*Syst. Biol.* 54(2):230–240, 2005

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150590923227

# A New Technique for Identifying Sequence Heterochrony

JONATHAN E. JEFFERY,[1] OLAF R. P. BININDA-EMONDS,[2] MICHAEL I. COATES,[3] AND MICHAEL K. RICHARDSON[1]

[1]*Institute of Biology, Leiden University, Kaiserstraat 63, 2311GP Leiden, The Netherlands*
[2]*Lehrstuhl für Tierzucht, Technical University of Munich, Alte Akademie 12, 85354 Freising-Weihenstephan, Germany; E-mail: Olaf.Bininda@tierzucht.tum.de*
[3]*Organismal Biology and Anatomy, University of Chicago, 1027 E. 57th Street, Chicago, Illinois 60637, USA*

*Abstract*.—Sequence heterochrony (changes in the order in which events occur) is a potentially important, but relatively poorly explored, mechanism for the evolution of development. In part, this is because of the inherent difficulties in inferring sequence heterochrony across species. The event-pairing method, developed independently by several workers in the mid-1990s, encodes sequences in a way that allows them to be examined in a phylogenetic framework, but the results can be difficult to interpret in terms of actual heterochronic changes. Here, we describe a new, parsimony-based method to interpret such results. For each branch of the tree, it identifies the least number of event movements (heterochronies) that will explain all the observed event-pair changes. It has the potential to find all alternative, equally parsimonious explanations, and generate a consensus, containing the movements that form part of every equally most parsimonious explanation. This new technique, which we call Parsimov, greatly increases the utility of the event-pair method for inferring instances of sequence heterochrony. [Developmental sequences; event pairing; heterochrony; Parsimov program; principle of parsimony.]

Developmental data from temporal sequences, especially those in embryonic development, are increasingly being analyzed in an evolutionary context (e.g., Mabee and Trendler, 1996; Smith, 1996, 1997, 2001a, 2001b; Velhagen, 1997; Nunn and Smith, 1998; Velhagen and Savitzky, 1998; Mabee et al., 2000; Chipman et al., 2000; Schlosser, 2001; Chipman, 2002; Jeffery et al., 2002a; Sánchez-Villagra, 2002; Bininda-Emonds et al., 2003; Cole et al., 2003). However, there are problems in optimizing such data onto phylogenetic trees in that the differences between the sequence of each species (sequence heterochrony) makes it difficult to "align" the data, and thus to determine which parts of the sequence have changed in a rigorous analytical fashion (see Bininda-Emonds et al., 2002).

Of the various methods that have been suggested for studying sequence heterochrony (e.g., Schlosser, 2001; Schulmeister and Wheeler, 2004), the most flexible is the "event-pair" (or "sequence unit") method, which was developed in the mid-1990s to study heterochrony in vertebrate embryonic development (Mabee and Trendler, 1996; Smith, 1997; Velhagen, 1997). This method encodes the relative position of each item in the sequence, thereby allowing comparisons between species to be made. The use of relative timing is necessary because variation in the overall rate of development (both between species and to some extent between individuals of the same species) prevents the use of absolute timing data (e.g., in hours, minutes, and seconds) in phylogenetic comparisons (for a full discussion, see Nunn and Smith, 1998; Bininda-Emonds et al., 2002; Jeffery et al., 2002a, 2002b).

The principle of event-pairing is simple—any developmental sequence can be broken down into a series of events. These may be, for example, the first appearance of a distinctive morphology or the first expression of a particular gene (Bininda-Emonds et al., 2002). Within a sequence, any two events can have only one of three relative timing relationships: event A occurs before event B, A and B occur simultaneously, or event A occurs after event B. By convention, these event-pairs are given the numerical scores 0, 1, and 2, respectively. Scoring every possible nonredundant event-pair will encode the entire developmental sequence (Fig. 1). Event-paired sequence data for different species can then be mapped onto a phylogenetic tree. Evolutionary changes in the developmental sequences are indicated by apomorphic event-pair changes (Fig. 2, Table 1). The evolutionary polarity of any changes can be inferred in the same way as conventional characters (e.g., using outgroup analysis to determine symplesiomorphies, synapomorphies, and autapomorphies).

Unfortunately, the very simplicity of the event-pair method makes interpretation of the apomorphic changes difficult (Jeffery et al., 2002b). For example, it is clear from the list of changes in Table 1 that there have been several changes in the developmental sequence from species X to species Y. However, as shown in Figure 3A and B, each change in event-pair score could have been the result of any of five different combinations of event movements. Thus, it is impossible without further information to determine the actual changes in the developmental sequence that caused the event-pair changes (i.e., which events have moved and in which direction; Jeffery et al., 2002b). Yet, in biological studies, it is precisely the movement of individual events that is relevant for determining phenomena such as evolutionary lability or event-interactions like induction or modularity (e.g., Wagner, 1996; Galis and Metz, 2001).

To overcome this limitation, we previously proposed a method called "event-pair cracking" which analyzes all the event-pair changes at a given node en bloc in an attempt to establish the underlying pattern of event-movement (the "solution") (Jeffery et al., 2002b). Briefly, cracking operates by identifying those events that move relative to the greatest number of other events and in a consistent direction (i.e., are always inferred as moving earlier/later in the developmental sequence). As such, the method is computationally quick and simple;

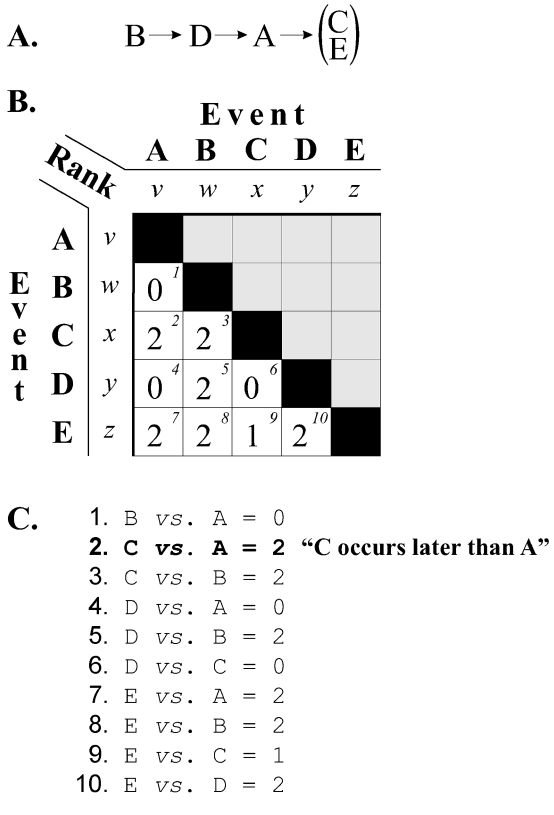



FIGURE 1. Conventions of event-paired data. (A) A sequence of five events (A to E). Events C and E occur simultaneously. (B) Event-pair table for the five events. There are 10 event-pairs, given by $1/2\,(n^2 - n)$ where $n$ is the number of events. The superscript in each cell gives the order in which event-pairs are listed when given as a string. This order is arbitrary, but must be used consistently for later calculations. The lowercase letters (v to z) represent the ranks in the sequence of events A to E, respectively. Including the 'upper triangle' of the table (gray cells) would produce redundant event-pairs, duplicated in the lower triangle. The black squares are trivial comparisons (e.g., A versus A, B versus B, etc.). (C) Event-pairs listed in order together with their scores. As an example, the score of event-pair number two (C versus A) is also given in bold. The particular event-pair scores shown would be given if the ranks v to z were 3, 1, 4, 2, and 4, respectively. (D) Event-pairs assembled as a string ready for analysis (e.g., with PAUP*).

software implementations usually take a few seconds to analyze an entire tree, even with a large number of events (e.g., Bininda-Emonds et al., 2003). However, the solution is dependent on a user-determined threshold level that is used to discriminate actively moving events (i.e., those showing large, coherent movement) from those that are only apparently moving in relation to these events. Determining this level is usually subjective and based on trial and error. Thus, depending on the stringency level, event-pair cracking might not identify all the event movements required to explain all the event-pair changes along a given branch (too conservative, leading to false negatives) or might identify too many (too liberal, leading to false positives). In fact, we previously noted that "The procedure is slightly conservative. [Event-movements] will be identified only if they are relatively large and coherent" (Jeffery et al., 2002b:482). Also, cracking can only find a single solution for each node for a given stringency level and there is no ready means to compare these solutions to determine if one is "better" than another.

TABLE 1. The 12 changes occurring along the branch from species X to species Y in Figure 2. Event-pair 2, for example, indicates that event C moved *early* with respect to event A.

| Event-pair | Involves… | Change |
|---|---|---|
| 1 | B versus A | 1 → 0 |
| 2 | C versus A | 2 → 0 |
| 4 | D versus A | 2 → 0 |
| 7 | E versus A | 2 → 0 |
| 16 | G versus A | 2 → 0 |
| 19 | G versus D | 2 → 0 |
| 20 | G versus E | 2 → 0 |
| 21 | G versus F | 2 → 0 |
| 29 | I versus A | 2 → 0 |
| 33 | I versus E | 2 → 1 |
| 34 | I versus F | 2 → 0 |
| 36 | I versus H | 2 → 0 |

In this paper, we use a worked example to describe a new method based on the principle of parsimony that identifies only the minimum number of events required to explain all observed event-pair changes along any given branch. Although the Parsimov method is computationally more intense than cracking (analyses can take hours rather than seconds), it has the potential to investigate all possible solutions to determine the best one based on a parsimony optimality criterion. Concordance and conflicts between multiple equally optimal solutions can be examined and a strict consensus solution can be calculated to include only those movements that form part of every equally most parsimonious solution.

## THEORETICAL UNDERPINNINGS OF THE PARSIMOV METHOD

Changes in an event-pair score imply that one or both events in the associated pair have shifted their positions in the developmental sequence (see Fig. 2 and Table 1). However, as noted above, there are always five possible ways (solutions) in which two events can move to produce the event-pair change (see Fig. 3A and B). Event-pair changes in isolation give no unequivocal evidence for any particular pattern of movement, and so adopting any particular solution in favor of the others will be an arbitrary decision unless further, independent information is available. In the absence of such evidence, each solution requires one or more ad hoc hypotheses that particular events have shifted in the developmental sequence to explain the observed event-pair changes. The principle of parsimony demands that such ad hoc decisions should be kept to a minimum, giving us a criterion with which to choose between the solutions. At the level of a single event-pair, this makes solutions (i) and (v) the most parsimonious of the alternatives in Figure 3A and B because both postulate the movement of only a single event. However, this principle can be extended to all the

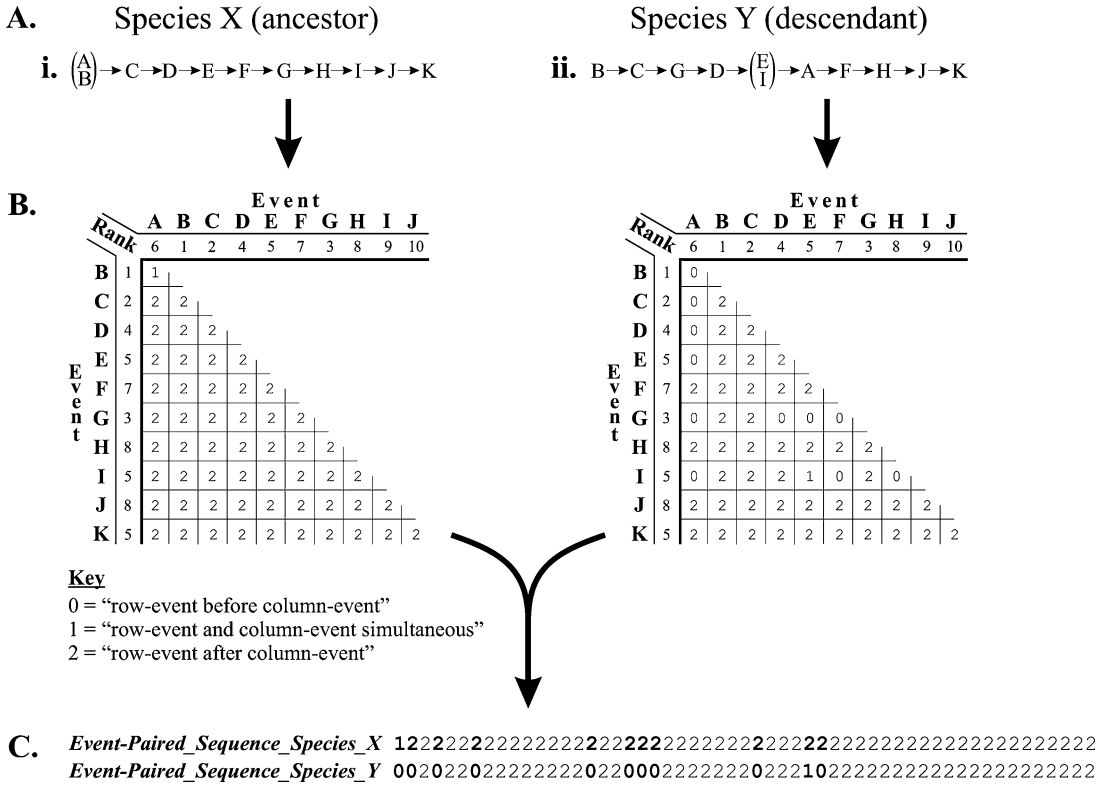



FIGURE 2. Hypothetical example of a heterochronic sequence change. (A) Two sequences of 11 events (A to K). Events that occur simultaneously are grouped in brackets. (B) Event-pair tables constructed for the sequences of species X and Y, respectively. (C) Fifty-five event-pairs listed for phylogenetic analysis. Event-pairs listed in bold show a change between species X and Y. See Figure 1 for the conventions of event-pair tables and lists.

event-pair changes along a given branch (i.e., for the entire data set): we seek the solution that explains all the observed event-pair changes with the smallest possible number of ad hoc hypotheses of event-movement. For example, Figure 4A shows four event-pair apomorphies involving five events. The solution in Figure 4B requires four ad hoc hypotheses (that B, C, D, and E shifted early), whereas that in Figure 4C requires only a single hypothesis (that A shifted late). The latter solution, therefore, is more parsimonious. It should be adopted, in preference

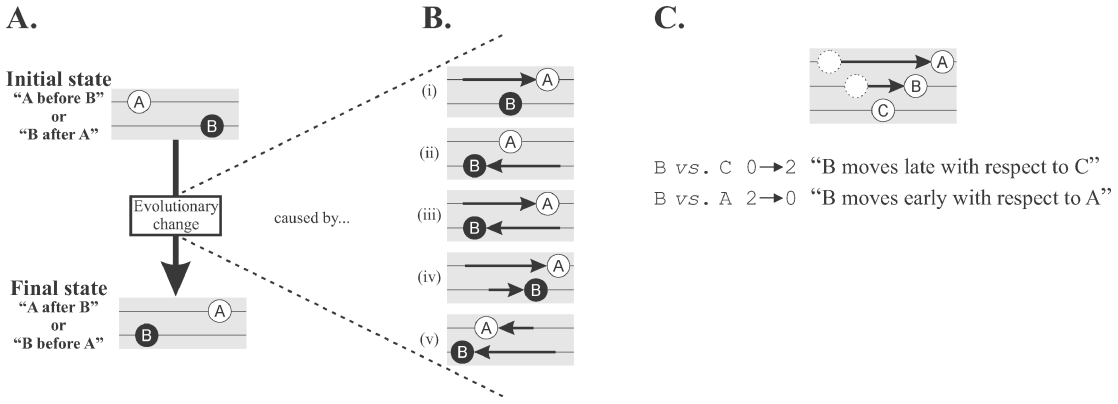
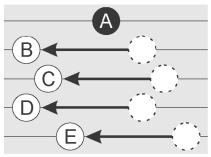



FIGURE 3. Possible movements causing an observed heterochronic change. (A) Two events (A and B) have a particular timing relationship (event A occurs before event B). However, an evolutionary transformation leads to event A occurring after event B. (B) There are five possible ways in which this change might have occurred: (i) event A moved later in development; (ii) event B moved earlier in development; (iii) event A moved later and event B moved earlier in development; (iv) event B moved later in development but event A moved even later; and (v) event A moved earlier in development, but event B moved even earlier. Figure after Jeffery et al. (2002b: fig. 2). (C) If a moving event (event B) is "overtaken" by another event (event A), it can generate incoherent event-pair changes, apparently implying that the event has moved both early and late. This must be accounted for when calculating a solution to a set of event-pair changes.

**A.**

| Event-pair | Change |
|---|---|
| B *vs.* A | 2 → 0 |
| C *vs.* A | 2 → 0 |
| D *vs.* A | 2 → 0 |
| E *vs.* A | 2 → 0 |

**B.**




**C.**

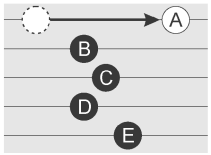



FIGURE 4. Using parsimony to choose between alternative schemes of movement. (A) Four event-pair changes are observed along a particular branch. Five events (A to E) are involved, but many different schemes of movement could produce the changes (see Fig. 3). (B) One scheme involves events B, C, D, and E moving early relative to event A. The four events must have moved in a coordinated fashion because any change in their positions relative to one another would have caused additional event-pair changes. (C) An alternative scheme involves event A moving late with respect to events B to E. This scheme requires only one ad hoc hypothesis (i.e., that event A moved) to explain all the observed event-pair changes in contrast to the four hypotheses required in (B) (events B, C, D, and E all moved early). This scheme is therefore more parsimonious and should be the preferred explanation of the observed event-pair changes.

to other solutions, as the working hypothesis of the event movements that gave rise to the observed pattern of event-pair change.

It should be noted that we have implemented Parsimov using a simple parsimony criterion that assumes a single step for each event along a given branch. However, other weighting schemes or even optimization criteria could be used. Likelihood or Bayesian versions of this method are easy to envisage given the development of appropriate probabilistic models of change (e.g., Lewis, 2001). Our implementation also assumes that, in the absence of information to the contrary, the movements of events are independent of one another. However, the realization that some events are linked to form developmental complexes or modules could affect the inference of event movement. For example, if events B, C, D, and E in Figure 4 were linked, then the solution in Figure 4B could be explained by only a single ad hoc hypothesis of movement (that of the complex as a whole) and thus would be equally parsimonious with the solution in Figure 4C.

## STEPS IN THE PARSIMOV METHOD

In this section, we first describe the basic procedure of the Parsimov method using a simple worked example based on the modest data set in Figure 2 and Table 1. We deliberately employ a naïve analytical approach in this example for illustration purposes. In reality, searching for the optimal set of event movements poses similar analytical difficulties to determining the optimal phylogenetic tree(s) for a set of morphological or molecular character data. Therefore, we later describe several more sophisticated search strategies in the next section that show high similarity with the more familiar tree searching heuristics. In all cases, we consider only a single branch on the example phylogeny; the same process is applied to each branch of the tree in turn to obtain the global solution.

### *Step 1—Collecting the Data*

As for event-pair cracking, the initial step proceeds as for the reconstruction of conventional characters. The event-pair–encoded data are mapped onto a phylogenetic tree and the apomorphic changes along each branch are determined. However, in an event-pair analysis, each character is a specific event-pair (e.g., B versus A, E versus G). Therefore, as with the cracking method, the characters need to be translated back into the event-pairs they represent. By convention, event-pairs are listed in the order shown in Figure 1. Thus, for example, we know that a change in character 2 represents a timing change between events C and A. The synapomorphic event-pair changes for our example are given in Table 1.

### *Step 2—Listing the Changes*

Once the event-pair changes are known, the distribution of the changes across the events is established. Some events might not be involved in any event-pair changes along the branch in question (in our example, events J and K) and so have not changed position with respect to any other event. They can therefore safely be ignored in subsequent steps. This is significant because it reduces the number of comparisons required, thereby speeding up the analysis at the node in question. This is similar to excluding invariant characters from a conventional parsimony-based phylogenetic analysis because they will not affect the resultant solution.

Although the polarity of all apomorphic changes can be determined in the same manner as for the reconstruction of conventional characters, event-pairs are inherently non-directional. In Table 1, for example, event-pair 2 describes the movement of event C relative to event A. This event-pair has its inverted equivalent (2′ in Table 2) describing the movement of event A relative to event C. This is the event-pair that would be recorded in the gray upper triangle of the event-pair matrix in Figure 1. Either is a valid description of the observed relative change: the different views are referred to as the perspective of the event-pair (see Jeffery et al., 2002b). Only one perspective is used when optimizing event-pairs onto trees to prevent a needless duplication of the data set. However, when searching for the best solutions, both perspectives must be considered, including the inverse data in Table 2.

When event-pair changes from both perspectives are compiled for each event in our example, the set of

TABLE 2. Because event-pairs are inherently nondirectional (see Fig. 3), the changes in Table 1 can be viewed from an alternate perspective. For example, event-pair 2 in Table 1 suggests that the change was caused by event C moving early. The alternate perspective, event-pair 2′ in this table, suggests that it was caused by event A moving late. When calculating the best overall solution to a set of event-pairs, both perspectives must be considered.

| Event-pair | Involves... | Change |
|---|---|---|
| 1′ | A versus B | 0 → 1 |
| 2′ | A versus C | 0 → 2 |
| 4′ | A versus D | 0 → 2 |
| 7′ | A versus E | 0 → 2 |
| 16′ | A versus G | 0 → 2 |
| 19′ | D versus G | 0 → 2 |
| 20′ | E versus G | 0 → 2 |
| 21′ | F versus G | 0 → 2 |
| 29′ | A versus I | 0 → 2 |
| 33′ | E versus I | 1 → 2 |
| 34′ | F versus I | 0 → 2 |
| 36′ | H versus I | 0 → 2 |

changes that any solution must account for is shown in Table 3.

### *Step 3—Determining Event Movements*

To generate a set of event movements, events are selected iteratively with the assumption that the event-pair changes they are involved in are "real" (that is, caused by the movement of the selected event, rather than by its event-pair partner). In so doing, the ad hoc hypothesis of the selected event moving will explain a set of event-pair changes that, therefore, do not need to be considered in further rounds. For example, in Table 3, the assumption in the first round that event A moved late explains six event-pair changes (A versus each of B, C, D, E, G, and I). These can be deleted from the table along with their inverted equivalents (each of B, C, D, E, G, and I versus A) to leave six event-pair changes to be explained (Table 4). We therefore repeat the process, selecting another event

TABLE 3. Event-pair changes summarized for each event in Tables 1 and 2. For each event-pair change, both perspectives are listed. For example, event H moving late with respect to event I is also listed as event I moving early with respect to event H. The last three columns list the number of event-pair changes each event was involved in. TRC (total relative change) gives –1 for each entry in the "early" column and +1 for each entry in the "late" column. TAC (total absolute change) gives +1 regardless of the direction of change. Abs. prod. is the absolute value of the product of the TRC and TAC.

| | Moves | | Total change | | |
|---|---|---|---|---|---|
| Event | Early with respect to... | Late with respect to... | TRC | TAC | Abs. prod. |
| A | | B, C, D, E, G, I | 6 | 6 | 36 |
| B | A | | −1 | 1 | 1 |
| C | A | | −1 | 1 | 1 |
| D | A | G | 0 | 2 | 0 |
| E | A | G, I | 1 | 3 | 3 |
| F | | G, I | 2 | 2 | 4 |
| G | A, D, E, F | | −4 | 4 | 16 |
| H | | I | 1 | 1 | 1 |
| I | A, E, F, H | | −4 | 4 | 16 |

TABLE 4. Recalculation of Table 3 after the first round of analysis where event A was assumed to have moved late with respect to events B, C, D, E, G, and I. These event-pairs have been removed (from both perspectives), along with any events that no longer show any changes (events B and C). The remaining events have had their total change scores recalculated.

| | Moves | | Total change | | |
|---|---|---|---|---|---|
| Event | Early with respect to... | Late with respect to... | TRC | TAC | Abs. prod. |
| D | | G | 1 | 1 | 1 |
| E | | G, I | 2 | 2 | 4 |
| F | | G, I | 2 | 2 | 4 |
| G | D, E, F | | −3 | 3 | 9 |
| H | | I | 1 | 1 | 1 |
| I | E, F, H | | −3 | 3 | 9 |

in the next round (in our example, event D, the next event on the list with any event-pair changes remaining). The process is repeated until all the event-pair changes are explained. Each round adds another ad hoc hypothesis of an event moving early or late.

*Input orders and event-pair coherence.*—We refer to the order in which events are examined as the "input order"; it is analogous to a taxon addition sequence in conventional tree-searching heuristics. In our example, taking the input order from the order in which the events are listed in Tables 3 and 4 required five rounds, and thus five ad hoc hypotheses of event movement, to explain the 12 event-pairs changes (Table 5a). We can test other solutions by using different input orders. Table 5b and c show the solutions generated by two alternative input orders. In Table 5b, we started with event B but then continued as before in the order the events are listed in Table 3. This resulted in a less optimal solution, requiring nine ad hoc hypotheses of movement. It also highlights the importance of the decisions made about the direction of event-movement. For example, it was easy to decide the direction of movement of event A from Table 3 because all its event-pairs suggest that it moved late (they are "coherent"; sensu Jeffery et al., 2002b). However, in Table 3, event D has two "incoherent" event-pairs, suggesting opposing directions of movement. Clearly, event D cannot be moving in both directions at the same time—one event-pair must give the correct movement of event D, whereas the other is caused by the movement of an event-pair partner (this pattern of event-pair changes can be caused by the movements shown in Fig. 3C).

Unfortunately, without additional evidence, we cannot determine which of the two event-pairs gives the correct movement of event D. Yet, the choice has further consequences. For example, the ad hoc hypothesis that event D moved early with respect to event A (as in Table 5b) explains one of the event-pairs involving D. However, the only way to explain the remaining event-pair is to assume that event G moved early with respect to event D, an additional hypothesis of movement (marked with an asterisk in Table 5b). Alternatively, assuming that event D moved late with respect to event G requires that event A moved late with respect to event D. Although

TABLE 5. Three alternative solutions for the data in Figure 2. Any solution requires some ad hoc hypotheses about the movement underlying the observed event-pair changes. The number of hypotheses required is sensitive to the order in which the events were assessed (the input order). The third solution requires only three hypotheses and is thus the most parsimonious. Asterisks show movements required by incoherent movements (see main text).

| Solution | Input order | Resultant hypotheses | Events with event-pairs left to be explained |
|---|---|---|---|
| a | A | A moved *late* relative to B, C, D, E, G, I | D, E, F, G, H, I |
| | D | D moved *late* relative to G | E, F, G, H, I |
| | E | E moved *late* relative to G, I | F, G, H, I |
| | F | F moved *late* relative to G, I | H, I |
| | H | H moved *late* relative to I | — |
| b | B | B moved *early* relative to A | A, C, D, E, F, G, H, I |
| | C | C moved *early* relative to A | A, D, E, F, G, H, I |
| | D | D moved *early* relative to A | A, D, E, F, G, H, I |
| | | G moved *early* relative to D* | A, E, F, G, H, I |
| | E | E moved *late* relative to G, I | A, E, F, G, H, I |
| | | A moved *late* relative to E* | A, F, G, H, I |
| | F | F moved *late* relative to G, I | A, G, H, I |
| | G | G moved *early* relative to A | A, H, I |
| | H | H moved *late* relative to I | A, I |
| | I | I moved *early* relative to A | — |
| c | A | A moved *late* relative to B, C, D, E, G, I | D, E, F, G, H, I |
| | G | G moved *early* relative to D, E, F | E, F, H, I |
| | I | I moved *early* relative to E, F, H | — |

both cases require two ad hoc hypotheses to explain the event-pairs involving D, they have different effects on the overall solution. In Table 5b, we arbitrarily chose to assume that D moved early (and therefore G moved early). However, a full search should also examine the solution produced when we assume that D moved late (and therefore A moves late).

### *Search Strategies*

Ideally, we would systematically assess every possible input order, including every alternative assumption of direction, to determine the optimal solution set. For example, even though the movement for event A in our worked example is coherent and always indicated to be late, we should test the solution produced if we assume that A moved early. Note that doing so effectively fixes the directions of movement for events B, C, D, E, G, and I as early (i.e., even earlier; solution (v) in Fig. 3) to explain the event-pair. In so doing, this might explain other event-pairs within the same round, which can then be deleted.

The number of possible input orders increases rapidly with the number of events considered (Table 6)—in our example there are a maximum of over 186 million input orders. Clearly, many different input orders might lead to the same solution, but it is not possible to predict the outcome of a particular input order. This problem is comparable to conventional phylogenetic analyses, which aims to find the optimal solution(s) based on a particular optimality criterion from millions of alternatives. Like phylogenetic analysis, an exhaustive search of all possible input orders is possible only when there are few events (in practice, about six or fewer) contributing to the event-pair synapomorphies. Beyond this, heuristic search strategies must be employed. We first describe a rapid heuristic that can be used as the basis for a more thorough, branch-and-bound-like strategy.

*A rapid heuristic using "total change" scores.*—Jeffery et al. (2002b) described the use of total absolute and total relative change scores (TAC and TRC, respectively) to select actively moving events in event-pair cracking. For each event, the TAC is simply the number of event-pair changes it is involved in. The TRC is the sum of the event-pairs changes with the direction of movement accounted for (i.e., giving +1 for a "late with respect to . . . " change, and −1 for an "early with respect to . . . " change). For example, in Table 3, for event A, TAC = 6 (it is involved in six event-pairs) and TRC = 6 (six late changes, each with a score of +1). By contrast, for event D, TAC = 2, but TRC = 0 (one early change with a score of −1 and one late change with a score of +1). Together, these scores were held to give a good picture of the "explanatory power" of each event; that is, how many event-pair changes each could explain if it was assumed to have moved, and how many additional ad hoc hypotheses this would entail. In the current context, the TAC and TRC scores can be used to determine the input order by using the highest scoring event in each round in the Parsimov procedure. We suggest using the absolute value of the product of the TRC and TAC, but other values, such as the TRC or TAC alone, could be used. Furthermore, instead of examining both directions of movement (early or late) for a given event, the direction is determined by the TRC value; events where TRC = 0 are given a direction at random. Altogether, this forms the basis of a fast, but greedy algorithm that seeks to explain the largest number of event-pair movements in a given round, but should still obtain a near-optimal solution.

Thus, in our example, event A is selected in the first round based on its high total change score (TAC × TRC = 36) and is inferred to be moving late based on its positive TRC score (+6). After the six event-pair changes explained by the movement of event A are removed, the TRC and TAC of the remaining events are recalculated

TABLE 6. Input orders used in Parsimov searches (decreasing in thoroughness from left to right). If only a few events are involved, every single possible explanation of the observed distribution of event-pairs can be assessed (column 2). This exhaustive search method is guaranteed to find shortest solution(s). However, the number of possible solutions increases rapidly with the number of events involved. Alternatively, every possible combination of event-movement can be examined, but with only a single direction of movement for each event that is determined heuristically ("unpolarized"; see main text). This strategy cannot guarantee that the shortest solution(s) will be recovered, but the number of solutions does not increase as rapidly with the number of events. Finally, all input orders for either the first two or three events can be examined, with the input orders of the remaining events (and all directions of movement) determined heuristically. Again, these strategies cannot guarantee that the shortest solution(s) will be recovered. However, because every possible combination of starting input order is investigated, they are unlikely to become stuck in an "island" of locally optimal solutions. In each column, the gray area gives the number of event-movements that can be investigated within a day (and within memory limits) using our Parsimov software on a standard desktop PC.

| | Input orders to be assessed using | | | |
|---|---|---|---|---|
| No. of events | All possible input orders (polarized) | All possible input orders (unpolarized) | All possible combinations of three unpolarized events | All possible combinations of two unpolarized events |
| 2 | 8 | 2 | — | 2 |
| 3 | 48 | 6 | 6 | 6 |
| 4 | 384 | 24 | 24 | 12 |
| 5 | 3 840 | 120 | 60 | 20 |
| 6 | 46 080 | 720 | 120 | 30 |
| 7 | 645 120 | 5 040 | 210 | 42 |
| 8 | 10 321 920 | 40 320 | 336 | 56 |
| 9 | $\sim 1.86 \times 10^{08}$ | 362 880 | 504 | 72 |
| 10 | $\sim 3.72 \times 10^{09}$ | 3 628 800 | 720 | 90 |
| 11 | $\sim 8.17 \times 10^{10}$ | 39 916 800 | 990 | 110 |
| 12 | $\sim 1.96 \times 10^{12}$ | 479 001 600 | 1 320 | 132 |
| 13 | $\sim 5.10 \times 10^{13}$ | 6 227 020 800 | 1 716 | 156 |
| 14 | $\sim 1.43 \times 10^{15}$ | 87 178 291 200 | 2 184 | 182 |
| 15 | $\sim 4.28 \times 10^{16}$ | $\sim 1.31 \times 10^{12}$ | 2 730 | 210 |
| 16 | $\sim 1.37 \times 10^{18}$ | $\sim 2.09 \times 10^{13}$ | 3 360 | 240 |
| 17 | $\sim 4.66 \times 10^{19}$ | $\sim 3.56 \times 10^{14}$ | 4 080 | 272 |
| 18 | $\sim 1.68 \times 10^{21}$ | $\sim 6.40 \times 10^{15}$ | 4 896 | 306 |
| 19 | $\sim 6.38 \times 10^{22}$ | $\sim 1.22 \times 10^{17}$ | 5 814 | 342 |
| 20 | $\sim 2.55 \times 10^{24}$ | $\sim 2.43 \times 10^{18}$ | 6 840 | 380 |
| 21 | $\sim 1.07 \times 10^{26}$ | $\sim 5.11 \times 10^{19}$ | 7 980 | 420 |
| 22 | $\sim 4.71 \times 10^{27}$ | $\sim 1.12 \times 10^{21}$ | 9 240 | 462 |
| 23 | $\sim 2.17 \times 10^{29}$ | $\sim 2.59 \times 10^{22}$ | 10 626 | 506 |
| 24 | $\sim 1.04 \times 10^{31}$ | $\sim 6.20 \times 10^{23}$ | 12 144 | 552 |
| 25 | $\sim 5.20 \times 10^{32}$ | $\sim 1.55 \times 10^{25}$ | 13 800 | 600 |
| 26 | $\sim 2.71 \times 10^{34}$ | $\sim 4.03 \times 10^{26}$ | 15 600 | 650 |
| 27 | $\sim 1.46 \times 10^{36}$ | $\sim 1.09 \times 10^{28}$ | 17 550 | 702 |
| 28 | $\sim 8.18 \times 10^{37}$ | $\sim 3.05 \times 10^{29}$ | 19 656 | 756 |
| 29 | $\sim 4.75 \times 10^{39}$ | $\sim 8.84 \times 10^{30}$ | 21 924 | 812 |
| 30 | $\sim 2.85 \times 10^{41}$ | $\sim 2.65 \times 10^{32}$ | 24 360 | 870 |
| 31 | $\sim 1.77 \times 10^{43}$ | $\sim 8.22 \times 10^{33}$ | 26 970 | 930 |

(see Table 4), and the new highest scoring event is selected. In this example, events G and I have the same total change score (nine). One is chosen arbitrarily (in our example, G, inferred to be moving early based on its negative TRC score), the (three) event-pair changes its movement explains are removed, and the TRC and TAC values are recalculated for the third round. The total change score of event I remains nine, and it is selected as the highest scoring event; its inferred early movement (from its negative TRC score) explains the remaining three event-pair changes. The final solution is given in Table 5c.

By itself, a search based on total change scores will yield near-optimal solution(s) and so often constitutes an effective search strategy. However, to explore the effects of arbitrary choices between events with equal scores (and of differing directions of movement in events with incoherent event-pairs), the process should ideally be repeated several times. In practice, we have found that 100 repetitions invariably recover at least one of the equally most parsimonious solutions.

*Branch-and-bound-like near-thorough searches.*—Although the total change heuristic will quickly find a very good solution, it has the potential to become caught in suboptimal "islands" of solutions. In our example, the repetitions will always select event A first because it has the highest total change score. But, if the optimal solution can only be found by starting with another event, it will never be recovered. However, the total change heuristic can be used to establish a threshold value for subsequent, more thorough searches (the equivalent of a "branch-and-bound" phylogenetic search; Kitching et al., 1998). Because assessing an input order is a stepwise process, with each round adding one or more ad hoc hypothesis of event movement, input orders can be abandoned when it is clear that they will not provide a good solution. For example, the input order in Table 5b could be aborted after the selection of event D because all the event-pair changes have not been explained and the number of events hypothesized to be moving is already higher than the threshold of three determined by the total change heuristic (Table 5c). Any further

rounds can only lead to additional ad hoc hypotheses and therefore a suboptimal solution. However, if the more thorough search finds a more optimal solution, the threshold is reset to the new value for the analysis of any additional input orders.

Even with the use of a branch-and-bound-like technique, the number of input orders can still be too large to allow complete searches in a reasonable amount of time (i.e., in less than a day). In these cases, the threshold can be used in conjunction with a near-thorough method that examines all possible input orders of the first few events only. If event-pair changes remain after these initial rounds, the remainder of the input order is determined using the mechanism in the total change heuristic. Because the effects of starting the input with even medium and low-scoring events are examined, the strategy is less likely to become stuck in locally optimal islands of solutions. Even for large data sets, it is possible to examine every possible starting combination if only the input orders of the first two or three events are set (Table 6). When used together, the threshold and near-thorough searches provide an effective heuristic for assessing the many possible solutions of a large problem in a reasonable amount of time.

### *Step 4—Consensus Solutions*

If the search recovers only one most parsimonious solution, this can be adopted as the hypothesis of event movements that occurred to produce the observed event-pair changes. If more than one most parsimonious solution is recovered, a consensus solution can be determined. For example, Table 7 shows the five equally most parsimonious solutions for the data in Figure 2 and Tables 1 and 2. The consensus lists all the changes common to all solutions. In our example, the solutions are very similar, and the consensus retains a lot of information. In effect, it gives the fewest number of steps needed to transform the ancestral sequence into the descendant sequence (cf. Fig. 2A). It is possible that the most parsimonious solutions are mutually incompatible, such that the consensus retains no information. However, we have found this to be rare in practice.

TABLE 7. Five equally most parsimonious solutions for the data from Figure 2 and their strict consensus. The conflict between the solutions (underlined) concerns which other events the "shifting events" moved relative to. Note that the conflicts only concern movements relative to other shifting events (e.g., event A relative to event G, etc.). The consensus therefore retains a lot of information—it is, in fact, the same as solution 2. Compare with Figure 2A.

| | |
|---|---|
| 1 | A moved *late* relative to B, C, D, E, G, I<br>G moved *early* relative to D, E, F<br>I moved *early* relative to E, F, H |
| 2 | A moved *late* relative to B, C, D, E<br>G moved *early* relative to D, E, F<br>I moved *early* relative to E, F, H |
| 3 | A moved *late* relative to B, C, D, E, I<br>G moved *early* relative to D, E, F, A<br>I moved *early* relative to E, F, H |
| 4 | A moved *late* relative to B, C, D, E, G<br>G moved *early* relative to D, E, F<br>I moved *early* relative to E, F, H, A |
| 5 | A moved *late* relative to B, C, D, E<br>G moved *early* relative to D, E, F, A<br>I moved *early* relative to E, F, H, A |
| Consensus | A moved *late* relative to B, C, D, E<br>G moved *early* relative to D, E, F<br>I moved *early* relative to E, F, H |

### THE PARSIMOV PROGRAM

One of us (JEJ) has written a Perl script that will read in the output log from PAUP* (Swofford, 2002) containing apomorphy lists (obtained using the "describetrees apolist = yes;" command under the Trees menu) and subjected the data to the method described above. It outputs a text file for each node, plus a summary of the results for the whole tree. This program can be downloaded from http://www.tierzucht.tum.de/Bininda-Emonds/ or can be obtained from either of the first two authors on request.

## DISCUSSION

### *Comparisons with Previous Methods*

In her analysis of the development of the craniofacial region of therian mammals, Smith (1997) mapped event-pairs involving 28 developmental events onto a reference phylogeny (Fig. 5) to highlight differences between metatherians (marsupials) and eutherians (placentals). This revealed 56 event pairs that differed in event-pair score between the two clades (Table 8A). We previously discussed the limitations of this and later, more statistical methods (Nunn and Smith, 1998)—namely that they could not localize changes to particular branches of the tree, nor determine the direction of the changes (i.e., whether an event had moved earlier in one clade or moved later in the other)—and reanalyzed the data set using the cracking method (Bininda-Emonds et al., 2002; Jeffery et al. 2002b). A consensus of ACCTRAN

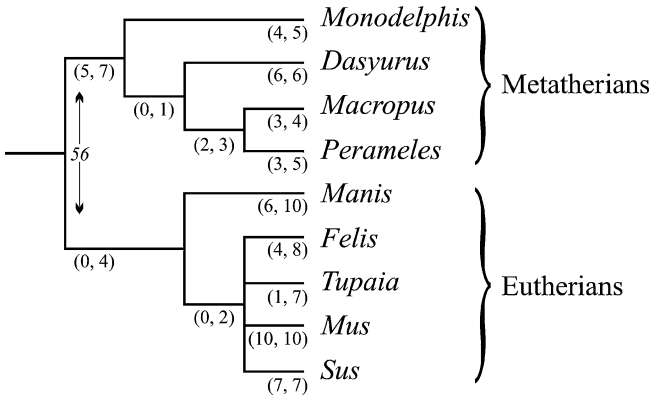



FIGURE 5. Cladogram of nine therian mammals used by Smith (1997). Numbers beneath each branch show the number of changes identified in a consensus of ACCTRAN and DELTRAN optimizations by the cracking method of Jeffery et al. (2002b) and the Parsimov method, respectively. Number in italics at the base of the tree is the number of event-pair changes separating eutherian and metatherian mammals as identified by Smith (1997).

TABLE 8. Comparison of three methods for reconstructing heterochronic changes from event-paired developmental data. (A) Smith (1997) mapped a data set derived from 28 cranial developmental events for nine species of mammal (Fig. 5) but was unable to localize changes to a particular node or to polarize them (i.e., as early in one clade or late in the other). She was also unable to estimate movements underlying changes. (B) Jeffery et al. (2002b) reanalyzed the data of Smith (1997) using their cracking technique. A consensus of the results from ACCTRAN and DELTRAN optimizations produced five hypotheses of event movement along the branch leading to the metatherians only. Eleven of the relative movements identified were among the 56 event-pairs highlighted by Smith (1997). A consensus of ACCTRAN and DELTRAN optimizations using the Parsimov method identified the same five event movements along the branch leading to the metatherians plus two additional movements. (C) The Parsimov method also uniquely identified four movements along the branch leading to the eutherians.

**A. Branch connecting common ancestors of Metatheria and Eutheria**

| Method (author) | Change |
|---|---|
| Character mapping (Smith, 1997) | 28 event-pair scores that differ consistently between all species in each of the two clades highlighted. A further 28 event-pair scores that differ between all species in each of the two clades with a single exception |

**B. Branch leading from root to common ancestor of Metatheria**

| Method (author) | Event | Moves | Relative to . . . |
|---|---|---|---|
| Cracking (Jeffery et al., 2002b) | First ossification of maxilla | Early | First appearance of tooth buds, differentiation of ear ossicles |
| | Evagination of telencephalon | Late | Cartilage in basioccipital region, first alignment of (tongue) myoblasts |
| | Layering in cortex | Late | First ossification of squamosal, craniofacial muscles distinguishable |
| | Swelling of thalamus and hypothalamus | Late | First ossification of frontal, first appearance of tooth buds, differentiation of ear ossicles, craniofacial muscles distinguishable |
| | Lens vesicle filled | Late | First ossification of frontal, differentiation of ear ossicles |
| Parsimov method | First ossification of maxilla | Early | Differentiation of ear ossicles |
| | Evagination of telencephalon | Late | Cartilage in basioccipital region, first alignment of (tongue) myoblasts |
| | Differentiation of retinal pigment | Late | Cartilage in basioccipital region |
| | Olfactory nerve connects with epithelium | Late | Differentiation of ear ossicles, differentiation of retinal pigment |
| | Layering in cortex | Late | First ossification of jugal, first ossification of squamosal, first appearance of muscle striations, craniofacial muscles distinguishable |
| | Swelling of thalamus and hypothalamus | Late | First ossification of frontal, first appearance of tooth buds, differentiation of ear ossicles, craniofacial muscles distinguishable, first appearance of muscle striations, olfactory nerve connects with epithelium |
| | Lens vesicle filled | Late | First ossification of frontal, differentiation of ear ossicles, first appearance of muscle striations |

**C. Branch leading from root to common ancestor of Eutheria**

| Method (author) | Event | Moves | Relative to . . . |
|---|---|---|---|
| Cracking (Jeffery et al., 2002b) | None detected | | |
| Parsimov method | First ossification of Jugal | Early | Secondary palate closes, craniofacial muscles distinguishable |
| | First ossification of Exoccipital | Late | First ossification of frontal, first ossification of basioccipital |
| | Membrane bones approach midline | Early | First ossification of basisphenoid, differentiation of condylar cartilage |
| | Twins (sensu Jeffery et al., 2002b): Swelling of thalamus and hypothalamus early relative to lens vesicle filled | | |

and DELTRAN optimizations of the cracked event-pair data showed five event movements that could account for most of the event-pair changes on the branch leading from the root to the common ancestor of metatherians (Jeffery et al., 2002b). The movements of these events were concordant with the event-pair changes identified by Smith (1997), but were more precise and informative in that they indicated the direction of movements and could be localized to a specific branch (see Table 8B; also Jeffery et al., 2002b). The consensus of ACCTRAN and DELTRAN optimizations also inferred 44 further event movements, localized along other branches of the phylogeny that were not examined by Smith (1997) (see Fig. 5).

We have again reanalyzed Smith's (1997) data using the Parsimov method described above. For each branch, we took a strict consensus of all the most parsimonious solutions. We compared these results with those produced by the cracking method. For reasons of space, we shall only discuss the comparisons of the consensus of ACCTRAN and DELTRAN optimizations. However, the conclusions are the same when each optimization is compared separately. Along every branch, the Parsimov method invariably found an equal or greater number of movements than cracking (a total of 89) because the cracking method can leave certain event-pair changes unexplained if they do not reveal a clear pattern of event movement that exceeds the threshold level. In fact, with

eight exceptions, the movements found by cracking were a subset of those found by the Parsimov method (for an example, see Table 8B). At some nodes, Parsimov found several changes where cracking found none (e.g., Table 8C). In the few cases where cracking identified movements not shown by the Parsimov method (e.g., the movement of the "First ossification of maxilla" early with respect to the "First appearance of tooth buds"; Table 8B), examination of the individual equally most parsimonious solutions found by the Parsimov method showed that, in each case, the cracking process had identified a movement that was only one of several equally parsimonious ways of explaining the event-pair changes. Thus, the movement would not be shown on a strict consensus of all the equally most parsimonious solutions from the Parsimov method. It also means that the hypothesis of movement from the cracking analysis was not robust because equally parsimonious alternatives to it exist.

The cracking method was also used by Jeffery et al. (2002a) to analyze heterochrony in the organogenetic period of amniote embryonic development. We reanalyzed these data using Parsimov and found that, as with the mammalian craniofacial data, the method invariably found an equal or greater number of movements than cracking had. Again, the results of cracking were usually a subset of those found by Parsimov (data not shown).

### *The Cracking and Parsimov Methods for Event-Paired Data*

Both the cracking and Parsimov methods deliver precise hypotheses about the movements underlying observed event-pair changes along a given branch of a tree (a great advantage over the statistical methods). As such, they provide the means to test hypotheses of developmental heterochrony (e.g., Smith, 1997; Cole et al., 2003), or of any change in sequential order (e.g., gene order), in a quantitative, analytical fashion. The Parsimov method delivers more robust hypotheses because it is less subjective in the sense that (1) it determines the minimal solution that accounts for every event-pair change and (2) it yields a consensus that contains all hypotheses of movement that must necessarily form part of any equally most parsimonious solution to the observed event-pair changes. As mentioned, the Parsimov method assumes the independence of events and will be misled by events forming linked complexes. However, the complementary use of the cracking method in conjunction with a low threshold value might be able to discover groups of events moving in parallel, thereby forming putative complexes. This information could then be used to direct a more robust Parsimov analysis.

The only disadvantage of the Parsimov method when compared to the cracking method is speed; using a desktop PC (with a 550-MHz processor) the cracking method took just over one second to analyze both optimizations of the Smith (1997) data set, whereas the Parsimov method took just under 15 hours, even using the heuristic options described above! However, even 15 hours is not an unreasonable time to wait for an analysis. Moreover, the continuous increase in desktop processor speed coupled with the possible development of better search heuristics means that this disadvantage will become less significant over time. For instance, the same analyses using either a 2.8-GHz Pentium IV Dell (with 512 Mb RAM) or a dual 2-GHz processor Mac G5 (with 1 Gb RAM) took under 7 hours. These increasing speed developments will also allow for more thorough searches, thereby improving the robustness of the result even further.

For both methods, the recovered solution forms the jumping-off point for an interpretation of the biological causes and implication of the timing shifts. For example, with Smith's (1997) data set, the inference of specific event movements can be used to examine questions such as how the particular craniofacial timing shifts relate to the very different reproductive strategies of eutherian and metatherian mammals, at what point in the developmental program the timing differences originate, and what the underlying genetic causes are.

## ACKNOWLEDGEMENTS

We thank Ariel Chipman, Kathleen Smith, and Gregory Wray for many helpful comments on an earlier draft of this paper. Support was provided by the Leverhulme Trust (JEJ), the van der Leeuw fonds (JEJ, OBE, MKR), and by the BMBF-funded project "Bioinformatics for the Functional Analysis of Mammalian Genomes (BFAM)" (OBE).

## REFERENCES

Bininda-Emonds, O. R. P., J. E. Jeffery, M. I. Coates, and M. K. Richardson. 2002. From Haeckel to event-pairing: The evolution of developmental sequences. Theory Biosci. 121:297–320.

Bininda-Emonds, O. R. P., J. E. Jeffery, and M. K. Richardson. 2003. Is sequence heterochrony an important evolutionary mechanism in mammals? J. Mamm. Evol. 10:335–361.

Chipman, A. D. 2002. Variation, plasticity and modularity in anuran development. Zoology 105:97–104.

Chipman, A. D., A. Haas, E. Tchernov, and O. Khaner. 2000. Variation in anuran embryogenesis: Differences in sequence and timing of early developmental events. J. Exp. Zool. B 288:352–365.

Cole, N. J., M. Tanaka, A. Prescott, and C. Tickle. 2003. Expression of limb initiation genes and clues to the morphological diversification of threespine stickleback. Curr. Biol. 13:R951–R952.

Galis, F., and J. A. J. Metz. 2001. Testing the vulnerability of the phylotypic stage: On modularity and evolutionary conservation. J. Exp. Zool. 291:195–204.

Jeffery, J. E., O. R. P. Bininda-Emonds, M. I. Coates, and M. K. Richardson. 2002a. Analysing evolutionary patterns in amniote embryonic development. Evol. Dev. 4:292–302.

Jeffery, J. E., M. K. Richardson, M. I. Coates, and O. R. P. Bininda-Emonds. 2002b. Analyzing developmental sequences within a phylogenetic framework. Syst. Biol. 51:478–491.

Kitching, I. J., P. L. Forey, C. J. Humphries, and D. M. Williams. 1998. Cladistics. 2nd edition. Oxford University Press, Oxford.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Mabee, P. M., K. L. Olmstead, and C. C. Cubbage. 2000. An experimental study of intraspecific variation, developmental timing, and heterochrony in fishes. Evolution 54:2091–2106.

Mabee, P. M., and T. A. Trendler. 1996. Development of the cranium and paired fins in *Betta splendens* (Teleosti: Percomporpha): Intraspecific variation and interspecific comparisons. J. Morphol. 227:249–287.

Nunn, C. L., and K. K. Smith. 1998. Statistical analyses of developmental sequences: The craniofacial region in marsupial and placental mammals. Am. Nat. 152:82–101.

Sánchez-Villagra, M. R. 2002. Comparative patterns of postcranial ontogeny in therian mammals: An analysis of relative timing of ossification events. J. Exp. Zool. B 294:264–273.
Schlosser, G. 2001. Using heterochrony plots to detect the dissociated coevolution of characters. J. Exp. Zool. B 291:282–304.
Schulmeister, S., and W. C. Wheeler. Comparative and phylogenetic analysis of developmental sequences. Evol. Dev. 6:50–57.
Smith, K. K. 1996. Integration of craniofacial structures during development in mammals. Am. Zool. 36:70–79.
Smith, K. K. 1997. Comparative patterns of craniofacial development in eutherian and metatherian mammals. Evolution 51:1663–1678.
Smith, K. K. 2001a. Early development of the neural plate, neural crest and facial region of marsupials. J. Anat. 199:121–131.
Smith, K. K. 2001b. Heterochrony revisited: The evolution of developmental sequences. Biol. J. Linn. Soc. 73:169–186.
Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Velhagen, W. A. 1997. Analyzing developmental sequences using sequence units. Syst. Biol. 46:204–210.
Velhagen, W. A., and A. H. Savitzky. 1998. Evolution of embryonic growth in thamnophiine snakes. Copeia 1998:549–558.
Wagner, G. P. 1996. Homologues, natural kinds and the evolution of modularity. Am. Zool. 36:36–43.

*First submitted 22 August 2003; reviews returned 8 February 2004; final acceptance 15 November 2004*
*Associate Editor: Greg Wray*
*Editor: Chris Simon*