Blackwell Publishing

Cladistics 23 (2007) 1–11

Cladistics

10.1111/j.1096-0031.2007.00168.x

Quantitative tests of primary homology

Ingi Agnarsson[1,2]* and Jonathan A. Coddington[1]

[1]*Department of Entomology, NHB-105, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA;* [2]*The University of British Columbia, Departments of Botany and Zoology, 3549-6270 University Blvd., Vancouver, BC V6T 1Z4, Canada*

Accepted 14 September 2006

Abstract

In systematic biology homology hypotheses are typically based on points of similarity and tested using congruence, of which the two stages have come to be distinguished as ''primary'' versus ''secondary'' homology. Primary homology is often regarded as prior to logical test, being a kind of background assumption or prior knowledge. Similarity can, however, be tested by more detailed studies that corroborate or weaken previous homology hypotheses before the test of congruence is applied. Indeed testing similarity is the only way to test the homology of characters, as congruence only tests their states. Traditional homology criteria include topology, special similarity, function, ontogeny and step-counting (for example, transformation in one step versus two via loss and gain). Here we present a method to compare quantitatively the ability of such criteria, and competing homology schema, to explain morphological observations. We apply the method to a classic and difficult problem in the homology of male spider genital sclerites. For this test case topology performed better than special similarity or function. Primary homologies founded on topology resulted in hypotheses that were globally more parsimonious than those based on other criteria, and therefore yielded a more coherent and congruent nomenclature of palpal sclerites in theridiid spiders than prior attempts. Finally, we question whether primary homology should be insulated as ''prior knowledge'' from the usual issues and demands that quantitative phylogenetic analyses pose, such as weighting and global versus local optima.

Homology—correspondence due to common ancestry—continues to be a vexing theoretical problem (Owen, 1843; Darwin, 1859; Remane, 1952; Hennig, 1966; Jardine, 1967; de Beer, 1971; Patterson, 1982; Farris, 1983; Pogue and Mickevich, 1990; de Pinna, 1991; Minelli and Schram, 1994; Nelson, 1994; Rieppel, 1994, 2001, 2006; Hall, 1995; Rieppel and Kearney, 2002; Brigandt, 2003; Kluge, 2003; Ghiselin, 2005; Hoßfeld and Olsson, 2005; Richter, 2005; Rutishauser and Moline, 2005; Scholtz, 2005; Wägele, 2005; Phillips, 2006). In phylogenetics, homology hypotheses are typically based on similarity and tested by congruence. Conventional phylogenetics accepts primary homology hypotheses, or characters (De Pinna, 1991) as given and only measures how well a tree accounts for the state variation encoded within each character. Rieppel and Kearney (2002) recently reviewed the topic of primary homology and reminded us that things can be similar to greater or lesser degrees, at different scales, and in different, arguably independent ways (see also Richter, 2005; Scholtz, 2005; Phillips, 2006; Rieppel, 2006). In theory, parsimony, or other criteria, could guide choice among competing homology hypotheses before their fit to a tree is assessed. Although testing the fit of data to trees using various optimality criteria is well-understood, how one would evaluate initial considerations of similarity during the formation of homology hypotheses, such as whether two structures should even be considered homologous, remains murky.

We follow De Pinna (1991) in distinguishing between ''primary'' and ''secondary'' homology: primary homology is the (theory-laden but untested) supposition that two parts are the same by inheritance, whereas secondary homologies have withstood the test of congruence, i.e., emerged as synapomorphies on a cladogram. We also follow Remane (1952) and most subsequent authors

*Corresponding author:
E-mail address: ingi@zoology.ubc.ca

in recognizing a variety of reasonably independent criteria (or tests) for primary homology: topology, special similarity, function and ontogeny (although ontogenetic sequence is akin to topology). Rieppel and Kearney (2002) suggested that closer and more detailed observation effectively tests primary homology (see also Richter, 2005; Scholtz, 2005; Rieppel, 2006; but see Kluge, 2003), and agree with many previous authors that similarity in structural detail, function and ontogeny are often seen as secondary criteria to topology.

Traditionally topology has been regarded as the most reliable criterion (e.g., Rieppel and Kearney, 2002; Richter, 2005). This amounts to an empirical claim that topology-driven homology schemes have been, and will continue to be, on average most successful in determining homology (Rieppel, 2006). It may be true. However, topology is not universally regarded as the "best" criterion (e.g., Scholtz, 2005) and assigning relative importance to homology criteria a priori is reminiscent of a priori character weighting. However well intentioned, skeptics usually want to see the unweighted solution, and in any case justifying the weights always burdens the analysis. For similar reasons, we see no particular need to anoint one criterion as better than another by giving it a larger role or greater weight, or primacy in the analysis.

Twenty years ago, many authors considered ontogeny as most important for deciphering homology (e.g., Nelson, 1978; Roth, 1984; Pearson et al., 1985; Larimer and Pease, 1990). Ontogenetic studies, however, apply other criteria (topology, correspondence) to various stages of the non-adult form; ontogeny is not a separate criterion, but extends our ability to apply other criteria. For example, ontogeny may prove the identity of two structures whose topological and particular similarities are obscured in the adult phenotype. Presumably, as our understanding of development matures, topology will become less pattern-based and more of a causal concept. The effectiveness of homology criteria may vary depending on the biological domain of the problem. In behavior, the analog to topology is arguably sequence of motor patterns, and organisms seem to scramble sequence more readily than change behavioral units as defined by special similarity (e.g., spider web construction, Eberhard, 1982; Coddington, 1986).

The problem of primary homology is not confined to morphological data. Aligning base positions in "homologous" sequences across different taxa is a crucial phase in molecular analysis (e.g., Simmons and Freudenstein, 2003; Wheeler, 2003; Phillips, 2006). As in morphological characters, similarity guides primary homology hypotheses of sequence data (Simmons and Freudenstein, 2003). Improvements in sequence alignment continue to appear (e.g., De Laet, 2005), but one method in particular, "Direct Optimization" (Wheeler, 1996, 1999, 2003, 2006), seeks optimal trees via direct optimization of the data, hence varying dynamically the primary homology hypotheses during tree search. Each resulting tree then has a different, implicit alignment, which can be calculated more or less precisely a posteriori. In morphology, dynamic homology analyses—simultaneously calculating costs of alternative homology schema and estimating tree topology—could also be used to improve propositions of homology (Ramírez, in press). Our approach, in contrast, attempts to use character complexity as information, and asks if *criteria* to judge homology can be evaluated *prior* to the tree search. This seems useful when characters are complex and alternative homology criteria may be in conflict. However, it is less important for DNA data as it is considerably simpler than morphology in that character state set is small and finite (A, C, G, T), so that topology becomes almost exclusively important.

Despite basic theoretical agreement about primary homology hypotheses and how to judge them, justifying arguments for morphological homologies remain overwhelmingly verbal rather than quantitative. Although cladistics tests congruence among character states, the absence of quantitative tests for the homology of the characters themselves is a serious issue. "Testing" primary homology still resembles precladistic systematics—more or less tendentious essays about which criteria are more important, gains or losses of features, and more or less lengthy expositions on morphological variation, position and function. For example, the identity of avian digits remains unresolved despite more than 150 years of study (e.g., Wagner, 2005). However, it is certainly possible to assess how well a given homology hypothesis applied to a set of taxa performs under the criteria of topology, similarity, function, ontogeny and even others. Even if that performance cannot usually be literally "measured", a given criterion may either, support, reject or be inapplicable to the hypothesis under consideration. Given a set of taxa whose morphology varies and alternative theories to explain that variation, the "fit" of those theories in each taxon under each criterion can be methodically examined and tallied in order to compare alternative primary homology hypotheses in a reasonably objective and reproducible way.

Here we propose such a method to measure whether topology, function or similarity, perform best in a didactic example and an empirical case: the homologies of male palpal genital sclerites in theridiid spiders. The method tallies the extent to which a given homology scheme preserves topology, special similarity and function (because ontogenetic data are very sparse, we discuss only the latter three criteria) with respect to a reference such as an outgroup or ground plan.

Tests of primary homology require a starting point from which the "cost" of the competing hypothesis can be assessed. Because the test is applied prior to

phylogenetic reconstruction, costs of alternative hypotheses cannot be estimated on a tree (or across competing trees), as is done, e.g., in the direct optimization method for DNA. Instead, costs are calculated based on agreement/disagreement with a reference taxon. For morphology, it makes sense to pick a ground plan as the starting point: changes from it to explain the observed diversity in form then constitute the cost or goodness of fit of the particular homology schema being considered. The reference could also be identified by outgroup comparison, or character optimization on a prior cladogram, or even be any real taxon in the analysis. The choice of reference is unlikely to affect methodologically the computation of costs, just as re-rooting trees does not change tree length. The reference scheme could be the optimized root character vector for the relevant features on the correct phylogeny, but, of course, the latter is not at hand, so in the worse case, stable homology hypotheses (as measured here) might take several rounds of adjustment. Furthermore, heuristically it helps if the reference has as many of the characters being tested as possible.

As we explain in more detail below, we propose to score a given homology hypothesis (e.g., "A" transforms to "B", with its implied changes) as either according (1), conflicting (0), or inapplicable (*) with respect to a given homology criterion, such as topology, special similarity or function for a series of taxa. The results of such scorings are then averaged so that the fit of hypotheses to criteria can be fairly compared. Alternatively the conflicts can be summed by step-counting (counting each incongruence as one step, as in a standard parsimony analysis of character states), where the most parsimonious hypotheses set is favored. This extends the idea that complexity (multiple independent comparisons) tests homology (Patterson, 1982).

For a test case we use a recent comparative study of the spider family Theridiidae (Fig. 1, see also Agnarsson et al., 2007, figs 4–41), whose male genitalia are difficult to homologize. Male genitalia have long been the most important character system for species identification and a rich source of characters for phylogenetic analyses (e.g., Griswold et al., 2005). As one might expect when a system that evolves fast enough to mark species is also used to establish large clades, the nomenclature of male palps has been contentious (Coddington, 1990). Although merely inconvenient for species identifications, it can wreak havoc in phylogenetic inference, where the names of parts usually guide primary homology hypotheses. Palpal sclerites seem to be among the most homoplasious of characters. Scharff and Coddington (1997, p. 422, fig. 103) found male genitalia on average, to be among the least reliable (in the sense of having highest homoplasy) characters in Araneidae compared with eyes, abdomen, female genitalia, spinnerets and behavior. The same pattern occurs in Araneoidea and Entelegynae (Griswold et al., 1998, 1999, 2005). Certain sclerites seem to perform especially poorly in almost all studies. For example, the consistency and retention indices of the median apophysis (MA) in orbicularians is much lower than average in eight of nine recent phylogenetic studies (Table 1). Although some of this homoplasy is certainly real—on theoretical grounds male genitalia may evolve faster than somatic morphology (e.g., Eberhard, 1985; but see Huber, 2003)—some may also be due to mistaken homologies. Unstable and complicated homology schemes for male sclerites have hindered phylogenetic analysis in theridiids (Levi and Levi, 1962; Saaristo, 1978; Coddington, 1990; Agnarsson, 2003, 2004, 2006a,b) and other spiders (Griswold et al., 1998). Perhaps because of this instability, recent taxonomic papers on theridiids avoid classic palpal sclerite names. Knoflach (e.g., 1991, 1992, 1993a,b) and Knoflach and van Harten (2000, 2001) explicitly chose names such as tegular apophysis I, II and III to avoid interfamilial homologies. We use a recent comparative study of theridiid palps (Agnarsson, 2004; Agnarsson et al., 2007) to test the method. We furthermore evaluated prior homology schema for theridiid palps (Levi and Levi, 1962; Saaristo, 1978; Coddington, 1990) insofar as we could reconstruct the various criteria those authors articulated for their decisions.

Method and results

Quantifying primary homology criteria

Figure 2 presents a simple example to illustrate the logic. Which "sclerites" in taxon A are homologous to which sclerites in the reference taxon? Collectively the three sclerites perform three different functions (F1–F3), but function and topology are discordant. Differences in special similarity are indicated by black or white. If function were the preferred primary homology criterion, then *r1* would be homologous to *a2*, *r2* to *a1*, and *r3* to *a3*. If sclerite *r2* was the "radix", then characters such as "position of radix: ectal (0); mesal (1)", and "radix color: (0) white; (1) black" are implied. If topology were preferred, then $r1 = a1$, $r2 = a2$, and $r3 = a3$. In this alternative, the previous two characters would compare states among non-homologous sclerites. In this case the character might instead state "radix function: (0) F1; (1) F2." The first primary homology hypothesis causes conflict in two other criteria, but the second causes conflict in only one. We propose to treat each homology criterion as primary, to compare implied homologs on each point of comparison, and to score that comparison as agreeing, different, or inapplicable (1, 0 or "*", respectively). In Fig. 2 with topology as the primary criterion, $r1 = a1$ agrees in similarity but differs in function, $r2 = a2$ also agrees in similarity and conflicts

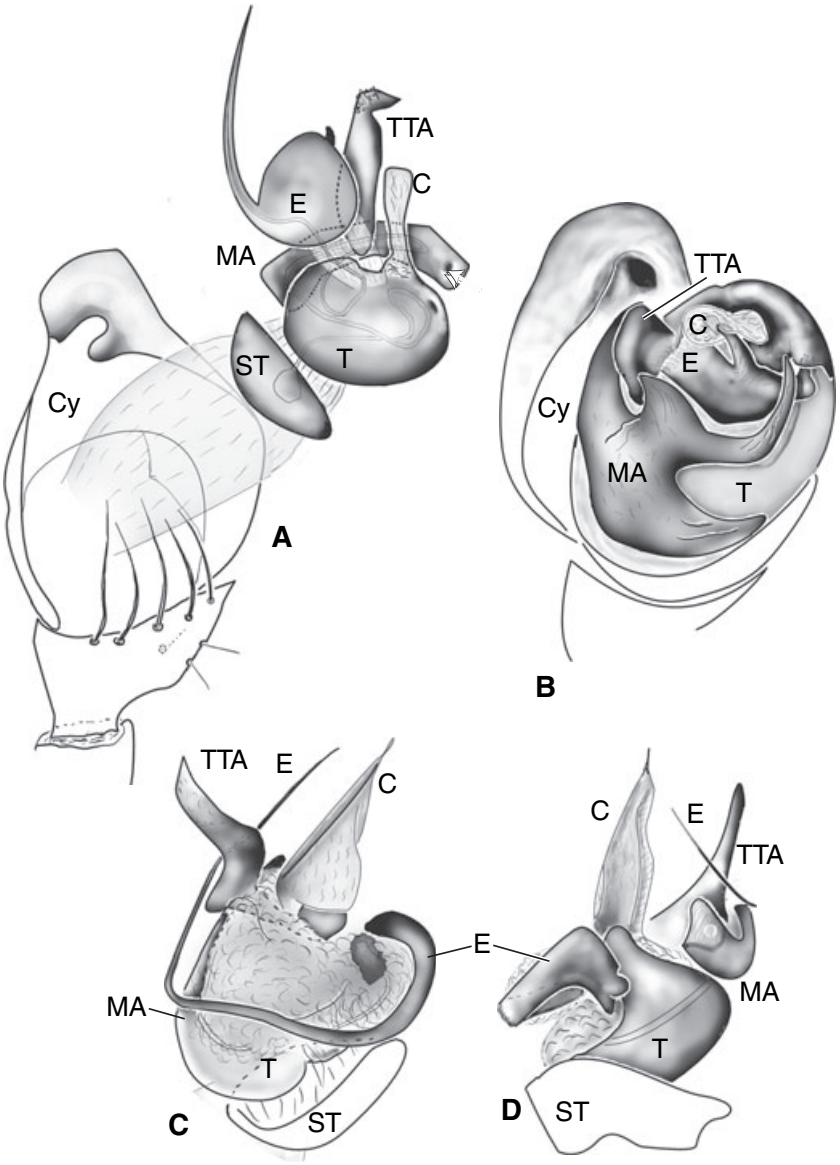

Fig. 1. Theridiid male pedipalpal bulbs. (A) Artificial reference groundplan, bulb (clasp of sclerites on the right) expanded from the cymbium (Cy); (B) *Theridion cochise* entire palp; (C,D) *Steatoda americana* bulb detached from cymbium; (C) ventral, (D) dorsal. While for some parts, e.g., the subtegulum (ST), tegulum (T), and embolus (E) homology is unproblematic—or at least uncontroversial—the homology of others, the median apophysis (MA), theridiid tegular apophysis (TTA) and conductor (C) is a difficult problem. Implied homologies are based on the current analysis.

in function, and $r3 = a3$ agrees in function but not similarity, for a total of three conflicts. If function is primary, both $r1 = a2$ and $r2 = a1$ create conflicts in topology and similarity, and $r3 = a3$ conflicts in similarity, for a total of five conflicts. Topology is preferred as the primary homology criterion because it minimizes conflict, or in other words, maximizes congruence.

Authors sometimes minimize conflicts in data by positing gains and losses of parts as necessary: taxon A has no *r1* and therefore *a1* is a novel sclerite. The reasoning is not unlike proposing gaps to explain indel events in molecular alignments. Although loss/gain hypotheses are certainly more plausible than drastic transformation in some cases, they should entail some cost, as do gap-opening or gap-extension penalties in alignment. A moderate view might be that if the *preponderance* of comparisons support homology, no gain/loss scenario should be invoked, and differences are scored as above. A *strict* view might be that if two potential homologs differ at all under a given criterion, then homology is rejected and gain/loss is invoked.

Figure 3 is more complex and more realistic. It encompasses the two views of gain/loss scenarios mentioned above, and also addresses ties in tallies.

Table 1
Consistency index (CI), retention index (RI) of the median apophysis (present or absent) in recent studies of orbicularian spiders, compared with the ensemble CI (ECI) and RI (ERI) of all characters in each study

Study	Focus taxon	CI	RI	ECI	ERI
Griswold et al. (1998)	Orbiculariae	14	44	64	81
Scharff and Coddington (1997)	Araneidae	16	50	33	74
Hormiga (1994b)	Linyphiidae	33	33	73	81
Hormiga et al. (1995)	Tetragnathidae	50	83	56	72
Hormiga (2000)	Erigoninae	50	67	41	68
Coddington (1990)	Orbiculariae	12	50	72	88
Hormiga (1994a)	Pimoidae	25	nc	71	87
Agnarsson (2004)	Theridiidae	25	40	37	73
Miller (2007)	Erigoninae	50	66	23	58
Average		30.6	54.1	52.2	75.8

Not calculated = nc.

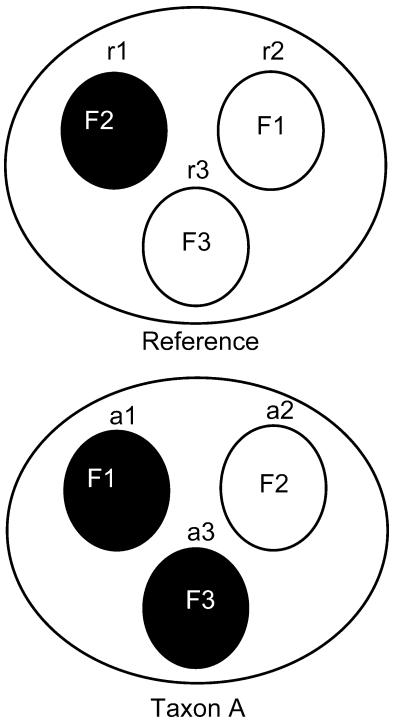

Fig. 2. Two taxa, a reference and A, each have three "sclerites", which have functions F1–F3 and different morphological attributes, black or white.

Table 2a represents the "preponderance" view, and 2b the "strict" view. Three taxa, the reference (e.g., a ground plan or outgroup), A and B, each have three sclerites named *r1–3*, *a1–3* and *b1–3*. Topology, three details of special similarity (white or black color, round or hexagonal shape, and "duct" or no), and function (F1–F4) differ in various ways.

Taking the preponderance view first (Table 2a), and setting topology as the first primary criterion, proposed homologies are $r1 = a1 = b1$, $r2 = a2 = b2$ and $r3 = a3 = b3$ for the following reasons. Considering just sclerite *r1* and its homologs, taxon A is uncontroversial because an "identical" sclerite is present in the same place, but *b1* has a duct, so the "duct" cell for sclerite 1, taxon B in Table 2a receives a 0, and all other cells receive 1, indicating congruence (indicated by "–" for visual clarity). Table 2 weights topology, special similarity and function equally (but see below), so the column "SS" averages the scores for each detailed comparison, resulting in a score of 1.0 for taxon A (no penalty) and 0.7 for taxon B. For sclerite $r2 = a2 = b2$, *a2* is black while *r2* is white (column "*col*"), and *b2* is round and lacks a duct compared with *r2* (columns "*dct*" and "*shp*"), giving average SS scores of 0.7 and 0.3 for taxa A and B, respectively. Also, *a2* has function F4 whereas *r2* has F2, thus scoring 0 in the FCN column for sclerite 2 in Taxon A in Table 2a. For $r3 = a3 = b3$, the only conflict is that *a3* is white.

If special similarity is primary, issues of tied scores arise. Considering *r1*, *a1* and *a3* are identical to it on the basis of special similarity alone (duct, color and shape), but choosing *a1* as a homolog does less violence to the topology criterion and therefore is preferred. Sclerite *r2* is most similar to *a2* (differing only in color) and *b1* (differing only in shape). Although *r2* and *b1* also differ topologically, and thus might be considered to be "tied" with the differences between *r2* and *b2* (shape, duct), that logic uses a secondary criterion, topology, to make a tie where the primary criterion is decisive. Sclerite *r2* differs only in shape from *b1*, whereas it differs in multiple ways from *b2* and *b3*, so the former homology is clearly preferable. Comparisons based on the primary criterion win, no matter the conflict imposed on other criteria, but ties in the primary criterion can be broken by considering secondary criteria. This reasoning exemplifies the "preponderance point of view because *r2* agrees with *a2* and *b1* in two of three similarity comparisons; gain/loss is not invoked.

If function is primary, issues of gain/loss arise. Sclerite $r1 = a1 = b1$, and differences are tallied as above, likewise for $r2 = a3 = b2$. Solely on the basis of function, however, taxon A has no sclerite performing

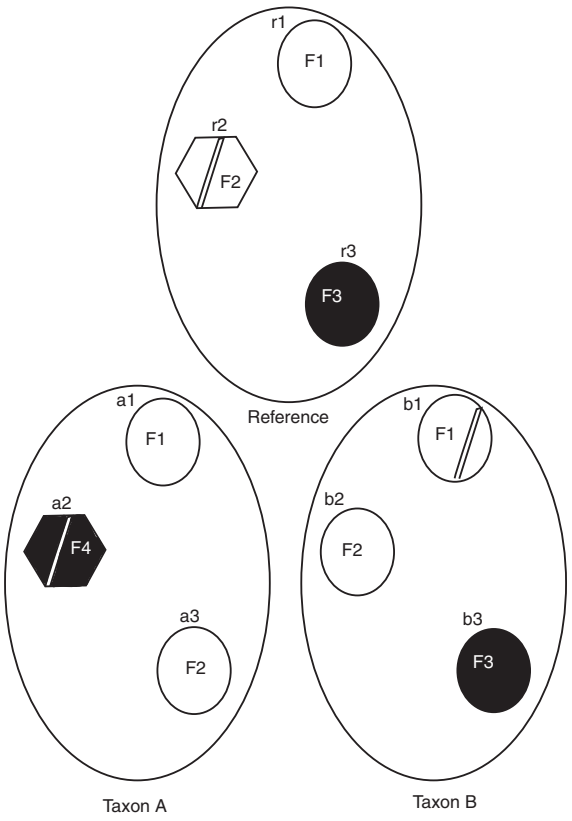

Fig. 3. Three taxa, a reference, A, and B, each have three sclerites that potentially in topology, detail (white or black, with or without a duct (parallel lines), round or hexagonal shape), and function (F1–F4).

F3, and so *r3* is presumed lost and *a2*, performing F4, gained. That hypothesis, however, entails the re-appearance of black color in *a2*. That similarity counts against the gain/loss scenario (thus 0 for *col* with FCN primary for sclerite 3 in Taxon A, Table 2a). On the other hand, differences between *a2* and *r3,* such as novel function F4, shape and presence of a duct, do not count against the gain/loss scenario because "new" structures are expected to have novel features. Because gain/loss is invoked, the G/L cell for sclerite 3 in Taxon A with function as the primary criterion in Table 2a scores 0.

The "Summary" matrix beneath Table 2a summarizes these comparisons across all criteria and sclerites. For the preponderance view, if topology is primary, topology scores 1.0, but special similarity and function suffer at 0.72, and 0.67, respectively, for a mean index value of 0.85. With special similarity and function as primary criteria, index means are 0.71 and 0.83. Topology performs best in this case.

The "strict" point of view holds that any differences under a particular criterion justify hypotheses of loss and gain (Table 2b). In this case, changes occur only if special similarity is primary, and only for sclerites 2 and 3. Taking taxon A first, even though *a2* differs from *r2* only in color, the strict view denies homology: *a2* is novel, *r2* lost, and 0 is entered in the G/L column. As potential falsifiers of the gain/loss scenario, any convergent similarities between *r2* and *a2* count against it, thus topology, duct presence and shape also score 0 for sclerite 2 in taxon A under the similarity criterion in Table 2b. Taxon B is similar; *b1* finds no exact match in the reference taxon and scores 0 in the G/L column. The two features that it shares convergently with *r2*, duct presence and color, count against the gain/loss hypothesis and score 0, but the shared features shape and topology do not. Note that the scoring of sclerite 3 with function F4 in taxon A is the same for the preponderance and strict points of views (Table 2). In effect, "strict" logic guided scoring in Table 2a; empirical cases may call for a mixture of the two views. Alternatively, one could argue that *a1* and *a3* found homologs in taxon A, that therefore *a2* as the sole unmatched sclerite in A is likely homologous to the unmatched *r3*, and that therefore function F4 is F3 transformed (with attendant changes in duct presence and shape), but that compromises the primacy of the functional criterion in that round of comparisons. Nevertheless, that scoring is possible. Philosophical purity is not the point, but rather to develop a method that fairly represents, yet systematizes and tests primary homology arguments.

In the above schemes, special similarity provided three points of comparison, but topology and function only one each, yet all were weighted equally. If differences in topological or functional congruence were evident, they would constitute additional points of comparison under their respective criteria. The number of such distinctions ultimately equals the number of distinct character states in a matrix. In conventional parsimony analysis, each state change counts as one step. Here the equivalent would be to count each difference as a step rather than to down-weight special similarity because criteria a priori are viewed a priori as coordinate or equivalent. In this case, special similarity has more points of comparison and happens to be more "complex" than the other criteria. Complexity is information, and it seems perverse to discriminate against it. The "Parsimony" columns in the Summary at the bottom of Table 2 give the results if all differences among sclerites in Fig. 3 are weighted equally (step-counting). Topology still performs best under either the preponderance or strict view by requiring just seven steps to explain the data, but weighting versus step-counting has an effect if a criterion has subcriteria, such as special similarity in this case. The "cost" of the strict point of view goes up if it has to account for several "coincidentally" similar features in a gain/loss scenario. This effect seems justified as these coincidences potentially falsify the gain/loss scenario and instead argue for transformation as an explanation of the difference in the primary homology criterion. Step-counting may favor transformational hypotheses in

Table 2
The scores for each sclerite (1–3) compared with the reference taxon in Fig. 3 for each pair-wise criterion comparison, with each criterion as primary in turn (sclerite homologies indicates in parentheses). For visual clarity, agreement between taxon and reference morphology is shown as a "–" but for calculations has a value of 1. SS averages the three columns (dct, col, shp). The six values (2 taxa × 3 sclerites) for each pair-wise comparison in the main table are averaged in the means matrix at bottom

	Sclerite 1							Sclerite 2							Sclerite 3						
	TOP	dct	col	shp	SS	FCN	G/L	TOP	dct	col	shp	SS	FCN	G/I	TOP	dct	col	shp	SS	FCN	G/L
(A) Preponderance																					
TOP Primary	(r1 = al = b1)							(r2 = a2 = b2)							(r3 = a3 = b3)						
Taxon A	–	–	–	–	1.0	–	–	–	–	0	–	0.7	0	–	–	–	–	–	1.0	0	–
Taxon B	–	0	–	–	0.7	–	–	–	0	–	0	0.3		–	–	–	0	–	0.7	–	–
SS Primary	(rl = al = b2)							(r2 = a2 = b1)							(r3 = a3 = b3)						
Taxon A	–	–	–	–	1.0	–	–	–	–	0	–	0.7	0	–	–	–	0	–	0.7	0	–
Taxon B	0	–	–	–	1.0	0	–	0	–	–	0	0.7	0	–	–	–	–	–	1.0	–	–
FCN Primary	(r1 = a1 = b1)							(r2 = a3 = b2)							(r3 = n/a = b3)						
Taxon A	–	–	–	–	1.0	–	–	0	0	–	0	0.3		–	–	–	0	–	0.7	–	0
Taxon B	–	0	–	–	0.7	–	–	–	0	–	0	0.3	–	–	–	–	–	–	1.0	–	–
(B) Strict																					
TOP Primary	(r1 = al = b1)							(r2 = a2 = b2)							(r3 = a3 = b3)						
Taxon A	–	–	–	–	1.0	–	–	–	–	0	–	0.7	0	–	–	–	–	–	1.0	0	–
Taxon B	–	0	–	–	0.7	–	–	–	0	–	0	0.3		–	–	–	0	–	0.7	–	–
SS Primary	(rl = al = b2)							r2 = n/a = n/a							(r3 = n/a = b3)						
Taxon A	–	–	–	–	1.0	–	–	0	0	–	0	0.3	0	0	–	0	0	–	0.3	–	0
Taxon B	0	–	–	–	1.0	0	–	–	0	0	–	0.3	–	0	–	–	–	–	1.0	–	–
FCN Primary	(r1 = a1 = b1)							(r2 = a3 = b2)							(r3 = n/a = b3)						
Taxon A	–	–	–	–	1.0	–	–	0	0	–	0	0.3	–	–	–	–	0	–	0.7	–	0
Taxon B	–	0	–	–	0.7	–	–	–	0	–	0	0.3	–	–	–	–	–	–	1.0	–	–

Summary	Preponderance						Strict					
Primary	TOP	SS	FCN	G/L	Index Means	Parsimony	TOP	SS	FCN	G/L	Index Means	Parsimony
TOP	1.00	0.72	0.67	1.00	0.85	7	1.00	0.72	0.67	1.00	0.85	7
SS	0.67	0.83	0.33	1.00	0.71	9	0.67	0.67	0.83	0.50	0.67	12
FNC	0.83	0.67	1.00	0.83	0.83	8	0.83	0.67	1.00	0.83	0.83	8

col = color, dct = duct, FCN = function, G/L = gain/loss, shp = shape, SS = special similarity, TOP = topology). "Parsimony" in the means matrix refers to step-counting; simply the number of zeros (conflict, homoplasy) in each column (see text for detail).

general. Below we discuss why indices and step-counting differ, and which might be preferred.

Theridiid palpal sclerites

Agnarsson (2004) analyzed theridiid phylogeny at the generic level using a matrix of 61 terminals (eight outgroup genera, 31 theridiid genera) and 242 characters, of which 88 (36%) pertained to the palpal organ. Based on these results, we chose 28 taxa (18 genera) from across the cladogram to exemplify theridiid palpal diversity, and to illustrate the method. If a set of taxa have identical palpal conformations with regards to the character under study, only one of the set needs to be included in a study such as this. A more detailed analysis including many more theridiid species is underway, but preliminary results are generally identical to those presented here (Agnarsson unpublished). For the complete list of material examined in this study see Agnarsson et al. (2007).

Orbiculariae (orb weaving spiders) commonly have three sclerites on the distal segment of the palpal bulb (tegulum) to which three names (embolus, conductor, median apophysis) have been applied. Some have a fourth, variously named. In Theridiidae, the basal condition is four (Agnarsson, 2004), but some taxa have five, and many others have three, two, or only one. Spider systematists consistently use four terms (conductor, embolus, median apophysis, radix) even while disavowing any implication of homology. In all palps the embolus contains the ejaculatory duct and conveys sperm to the female. It is the only sclerite whose homology is not controversial. Other homologies, such as the conductor, median apophysis, radix, theridiid tegular apophysis, paramedian apophysis, suprategulum, conductor II, etc., are problematic.

In theridiid palps the identity of the embolus is not controversial. Three sclerites remain in the reference ground plan (Fig. 1). To test their homology and the performance of homology criteria, we score a selection of

taxa illustrated in Agnarsson et al. (2007, figs 4–41) that could affect homology assignments. Topology is self-explanatory (Fig. 1). Special similarity includes three comparisons: presence/absence of ducts within sclerites (dct); fused or flexible attachment to the tegulum; and membranous or sclerotized sclerite texture (see also Agnarsson, 2004). Function includes two comparisons: whether the sclerite locks to the cymbium (Agnarsson, 2004); and whether it "conducts" the embolus. Conflict with other criteria occurs if the conducting sclerite differs topologically, or is membranous rather than sclerotized, or is flexibly attached rather than fused to the tegulum. The primary criterion therefore necessarily imposes lower scores on the secondary criteria, but every criterion has a chance to be primary.

This method can also evaluate different studies of the same homology problem. For example, Saaristo (1978) concluded that the presence of a duct inside theridiid sclerites outweighed any other evidence of homology. Hence he labeled the sclerite that functioned to lock the bulb to the cymbium as "locking apophysis A" if it contained a duct, but as "locking apophysis B" if not. He emphasized one special similarity, and introduced novel sclerites as necessary, regardless of topology or function. We compare Saaristo's hypothesis to that adopted here, as well as to that of Levi and Levi (1962) and Coddington (1990). Usually it was relatively straightforward to infer what primary criterion guided homology hypotheses, but in some cases the logic guiding choices in prior work will be difficult to understand in this context. Actual scores for sclerites of 28 genera would be tedious to present and meaningless without extensive figures, but Table 3 presents the summary results. The best overall indicator of the ability of a primary homology criterion to explain the data is its summary score. It is also illuminating to visualize graphically how well each criterion or homology scheme performed on each "axis" of comparison: topology, function and special similarity (Fig. 4). As one would expect the scores from Table 2 for each criterion, topology (TOP), special similarity (SS) and function (FCN), are each highest on its own axis, but topology and similarity group more closely with each other than either does to function. The former two conflict less with each other than with function. Among historical approaches to the problem, Coddington (1990) scored higher on considerations of special similarity than Levi and Levi (1962) and Saaristo (1978), but none of these were as efficient an explanation of the data as the topological solution (TOP) preferred here.

Table 3
Comparison of four different primary homology hypotheses for theridiid palpal sclerites (A04 = Agnarsson, 2004; C90 = Coddington, 1990; L62 = Levi and Levi, 1962; S78 = Saaristo, 1978), and three primary homology criterion applied to Agnarsson (2004) (A04 = TOP). Highest scores for each are shown in bold

	A04	C90	L62	S78	TOP	FCN	SS
Topology	**1.00**	0.61	0.68	0.67	**1.00**	0.77	0.93
Similarity	0.92	**0.86**	0.72	**0.70**	0.92	0.79	**0.94**
Function	0.77	0.48	0.61	0.41	0.77	**1.00**	0.75
Gain/Loss	0.58	0.58	**0.86**	0.31	0.58	0.59	0.52
Total w G/L	**0.82**	0.63	0.72	0.52	**0.82**	0.79	0.78
Parsimony	**66**	222	121	181	**66**	87	74

Abbreviations and calculations as in Table 2.

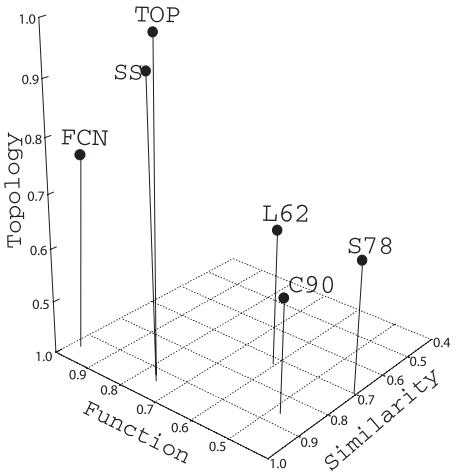

Fig. 4. Values from Table 3 with topology, function and special similarity as orthogonal axes. Abbreviations as in Table 2.

Discussion

This procedure uses a simple scoring system of pairwise comparisons to compare and clarify criteria guiding primary homology hypotheses. Although it can certainly be applied in many contexts and to many groups, comparative morphology is so splendidly diverse that it is pointless to try to anticipate every nuance and puzzle that might arise in its application. We outlined two ways to summarize incongruities, index and step-counting, and two points of view about the trade-off between transformational and gain/loss theories of morphological change, the "preponderance" and "strict" views. The index presented here weights topology, function and similarity equally, so that multiple points of comparison within each criterion, if they exist, are averaged or otherwise down-weighted. Obviously other indices are possible. Step-counting weights each implied change equally, regardless of the sort of difference. Nearly all previous discussions of homology criteria lean towards the former method: *criteria* are considered the comparable units. Step-counting, in contrast, emulates the usual cladistic equal-weights parsimony analysis of characters: each observation is an independent, equally valid falsifier. However, complex structures might provide so many

points of comparison that special similarity might "overwhelm" topology and function as criteria. Weighting is an old issue in systematics, and we have nothing new to add, except to note that most arguments about weighting apply to these considerations as well. The method presented here seems flexible enough to incorporate various weighting schemes.

The preponderance and strict points of view are only two of probably many divergent points of view on how to interpret morphological change. The former prefers transformation, and the latter prefers gain and loss. In our view, although the former may suffer from occasionally tending to rule against the "primary" criterion during a round of scoring (although we did not apply that here), it better preserves homology, produces more broadly applicable homology hypotheses, and will therefore be intrinsically more testable. Gain/loss hypotheses, on the other hand, can be taken to ridiculous lengths, but are constrained by counting against them features that would be obviously evidence of homology under the transformational point of view. In theory it is simple enough to construct a case with many concordant features and only one difference between two structures; in such a case gain/loss would clearly be an inferior explanation. Trends in empirical studies are more difficult to predict.

Step-counting avoids many of the above pitfalls, but seems to be a novel way to approach problems in primary homology. Despite the historical trend to view criteria of primary homology as co-ordinate and somehow equal in explanatory power, a priori equality seems like a poor analytical strategy. Presumably some sort of more sophisticated weighting, perhaps analogous to successive or implied weights is possible, but we do not explore that avenue here.

Topology best explained sclerite homologies in theridiid palps (Fig. 4, see also Table 3). Even when secondary to special similarity, it is nearly as effective in explaining items of special similarity. Homology based strictly on function, on the other hand, does considerable violence to topology and special similarity. Although the generality of that result needs further study, it does support the traditional view that topology is often the most reliable criterion. Historical approaches to the theridiid sclerite problem did not consistently adopt any particular criterion in determining homologies. In contrast we found that strict topology as the primary criterion produced homology statements that explained similarity best in terms of inheritance (Fig. 4). In this case topology was fully congruent with at least one other criterion in every comparison (Agnarsson et al., 2007, figs 4–39). It worked well specifically and generally. Topology will not always perform this well, however; sometimes other criteria will be preferred. The spider embolus, for example, is identified by function (sperm transmission) and special similarity (internal duct). Its insertion on the palp varies wildly and strict topology would fail miserably. Hence, the efficacy of criteria will vary on a case-by-case basis and this method allows testing the criteria in each case separately.

Although tree-thinking may not be *required* to form primary homology hypotheses, what ultimately matters is the overall fit of data to explanation, which *is* a tree. The distinction between primary and secondary homology may therefore be more operational than fundamental. Locally "optimal" primary homology hypotheses may wreak global havoc. The same problem, of course, afflicts tree-searches and sequence alignment. As in those cases, we think the trees implied by primary homology hypotheses can and should feed back on to those hypotheses. Nevertheless, tree-like hypotheses of primary homology are more like character state trees than taxon trees. They account only for transformation between and origin of characters and states as "terminals"; the number of taxa with a particular character or state is immaterial. Evolutionary diversification rates that affect taxon numbers have nothing to do with the validity of homology hypotheses, and therefore should not affect relative scores. Despite these differences, we conclude that quantitative methods to evaluate primary homology hypotheses are probably subject to the same considerations as other quantitative phylogenetic operations.

Strikingly, in the empirical example, strict adherence to a single primary criterion outperformed previous studies based on non-quantitative approaches in almost all cases. Furthermore, previous and novel hypotheses could be compared and ranked in terms of their fit to the data. The method thus seems useful. However, we freely admit that homology is a difficult and subtle problem; atomizing it into categories that receive integer scores does not capture every important consideration. In case it needs to be said, neither will this method *per se* solve long-standing homology problems: the avian digit conundrum will probably persist. However, by analogy to the benefits (and costs!) that quantitative systematics had on classical narrative approaches, quantification and test of congruence in primary homology may also clarify and advance debate.

Acknowledgments

Thanks to Barbara Knoflach for her various contributions to this paper. The manuscript was improved by comments from Maureen Kearney, Philippe Grandcolas, Matjaz Kuntner, Martin Ramírez, and two anonymous reviewers. Support for this work came from came from a Killam Postdoctoral Fellowship, and the USIA Fulbright program (both to I. Agnarsson) and NSF grant EAR-0228699 to J. A. Coddington.

References

Agnarsson, I., 2003. The phylogenetic placement and circumscription of the genus *Synotaxus* (Araneae: Synotaxidae) with a description of a new species from Guyana, and notes on theridioid phylogeny. Invertebr. Syst. 17, 719–734.

Agnarsson, I., 2004. Morphological phylogeny of cobweb spiders and their relatives (Araneae, Araneoidea, Theridiidae). Zool. J. Linn. Soc. 141, 447–626.

Agnarsson, I., 2006a. A revision of the New World eximius lineage of *Anelosimus* (Araneae, Theridiidae) and a phylogenetic analysis using worldwide exemplars. Zool. J. Linn. Soc. 146, 453–593.

Agnarsson, I., 2006b. Asymmetric female genitalia and other remarkable morphology in a new genus of cobweb spiders (Theridiidae, Araneae) from Madagascar. Biol. J. Linn. Soc. 87, 211–232.

Agnarsson, I., Coddington, J.A., Knoflach, B. 2007. The morphology and evolution of cobweb spider male genitalia (Araneae: Theridiidae). J. Arachnol. 37, 334–395.

de Beer, G., 1971. Homology: an Unsolved Problem. Oxford University Press.

Brigandt, I., 2003. Homology in comparative, molecular, and evolutionary developmental biology: The radiation of a concept. J. Exp. Zool. B. Mol. Dev. Evol. 299, 9–17.

Coddington, J.A., 1986. Orb webs in ''non-orb weaving'' ogre faced spiders (Araneae: Deinopidae): a question of genealogy. Cladistics 2, 53–67.

Coddington, J.A., 1990. Ontogeny and homology in the male palpus of orb weaving spiders and their relatives, with comments on phylogeny (Araneoclada: Araneoidea, Deinopoidea). Smithson. Contrib. Zool. 496, 1–52.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, 1st edn. John Murray, London.

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

Eberhard, W.G., 1982. Behavioral characters for the higher classification of orb-weaving spiders. Evolution 36, 1067–1095.

Eberhard, W.G., 1985. Sexual Selection and Animal Genitalia. Harvard University Press, Cambridge, MA.

Farris, J.S., 1983. The logical basis of phylogenetic analysis. In: Platnick, N., Funk, V. (Eds.), Advances in Cladistics, Vol. 2. Columbia University Press, New York, pp. 7–36.

Ghiselin, M.T., 2005. Homology as a relation of correspondence between parts of individuals. Theor. Biosci. 124, 91–103.

Griswold, C.E., Coddington, J.A., Hormiga, G., Scharff, N., 1998. Phylogeny of the orb-web building spiders (Araneae, Orbiculariae: Deinopoidea, Araneoidea). Zool. J. Linn. Soc. 123, 1–99.

Griswold, C.E., Coddington, J.A., Platnick, N.I., Forster, R.R., 1999. Towards a phylogeny of entelegyne spiders (Araneae, Araneomorphae, Entelegynae). J. Arachnol. 27, 53–63.

Griswold, C.E., Ramírez, M.J., Coddington, J.A., Platnick, N.I., 2005. Atlas of phylogenetic data for entelegyne spiders (Araneae: Araneomorphae: Entelegynae) with comments on their phylogeny. Proc. Calif. Acad. Sci. 56, 1–324.

Hall, B.K., 1995. Homology and embryonic development. Evol. Biol. 28, 1–37.

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

Hormiga, G., 1994a. A revision and cladistic analysis of the spider family Pimoidae (Araneoidea, Araneae). Smithson. Contrib. Zool. 549, 1–104.

Hormiga, G., 1994b. Cladistics and the comparative morphology of linyphiid spider and their relatives (Araneae, Araneoidea, Linyphiidae). Zool. J. Linn. Soc. 111, 1–71.

Hormiga, G., 2000. Higher level phylogenetics of erigonine spiders (Araneae, Linyphiidae, Erigoninae). Smithson. Contrib. Zool. 609, 1–160.

Hormiga, G., Eberhard, W.G., Coddington, J.A., 1995. Web construction behavior in Australian *Phonognatha* and the phylogeny of nephiline and tetragnathid spiders (Araneae, Tetragnathidae). Austral. J. Zool. 43, 313–364.

Hoßfeld, U., Olsson, L., 2005. The history of the homology concept and the ''Phylogenetisches Symposium''. Theor. Biosci. 124, 243–253.

Huber, B.A., 2003. Rapid evolution and species-specificity of arthropod genitalia: fact or artifact? Org. Divers. Evol. 3, 63–71.

Jardine, N., 1967. The concept of homology in biology. Br. J. Philos. Sci. 18, 125–139.

Kluge, A.G., 2003. The repugnant and the mature in phylogenetic inference: atemporal similarity and historical identity. Cladistics 19, 356–368.

Knoflach, B., 1991. *Achaearanea tabulata* Levi, eine für Österreich neue Kugelspinne (Arachnida, Aranei: Theridiidae). Ber. Nat.-Med. Verein Innsbruck 78, 59–64.

Knoflach, B., 1992. Neue *Robertus*-Funde in den Alpen: *R. mediterraneus* Eskov und *Robertus* sp. (Arachnida, Aranei: Theridiidae). Ber. Nat.-Med. Verein Innsbruck, 79, 161–171.

Knoflach, B., 1993a. *Theridion conigerum* Simon-rediscovered in Austria (Araneida: Theridiidae). Bull. Br. Arachnol. Soc. 9, 205–208.

Knoflach, B., 1993b. Das Mannchen von *Episinus theridioides* Simon (Arachnida: Araneae, Theridiidae). Mitt. Schweizerischen Entomologischen Gesellschaft 66, 359–366.

Knoflach, B., van Harten, A., 2000. Palpal loss, single palp copulation and obligatory mate consumption in *Tidarren cuneolatum* (Tullgren, 1910) (Araneae, Theridiidae). J. Nat. Hist. 34, 1639–1659.

Knoflach, B., van Harten, A., 2001. *Tidarren argo* sp. nov. (Araneae: Theridiidae) and its exceptional copulatory behaviour: emasculation, male palpal organ as a mating plug and sexual cannibalism. J. Zool. Lond. 254, 449–445.

Larimer, J.L., Pease, C.M., 1990. Unexpected divergence among identified interneurons in different abdominal segments of the crayfish, *Procambarus clarkii*. J. Exp. Zool. 253, 20–29.

Levi, H.W., Levi, L.R., 1962. The genera of the spider family Theridiidae. Bull. Mus. Comp. Zool. 127, 1–71.

Miller, J.A., 2007. Review of erigonine spider genera in the Neotropics (Araneae: Linyphiidae, Erigoninae). Zool. J. Linn. Soc. 149, 1–263.

Minelli, A., Schram, F.R., 1994. Owen revisited: a reappraisal of morphology in evolutionary biology. Bijdragen Tot Dierkunde 64, 65–74.

Nelson, G.J., 1978. Ontogeny, phylogeny, paleontology and the biogenetic law. Syst. Zool. 27, 324–345.

Nelson, G.J., 1994. Homology and systematics. In: Hall, B.K. (Ed.), Homology: the Hierarchical Basis of Comparative Biology. Academic Press, San Diego, pp. 101–149.

Owen, R., 1843. Lectures on the 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.

Pearson, K.G., Boyan, G.S., Bastiani, M., Goodman, C.S., 1985. Heterogeneous properties of segmentally homologous interneurons in the ventral nerve cord of locusts. J. Comp. Neurol. 233, 133–145.

Phillips, A.J., 2006. Homology assessment and molecular sequence alignment. J. Biomed. Informat. 39, 18–33.

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

Pogue, M.G., Mickevich, M.F., 1990. Character definitions and character state delineation—the bete-noir of phylogenetic inference. Cladistics 5, 319–361.

Ramírez, M.J., in press. Homology as a parsimony problem: a dynamic homology approach for morphological data. Cladistics, doi: 10.1111/j.1096-0031.2007.00162.x.

Remane, A., 1952. Die Grundlagen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik. Geest und Portig, Leipzig.

Richter. S., 2005. Homologies in phylogenetic analyses—concept and tests. Theor. Biosci. 124, 105–120.

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, pp. 63–100.

Rieppel, O., 2001. Preformationist and epigenetic biases in the history of the morphological character concept. In: Wagner, G.P. (Ed.), The Character Concept in Evolutionary Biology. Academic Press, San Diego, pp. 55–75.

Rieppel, O., 2006. The merits of similarity reconsidered. Syst. Biodiversity, 4, 137–147.

Rieppel, O., Kearney, M., 2002. Similarity. Biol. J. Linn. Soc. 75, 59–82.

Roth, V.L., 1984. On homology. Biol. J. Linn. Soc. 22, 13–29.

Rutishauser, R., Moline, P., 2005. Evo-devo and the search for homology (''sameness'') in biological systems. Theor. Biosci. 124, 213–241.

Saaristo, M.I., 1978. Spiders (Arachnida, Araneae) from Seychelles Islands, with notes on taxonomy. Ann. Zool. Fenn. 15, 99–126.

Scharff, N., Coddington, J.A., 1997. A phylogenetic analysis of the orb-weaving spider family Araneidae (Arachnida, Araneae). Zool. J. Linn. Soc. 120, 355–434.

Scholtz, G., 2005. Homology and ontogeny: pattern and process in comparative developmental biology. Theor. Biosci. 124, 121–143.

Simmons, M.P., Freudenstein, J.V., 2003. The effects of increasing genetic distance on alignment of, and tree construction from, rDNA internal transcribed spacer sequences Mol. Phylogenet. Evol. 26, 444–451.

Wägele, J.-W., 2005. Foundations of Phylogenetic Systematics. Verlag Dr. Friedrich Pfeil, München.

Wagner, G.P., 2005. The developmental evolution of avian digit homology: an update. Theor. Biosci. 124, 165–183.

Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics. Cladistics 12, 1–9.

Wheeler, W.C., 1999. Fixed character states and the optimization of molecular sequence data. Cladistics 15, 379–385.

Wheeler, W.C., 2003. Search-based optimization. Cladistics 19, 348–355.

Wheeler, W.C., 2006. Dynamic homology and the likelihood criterion. Cladistics 22, 157–170.