called this “inapplicable data” coding, and listed it with several other coding patterns as a “non-conventional” method. Forey and Kitching (2000) mistakenly claimed that Hawkins (2000) termed this coding scheme “conventional” coding. Hawkins, to the contrary, used “conventional” coding as a general term for transformational coding (Hawkins et al., 1997, p. 277; Hawkins, 2000). “Contingent” coding involves the assignment of the presence or absence of a structure to its own character. But this is the extent to which an otherwise typical transformational character is subdivided. When taken to an extreme, subdividing all transformational character states into binary presence–absence character statements has been called “presence–absence” (Pleijel, 1995) or

“reductive” coding (Wilkinson, 1995). As reductive coding clearly violates both precepts I and II, it will not be discussed further here.

The quantitative consequences of “contingent” versus “multistate” coding are most easily evaluated in the context of a pair of character statements that may be combined into a single multistate character. Figure 1 uses three-taxon diagrams to compare the effect of coding “absent” with a pair of contingent character statements or as a single multistate character. Consider three taxa, X without brow horns, Y with short brow horns, and Z with long brow horns, and the following three character statements (labeled characters I–III, respectively, in Fig. 1):

23. Brow horns: absent (0); present (1).

24. Brow horns, length relative to nasal horn: subequal (0); longer (1).

25. Brow horns: absent (0); subequal to nasal horn (1), longer than nasal horn (2).

Character statements 23 and 24 (Table 9, characters I, II; Fig. 1) are neomorphic and transformational, respectively. Character statement 25 (Table 9, character III), in contrast, is a chimera; the character is neomorphic in form (i.e., it has no variable such as ‘length’) but also includes transformational (subequal, longer) character states.

First, ‘absent’ as a primitive character state is considered (Fig. 1A–C). In this case, outgroup taxa do not have the horns of interest. Coding neomorphic and

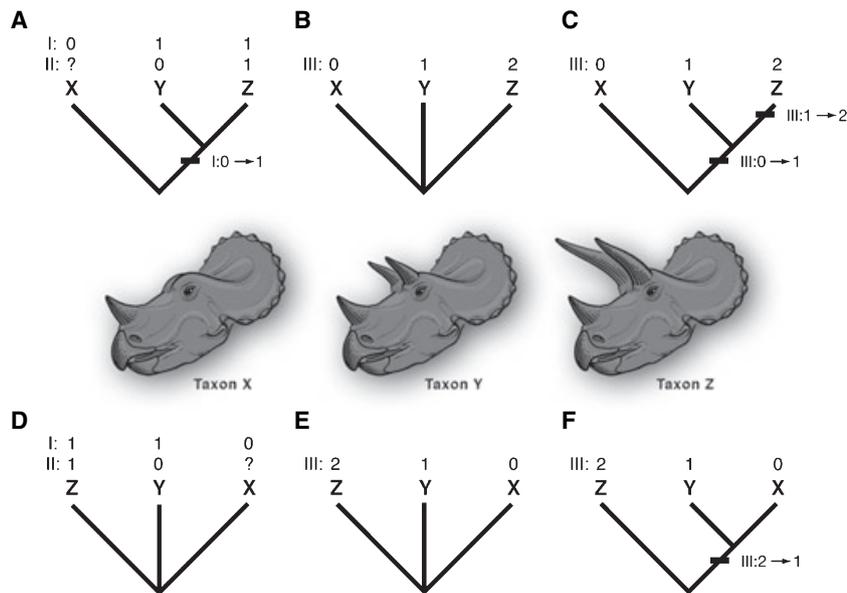


Fig. 1. The phylogenetic effects of separating or combining neomorphic and transformational states, using three hypothetical frilled dinosaurs (Taxon X, Y, Z) and character data on the presence and length of brow horns as scored in Table 9. (A–C) ‘Absent’ as a primitive condition: (A) separate neomorphic and transformational characters; (B) single unordered character; (C) single ordered character. (D–F) ‘Absent’ as a derived condition (i.e., loss): (D) separate characters; (E) single unordered character; (F) single ordered character. Character state distributions are shown above cladogram; unambiguous character state transformations are indicated on each cladogram.

Table 9

Coding “absent” within a separate neomorphic character statement (I) paired with a transformational character statement (II) versus within a single multistate character statement (III). Example coding for text character statements 23–25 and Figure 1 describing the presence and length of brow horns”

Taxon	Character states	Character scoring		
		Separate		Mixed
		I	II	III
X	Absent	0	?	0
Y	Present, subequal	1	0	1
Z	Present, longer	1	1	2

transformational characters separately (character statements 23, 24) results in an unambiguous synapomorphy (origin of brow horns) (Fig. 1A). The relative length of the brow horns has an ambiguous distribution at the internal node, because there are no shared character states. Coding this data as a single multistate character (character statement 25) yields no synapomorphies if left unordered (Fig. 1B). Thus, even though two of the taxa bear brow horns (Taxon Y, Z), no synapomorphy is recorded. This undesirable result can be partially remedied by ordering the character states (Fig. 1C), which then yields two unambiguous character state transformations (Fig. 1C). The first transformation records the appearance of a relatively short horn and later the elongation of that horn in Taxon Z, a result that also obtains with binary additive coding. But what if sequential ordering of the derived character states 1 and 2 is considered undesirable or unknowable? There is no way to duplicate the results of the logical separation of neomorphic and transformational character statements that allow, in this example, unambiguous recognition of the origin of a structure separate from a consideration of its transformation. The view taken here is that character statement 25 is a logical chimera; ordering alone cannot effectively separate neomorphic and transformational components.

Second, “absent” as a derived character state (i.e., as “loss”) is considered (Fig. 1D–F). In this case, presence of brow horns of some length is regarded as the primitive condition. Coding neomorphic and transformational characters separately (character statements 23, 24) results in an unambiguous autapomorphy for Taxon X that lacks a horn (Fig. 1D). A similar result is obtained when coding these data as single multistate characters (character statement 25; Fig. 1E). However, if this multistate character was ordered to partially remedy ambiguity during gain of a horn (Fig. 1C), then an unambiguous synapomorphy is present, one that links short-horned Taxon Y and hornless Taxon X (Fig. 1F). Again, if one regarded as unknown the specific antecedent horned condition (short- or long-horned) of the hornless Taxon X, ordering would need to be eliminated during size

reduction or loss. An “easy loss” step matrix for character state transformation would mimic the results of coding neomorphic and transformational characters separately.

### Rejection criteria for character statements

Poe and Wiens (2000) discussed “character selection criteria” that authors have used to include or reject character data. There seems little sense, nevertheless, to consider *positive* “selection criteria”, when *negative* rejection of data is the heart of the matter. If an author states that characters with low variability were included, one is left to infer that characters with high variability were excluded. Data partition analysis aside, there is no logical reason to exclude completely compatible, available data of relevant type (e.g., morphological); one needs justification to do so. Building a data matrix, in theory at least, should not amount to positively “selecting” agreeable character data. “Rejection criteria” for character data, then, are all that need be considered, and these are enumerated below.

All current criteria discussed in recent reviews (Poe and Wiens, 2000; Rieppel and Kearney, 2002) involve *comparative* aspects of character data that are intimately linked to the morphological nature of the character itself. I argue here that two other categories of rejection criteria exist, logical and operational. These relate to the content of this paper, the logical and operational aspects of character statements, which are usually brought to bear on potential character data by phylogeneticists before deeper-seated comparative considerations (Table 10).

#### Comparative

Poe and Wiens (2000, p. 33) identified: (1) high variation, (2) high homoplasy, (3) substantial missing

Table 10  
Rejection criteria for morphological characters

Category	No.	Criteria
Comparative	1	High variation
	2	High homoplasy
	3	Substantial missing data
	4	Unknown polarity
	5	Continuous variation
	6	Doubtful topology
	7	Doubtful connection
	8	Absence of special features
	9	Absence of intermediate forms
Logical	10	Character correlation
	11	Mixed character statements
Operational	12	Ambiguous or imprecise descriptions
	13	Incomplete quantitative-relative character statements

data, (4) unknown polarity and (5) continuous variation as commonly used, if poorly justified, rejection criteria. Likewise, Rieppel and Kearney (2002, p. 59) introduced other rejection criteria related to classical establishment of “primary homology”, such as (6) topology, (7) connection, (8) special similarity and (9) intermediate forms. Failure to establish characters and character states that have a solid or convincing comparative grounding (6–9) or that avoid what many regard as undesirable qualities (1–5) may result in rejection of character data by subsequent phylogeneticists.

### *Logical*

Rejection of character data on the basis of logical considerations involve perceived violations of precepts I and II, namely, character interdependence and the overlap/conjunction of character states (Table 10, numbers 10, 11). (10) Character correlation, for example, violates character independence and is a common reason for rejection of character data. (11) Character statements that intermix neomorphic and transformational states (i.e., “multistate” coding) also violate precepts I and II as discussed above. These are logical shortcomings, distinct from the comparative considerations above, and may result in rejection or at least reorganization of character data by subsequent phylogeneticists.

### *Operational*

*Operational* shortcomings include difficulty and/or inability to discern and/or score character statements (Table 10, numbers 12, 13). This is a very common reason for rejection of character data and can arise in at least two ways. Many operational deficiencies originate in (12) ambiguous or imprecise descriptions. Qualitative or quantitative character states may not clearly distinguish the relevant condition or may apply only to a particular vantage point that is not specified. For example, “sigmoid” versus “inturned” or “prominent” versus “rounded” as character states may not be informative enough for verification by another author. At issue here is the inability of one author to interpret the character statements of another as the reason for character statement rejection. This does not concern perceived variability of specimens or taxa (criteria 1–5), disagreements over “primary homology” (criteria 6–9), logical shortcomings (criteria 10, 11), or even differing interpretations of the same morphology among specialists considering the same taxa (Hawkins et al., 1997; Swiderski et al., 1998; Hawkins, 2000; Stevens, 2000).

(13) Incomplete quantitative-relative character statements (Table 8, example 2), commonplace in character data, are often rejected on the grounds that the

relationship is not apparent. Most quantitative-relative characters are relative measurements or ratios. Omission of a variable qualifier often eliminates the possibility of effective evaluation of the character states. Character statement 25, for example, involves relative horn length (short, long) but, unlike character statement 24, does not specify the relative relationship. What constitutes “short”? What if a taxon is found with an intermediate condition? Relative character states, simply put, are difficult to evaluate without specification of the relative relationship. The usual critique of ratios as character data is of secondary importance (i.e., uncertainty as to which feature is changing, failure to account for allometric scaling; Atchley et al., 1976). In the description of a character statement, such as number 24, there are often multiple additional comparisons (e.g., snout height, skull length) that confirm the locus of relative size change. The problem here involves the interpretation or testability of an incomplete quantitative-relative character statement.

### **Conclusions**

When Farris et al. (1970) ushered in the era of quantitative cladistics, I have argued that characters were properly viewed as independent variables composed of mutually exclusive conditions (Farris, 1983; Felsenstein, 1983). This has not been fully articulated or implemented by morphology-based cladists, who thus have not widely appreciated the following points: (1) character statements can be divided into character and statement, the former defined as an organismal feature expressed as an independent variable and the latter as its variable conditions; (2) character statements are composed of four logical components (locator, variable, variable qualifier, character state) that compose two fundamental patterns, neomorphic and transformational; (3) there are a small handful of “production rules” that compose a “generative grammar” for morphology-based character statements, the logical components of which may be visually diagrammed as a branching character-statement tree; (4) construction of complete characters that follow a few basic heuristic patterns (secondary locator(s) first, variable and variable qualifier last) stands to substantially reduce variation in morphology-based character data and enhance comparison between analyses; and (5) the exclusion of “absent” as a possible character state in characters that are truly transformational clarifies longstanding controversy over how to score an inapplicable state. Contingent coding, which uses a “?” for taxa lacking the structure of interest, is recommended for transformational characters that have neomorphic counterparts. The character state “absent” is logically at home only within neomorphic character statements.

## Acknowledgments

I thank Carol Abraczinskas for execution of Fig. 1. I thank Kurt Pickett and Steve Brusatte for comments on an earlier draft of the paper and especially Jeffrey Wilson and Yann Bertrand for detailed critiques and suggestions.

## References

- Atchley, W.R., Gaskins, C.T., Anderson, D., 1976. Statistical properties of ratios. I. Empirical results. *Syst. Zool.* 25, 137–148.
- Baum, B.R., 1988. A simple procedure for establishing discrete characters from measurement data applicable to cladistics. *Taxon* 37, 63–70.
- de Beer, G.R., 1971. Homology, an unsolved problem. *Oxford Biol. Readers* 11, 1–16.
- Bock, W.J., 1974. Philosophical foundations of classical evolutionary classification. *Syst. Zool.* 22, 375–392.
- Bolker, J.A., Raff, R.A., 1996. Developmental genetics and traditional homology. *Bioessays* 18, 489–494.
- Bonde, N., 1977. Cladistic classification as applied to vertebrates. In: Hecht, M.K., Good, P.C., Hecht, B.M. (Eds.), *Major Patterns in Vertebrate Evolution*. Plenum Press, New York, pp. 741–804.
- 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*. Special Volume Series 58. Systematics Association, London, pp. 10–21.
- Brower, A.V.Z., Schawaroch, V., 1996. Three steps of homology assessment. *Cladistics* 12, 265–272.
- Cain, A.J., Harrison, G.A., 1958. An analysis of the taxonomists' judgement of affinity. *Proc. Zool. Soc. London* 131, 85–98.
- Chomsky, N., 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.
- Colless, D.H., 1967. An examination of certain concepts in phenetic taxonomy. *Syst. Zool.* 16, 6–27.
- Colless, D.H., 1985. On "character" and related terms. *Syst. Zool.* 34, 229–233.
- Cracraft, J., 1978. Science, philosophy, and systematics. *Syst. Zool.* 27, 213–216.
- Crovello, T.J., 1970. Analysis of character variation in ecology and systematics. *Annu. Rev. Ecol. Syst.* 1, 55–98.
- Diederich, J., 1997. Basic properties of biological databases: character development and support. *Math. Comput. Modelling* 25, 109–127.
- Dodson, P., 1978. On the use of ratios in growth studies. *Syst. Zool.* 27, 62–67.
- Eble, G.J., 2000. Theoretical morphology: state of the art. *Paleobiology* 26, 520–528.
- Farris, J.S., 1970. Methods for computing Wagner Trees. *Syst. Zool.* 19, 83–92.
- Farris, J.S., Kluge, A.G., Eckardt, M.J., 1970. A numerical approach to phylogenetic systematics. *Syst. Zool.* 19, 172–191.
- Farris, J.S., 1983. The logical basis of phylogenetic analysis. In: Platnick, N.I., Funk, V.A. (Eds.), *Advances in Cladistics*. Columbia University Press, New York, pp. 7–36.
- Farris, J.S., Källersjö, M., Albert, V.A., Allard, M., Anderberg, A., Bowditch, B., Bult, C., Carpenter, J.M., Crowe, T.M., De Laet, J., Fitzhugh, K., Frost, D., Goloboff, P., Humphries, C.J., Jondelius, U., Judd, D., Karis, P.O., Lipscomb, D., Luckow, M., Mindell, D., Muona, J., Nixon, K., Presch, W., Seberg, O., Siddal, M.E., Struwe, L., Tehler, A., Wenzel, J., Wheeler, Q., Wheeler, W., 1995. Explanation. *Cladistics* 11, 211–218.
- Felsenstein, J., 1983. Parsimony in systematics: biological and statistical issues. *Annu. Rev. Ecol. Syst.* 14, 313–333.
- Fistrup, K., 1992. Character: current usages. In: Keller, E.F., Lloyd, E.A. (Eds.), *Keywords in Evolutionary Biology*. Harvard University Press, Cambridge, MA, pp. 45–51.
- Fistrup, K., 2001. A history of character concepts in evolutionary biology. In: Wagner, G.P. (Ed.), *The Character Concept in Evolutionary Biology*. Academic Press, San Diego, CA, pp. 13–35.
- Forey, P.L., Kitching, I.J., 2000. Experiments in coding multistate characters. In: Scotland, R., Pennington, R.T. (Eds.), *Homology and Systematics*. Special Volume Series 58. Systematics Association, London, pp. 54–80.
- Freudenstein, J.V., 2005. Characters, states, and homology. *Syst. Biol.* 54, 965–973.
- Ghiselin, M.T., 1984. "Definition", "character", and other equivocal terms. *Syst. Zool.* 33, 104–110.
- Grande, L., Rieppel, O. (Eds.), 1994. *Interpreting the Hierarchy of Nature: From Systematic Patterns to Process Theories*. Academic Press, San Diego, CA.
- Grant, T., Kluge, A.G., 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. *Cladistics* 19, 379–418.
- Grant, T., Kluge, A.G., 2004. Transformation series as an ideographic character concept. *Cladistics* 20, 23–31.
- Hall, B.K., 1994a. Introduction. In: Hall, B.K., (Ed.), *Homology: The Hierarchical Basis of Comparative Biology*. Academic Press, San Diego, CA, pp. 1–19.
- Hall, B.K. (Ed.), 1994b. *Homology: The Hierarchical Basis of Comparative Biology*. Academic Press, San Diego, CA.
- Harvey, P.H., Pagel, M.D., 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.
- 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*. Special Volume Series 58. Systematics Association, London, pp. 22–53.
- Hawkins, J.A., Hughes, C.E., Scotland, P.W., 1997. Primary homology assessment, characters and character states. *Cladistics* 13, 275–283.
- 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.
- Hecht, M.K., Edwards, J.L., 1977. The methodology of phylogenetic inference above the species level. In: Hecht, M.K., Good, P.C., Hecht, B.M. (Eds.), *Major Patterns in Vertebrate Evolution*. Plenum Press, New York, pp. 3–51.
- Hennig, W., 1966. *Phylogenetic Systematics*. University of Illinois Press, Urbana, IL.
- Kitching, I., Williams, D., Forey, P.L., Humphries, C.J., 1998. *Cladistics: the Theory and Practice of Parsimony Analysis*. Oxford University Press, Oxford.
- 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.
- Lauder, G.V., 1994. Homology, form, and function. In: Hall, B.K. (Ed.), *Homology: The Hierarchical Basis of Comparative Biology*. Academic Press, San Diego, CA, pp. 152–196.
- Lee, M.S.Y., 1998. Convergent evolution and character correlation in burrowing reptiles: towards a resolution of squamate relationships. *Zool. J. Linn. Soc.* 65, 369–543.
- Lee, M.S.Y., Caldwell, M.W., 2000. *Adriosaurus* and the affinities of mosasaurs, dolichosaurs, and snakes. *J. Paleontol.* 74, 915–937.
- Li, J.-L., Li, M.-X., Deng, H.-Y., Duffy P.E., Deng, H.-W., 2005. PhD: web database application for phenotype data management. *Bioinform.* 21, 3443–3444.
- Liem, K.F., Bemis, W.E., Walker, W.F., Grande, L., 2001. *Functional Anatomy of the Vertebrates*. Thomson Learning, Inc., Belmont.

- Little, W., Fowler, H.W., Coulson, J., 1955. *The Oxford Universal Dictionary: On Historical Principles*. Oxford University Press, Oxford.
- Mabee, P.M., 1989. Assumptions underlying the use of ontogenetic sequences for determining character state order. *Trans. Am. Fish. Soc.* 118, 151–158.
- Mabee, P.M., 2000. The usefulness of ontogeny in interpreting morphological characters. In: Wiens, J.J. (Ed.), *Phylogenetic Analysis of Morphological Data*. Smithsonian Institution Press, Washington, DC, pp. 84–114.
- Mabee, P.M., Ashburner, M., Cronk, Q., Gkoutos, G.V., Haendel, M., Mungall, C., Segerdell, E., Westerfield, M., in press. Phenotype ontologies: the bridge between genomics and evolution. *TRENDS in Ecology and Evolution*.
- Mabee, P.M., Humphries, J., 1993. Coding polymorphic data: Examples from allozymes and ontogeny. *Syst. Biol.* 42, 166–181.
- Maddison, W., 1993. Missing data versus missing characters in phylogenetic analysis. *Syst. Biol.* 42, 576–581.
- Maddison, W.P., Maddison, D.R., 1992. *Macclade: Analysis of Phylogeny and Character Evolution, Version 3*. Sinauer Associates, Inc., Sunderland, MA.
- Mayr, E., Linsley, E.G., Usinger, R.L., 1953. *Methods and Principles of Systematic Zoology*. McGraw-Hill Book Co., Inc., New York.
- McGhee, G.R. Jr, 1999. *Theoretical Morphology: The Concept and its Applications*. Columbia University Press, New York.
- Michener, C.D., Sokal, R.R., 1957. A quantitative approach to a problem in classification. *Evolution* 11, 130–162.
- Miyamoto, M.M., Cracraft, J. (Eds.), 1991. *Phylogenetic Analyses and DNA Sequences*. Oxford University Press, New York.
- Müller, G.B., Wagner, G.P., 1991. Novelty in evolution: restructuring the concept. *Annu. Rev. Ecol. Syst.* 22, 229–256.
- Nelson, G., 1978. Ontogeny, phylogeny, paleontology, and the biogenetic law. *Syst. Zool.* 27, 324–345.
- Nelson, G.J., Platnick, N.I., 1981. *Systematics and Biogeography*. Columbia University Press, New York.
- O'Grady, R.T., Deets, G.B., 1987. Coding multistate characters, with special reference to the use of parasites as characters of their hosts. *Syst. Zool.* 36, 268–279.
- Owen, R., 1843. *Lectures on Comparative Anatomy and Physiology of the Invertebrate Animals*. Delivered at the Royal College of Surgeons, in 1843. Longman, Brown, Green, and Longmans, London.
- Patterson, C., 1982. Morphological characters and homology. In: Joysey, K.A., Friday, A.E. (Eds.), *Problems of Phylogenetic Reconstruction*. Academic Press, London, pp. 21–74.
- Paterson, T., Kennedy, J.B., Pullan, M.R., Cannon, A., Armstrong, K., Watson, M.F.A., Raguenaud, C., McDonald, S.M., Russell, G., 1997. A universal character model and ontology of defined terms for taxonomic description. In: Rahm, E. (Ed.), *Data Integration in the Life Sciences*, Springer-Verlag, Berlin, pp. 63–78.
- Pimentel, R.A., Riggins, R., 1987. The nature of cladistic data. *Cladistics* 3, 201–209.
- de Pinna, M.C.C., 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7, 367–394.
- Platnick, N.I., 1979. Philosophy and the transformation of cladistics. *Syst. Zool.* 28, 537–546.
- Pleijel, F., 1995. On character coding for phylogeny reconstruction. *Cladistics* 11, 309–315.
- 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., Micevich, M.F., 1990. Character definitions and character state delineation: the bête noir of phylogenetic inference. *Cladistics* 6, 319–361.
- Richards, R., 2003. Character individuation in phylogenetic inference. *Philos. Sci.* 70, 264–279.
- Rieppel, O., 1994. Homology, topology, and typology: the history of modern debates. In: Hall, B.K. (Ed.), *Homology: The Hierarchical Basis of Comparative Biology*. Academic Press, San Diego, CA, pp. 63–100.
- Rieppel, O., Kearney, M., 2002. Similarity. *Biol. J. Linn. Soc.* 75, 59–82.
- Rodrigues, P.D., 1986. On the term character. *Syst. Zool.* 35, 140–141.
- Sattler, R., 1984. Homology—a continuing challenge. *Syst. Bot.* 9, 382–394.
- Scotland, R., Pennington, R.T. (Eds.), 2000. *Homology and Systematics*. Special Volume Series 58. Systematics Association, London.
- Slowinski, J.B., 1993. “Unordered” versus “ordered” characters. *Syst. Biol.* 42, 155–165.
- Sneath, P.H.A., Sokal, R.R., 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. W.H. Freeman, San Francisco, CA.
- Sokal, R.R., Sneath, P.H.A., 1963. *Numerical Taxonomy*. W.H. Freeman, San Francisco, CA.
- 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*. Special Volume Series 58. Systematics Association, London, pp. 81–105.
- Stuessy, T.F., 1990. *Plant Taxonomy: The Systematic Evaluation of Comparative Data*. Columbia University Press, New York.
- Swiderski, D.L., Zelditch, M.L., Fink, W.L., 1998. Why morphometrics isn't special: coding quantitative data for phylogenetics. *Syst. Biol.* 47, 508–519.
- Swofford, D.L., Maddison, W.P., 1992. Parsimony, character-state reconstructions, and evolutionary inferences. In: Mayden, R.L. (Ed.), *Systematics, Historical Ecology and North American Freshwater Fishes*. Stanford University Press, Stanford, pp. 187–223.
- Thiele, K., 1993. The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics* 9, 275–304.
- Wagner, G.P., 1989. The biological homology concept. *Annu. Rev. Ecol. Syst.* 20, 51–69.
- Wagner, G.P., 2001a. Characters, units and natural kinds: an introduction. In: Wagner, G.P. (Ed.), *The Character Concept in Evolutionary Biology*. Academic Press, San Diego, CA, pp. 1–12.
- Wagner, G.P., ed., 2001b. *The Character Concept in Evolutionary Biology*. Academic Press, San Diego, CA.
- Werdelin, L., Tullberg, B.S., 1995. A comparison of two methods to study correlated discrete characters on phylogenetic trees. *Cladistics* 11, 265–277.
- Wheeler, W.C., 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Syst. Biol.* 44, 321–331.
- Wiens, J.J., ed., 2000. *Phylogenetic Analysis of Morphological Data*. Smithsonian Institution Press, Washington, DC.
- Wiley, E.O., 1975. Karl R. Popper, systematics, and classification: a reply to Walter Bock and other evolutionary taxonomists. *Syst. Zool.* 24, 233–243.
- Wiley, E.O., 1981. *Phylogenetics: the Theory and Practice of Phylogenetic Systematics*. John Wiley & Sons, New York.
- Wilkinson, M., 1995. A comparison of two methods of character reconstruction. *Cladistics* 11, 297–308.
- Williams, D.M., Seibert, D.J., 2000. Characters: homology and three-item analysis. In: Scotland, R., Pennington, R.T., eds. *Homology and Systematics*. Special Volume Series 58. Systematics Association, London, pp. 183–208.
- Woodger, J.H., 1952. Science without properties. *Br. J. Philos. Sci.* 2, 193–216.
- Zelditch, M.L., Fink, W.L., Swiderski, D.L., 1995. Morphometrics, homology, and phylogenetics: quantified characters as synapomorphies. *Syst. Biol.* 44, 179–189.