*Syst. Biol.* 51(3):478–491, 2002
DOI: 10.1080/10635150290069904

# Analyzing Developmental Sequences Within a Phylogenetic Framework

JONATHAN E. JEFFERY,[1,3] MICHAEL K. RICHARDSON,[1,3] MICHAEL I. COATES,[2] AND OLAF R. P. BININDA-EMONDS[3,4]

[1]*Department of Anatomy and Developmental Biology, St. George's Hospital Medical School, Tooting, London SW17 ORE, UK*
[2]*Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois, 60637–1508, USA*
[3]*Institute of Evolutionary and Ecological Sciences, Leiden University, 2300 RA Leiden, The Netherlands; E-mail: bininda@rulsfb.leidenuniv.nl*

*Abstract.*— Heterochrony is important as a potential mechanism of evolutionary change. However, the analysis of developmental timing data within a phylogenetic framework to identify important shifts has proven difficult. In particular, analytical problems with sequence (event) heterochrony revolve around the lack of an absolute time frame in development to allow standardization of timing data across species. An important breakthrough in this regard is the method of "event-pairing," which compares the relative timing of developmental events in a pairwise fashion. The resulting event-pair-encoded data can be mapped onto a phylogeny, which can provide important biological information. However, event-paired data are cumbersome to work with and lack a rigorous quantitative framework under which to analyze them. Critically, the otherwise advantageous relativity of event-pairing prevents an assessment of whether one or both events in a single event-pair have changed position during evolutionary history. Building on the method of event-pairing, we describe a protocol whereby event-pair transformations along a given branch are analyzed en bloc. Our method of "event-pair cracking" thereby allows developmental timing data to be analyzed quantitatively within a phylogenetic framework to infer key heterochronic shifts. We demonstrate the utility of event-pair cracking through a worked example and show how it provides a set of desired features identified by previous authors. [Character mapping; development; event-pairing; evolutionary transformations; heterochrony; phylogenetic framework.]

The evolution of developmental mechanisms is a key field of research in modern biology (Holland, 1999; Raff, 2000). Change in developmental timing (heterochrony) is often cited as a potential mechanism of evolutionary change (Gould, 1977, 1982; McKinney and McNamara, 1991; Raff, 1996) and usually is discussed in the context of large-scale timing changes that affect the entire organism (e.g., as formalized as the processes of neotony or paedomorphosis). However, the full extent of heterochrony as an evolutionary mechanism may be underestimated, particularly with respect to smaller-scale heterochronies affecting single organs. This potential underestimation is related in part to the methodological problems in analyzing heterochrony under a phylogenetic framework (Alberch, 1985; Mabee and Trendler, 1996; Mabee et al., 2000).

Recently, there has been renewed interest in the analysis of developmental sequences, or the order in which events occur during development (for an excellent summary, see Smith, 2001). This interest derives from two sources. First, the debate concerning the existence of a conserved, "phylotypic" stage in vertebrate development depends critically on how much developmental events change their times of appearance (sequence or event heterochrony) during this period (reviewed in Richardson, 1995, 1999; Smith, 2001). Second, the development of "event-pairing" (Mabee and Trendler, 1996; Smith, 1996), which compares the relative timing of developmental events in a pairwise fashion, has revolutionized the analysis of developmental sequences. Using this method, several studies have demonstrated sequence heterochrony by optimizing event-paired data onto existing phylogenetic trees (e.g., Smith, 1996, 1997; Velhagen, 1997; Jeffrey et al., in press).

In this paper, we examine the issue of analyzing developmental sequences within a phylogenetic framework to elucidate instances of sequence heterochrony. In particular, we review general analytical problems inherent to heterochronic data and the advantages that event-pairing provides in comparison with other methods. Although event-pairing represents an

[4]Corresponding author.

important breakthrough, its full potential has not been realized because of the lack of a rigorous quantitative framework under which to analyze event-paired data (Nunn and Smith, 1998). We therefore introduce a protocol for full interpretation of event-paired data in a phylogenetic context to reveal patterns of heterochrony. Using our procedure of "event-pair cracking," timing shifts of individual developmental events can now be identified precisely and localized on a tree. We demonstrate this with a worked example based on published data. Our protocol allows the rigorous examination of hypotheses of heterochrony and can be applied to other instances in which event-pairing can be useful (e.g., analyses of changes in gene order).

## Measuring the Progress of Development

Any analysis of developmental sequences requires that the data be quantified so as to allow comparison between species. However, two factors make this difficult. First, the very existence of heterochrony means that homologous structures may occur at different times in different species. Second, there is no absolute time frame for cross- species comparisons because the rate of development varies between species (Hall and Miyake, 1995; Smith, 2001). Therefore, a crucial methodological issue is to find a common metric for the progress of development (Richardson, 2000). We briefly review the issues in this area as a prelude to showing how event-pairing overcomes many of the analytical problems inherent to developmental sequence data. Additional detail and arguments are given in Bininda-Emonds et al. (in press).

The most obvious metrics of developmental progress are chronological age and size measurements. However, factors such as temperature, nutrition, and intraspecific (genetic) variation all affect the rate of development (Hall and Miyake, 1995). Thus, age and size usually correlate poorly with morphological maturity (e.g., as noted for the Rhesus monkey, *Macaca mulatta*; Gribnau and Geijsberts, 1981). This problem is compounded by the lack of a universal time frame, necessitating elaborate normalization of the timing data (e.g., Dettlaff and Dettlaff, 1961). However, this normalization is itself hindered by the fact that the total developmental time span can differ widely even in closely related organisms and by the lack of homologous endpoints for calibration.

To circumvent the problems with chronological age, comparisons of developmental maturity now tend to rely on broad "stages" based on the occurrence of one or more key (morphological) events. However, the selection of any particular event or events as the key definition of a stage is difficult to justify on purely biological grounds. Although arbitrary, stages are useful for standardizing studies of a single species, especially when based on a suite of events (Hall and Miyake, 1995). However, stages are unlikely to provide a common metric of developmental maturity across species because of the potential for heterochrony, wherein the key events shift their timing relative to the development of other structures (Richardson, 1995).

## Desirable Features of a Method and Event-Pairing

Ideally, any method that uses developmental timing data to infer instances of sequence heterochrony should possess the following three features. First, the method should use the *relative* timing of developmental events to overcome any problems associated with chronological age or the lack of an absolute time frame in development. Second, the developmental timing data should be analyzable within a phylogenetic framework to distinguish shared primitive from shared derived similarity (Mabee and Trendler, 1996). Third, the method should be amenable to a rigorous quantitative framework under which to infer significant timing shifts (Nunn and Smith, 1998).

The first two of these features can be realized immediately through the pairwise comparison of individual events in a developmental sequence. As formalized by Smith (1996, 1997), event-pairing involves constructing a matrix derived from the developmental sequence of each species under consideration (see Alberch [1985] for a discussion of developmental sequences). Each developmental event is scored according to whether it occurred before (0), simultaneously with (1), or after (2) each of the remaining events; missing data are included as "?" (Fig. 1; Tables 1 and 2). The application of event-pairing was a breakthrough in the study of sequence heterochronies because the resulting, discrete scores are both

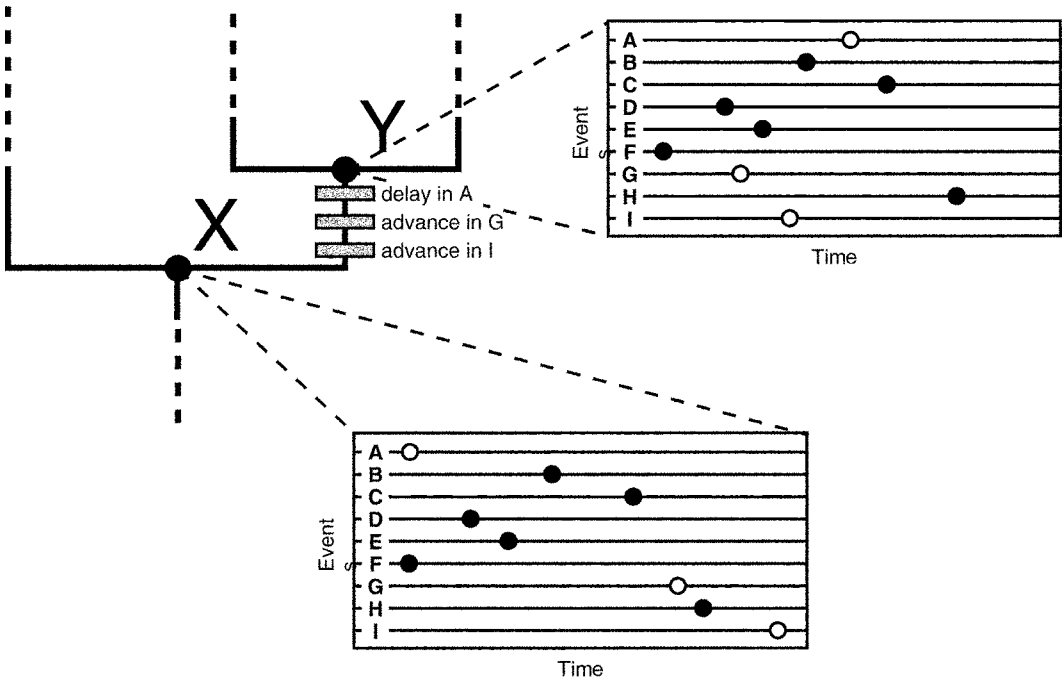


FIGURE 1. Hypothetical example showing changes in developmental sequence between nodes X and Y on a partial phylogeny. As displayed in the form of "abacus diagrams" (Richardson, 1995, 2000), only events A (delay), G, and I (both advances) have changed their absolute developmental timing.

a stage- and time-independent description of development and suitable for analysis in a phylogenetic context. Event-paired heterochronic data have been used to examine the evolution of craniofacial and central nervous system development in marsupial versus placental mammals (Smith, 1996, 1997), the ossification sequence of five cranial bones in six species of thamnophiine snakes (Velhagen, 1997), and heterochronic changes during tetrapod evolution (Jeffrey et al., in press).

However, the ability to interpret what synapomorphies of event-paired data mean in the context of heterochronic changes to the developmental sequence (our third desired feature) has remained elusive. Mapping event-paired data onto an independently established phylogeny (e.g., Smith, 1996, 1997; Velhagen, 1997) highlights changes in the relative timing relationship of pairs of developmental events only. Without further information, one cannot determine whether one event or both have moved, nor discern their direction of movement (Fig. 2). Further, the many changes in entire developmental sequences make it difficult to distinguish events that have actively moved from those that have only apparently done so (e.g., events B and C in Fig. 1) because other events

TABLE 1. Ranked developmental sequences for the nine developmental events in Figure 1.

| | Event | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| Node X | 1 | 4 | 5 | 2 | 3 | 1 | 6 | 7 | 8 |
| Node Y | 7 | 6 | 8 | 2 | 4 | 1 | 3 | 9 | 5 |

TABLE 2. The event-pair matrix for the timing data for node X in Table 1. Standard coding as formalized by Smith (1997) was used: 0 = row-event occurs before column-event, 1 = row and column-events occur simultaneously, 2 = row-event occurs after column-event. Character numbers are obtained by reading across rows.

| | Column-event | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Row-event | A | B | C | D | E | F | G | H |
| B | 2 | | | | | | | |
| C | 2 | 2 | | | | | | |
| D | 2 | 0 | 0 | | | | | |
| E | 2 | 0 | 0 | 2 | | | | |
| F | 1 | 0 | 0 | 0 | 0 | | | |
| G | 2 | 2 | 2 | 2 | 2 | 2 | | |
| H | 2 | 2 | 2 | 2 | 2 | 2 | 2 | |
| I | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

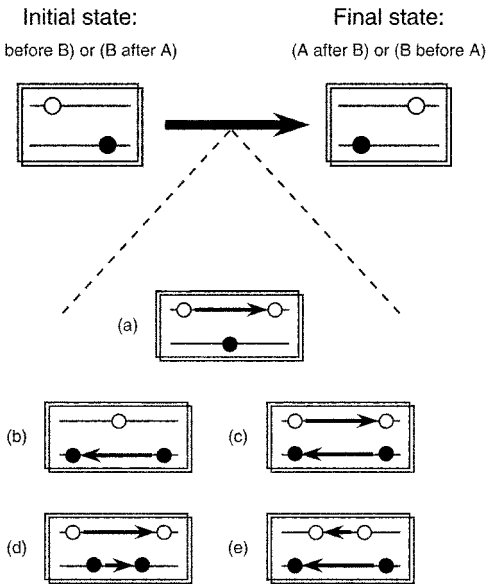


FIGURE 2. Relative timing of event-pairs. The events pictured (A, open circles; B, filled circles) show a change in their relative timing. However, this change could be achieved in five ways: (a) A moves late, B remains static; (b) A remains static, B moves early; (c) A moves late, B moves early; (d) B moves late, A moves even later; and (e) A moves early, B moves even earlier. Without further information, it is impossible to distinguish between these alternatives.

have actively changed their timing relationship with them.

For example, Smith (1996, 1997) restricted her analyses to identifying developmental events that distinguish marsupial and placental mammals. She arrived at two equally plausible scenarios. One was that marsupials were characterized by advanced development of the cranial skeleton and musculature compared with that of structures of the central nervous system. The other was the reciprocal, namely, that placentals were characterized by advanced development of the central nervous system relative to the craniofacial region. She could not conclusively determine whether the heterochronic shift or shifts producing this pattern (1) occurred in marsupials or placentals or both, and (2) involved one system, the other, or both. The addition of an outgroup taxon would address the first problem by localizing where the changes occurred on the tree (Smith, 1996). However, solving the second problem would require a method to derive the timing shift of an event from any inferred evolutionary transformations in the event-pairs (= "event-pair synapomorphies") to which it contributes. Such a methodology has been lacking.

## A Quantitative Framework for Analysis of Developmental Sequences

In studying sequence heterochrony, one must distinguish between an event and the relative timing of it. The developmental events are not the features directly under study. Rather, they form the substrate for the real character set, namely, the heterochrony in those events. In this, we make the assumption that both the events and their relative timing in development are heritable, homologous features of organisms and therefore good phylogenetic characters (Wiley, 1981; Grandcolas et al., 2001). The situation is complicated by the fact that although we are most interested in the timing shifts of single events, it is the event-paired data that form the basis of the analysis. To avoid confusion, we follow Smith (1997) in using "character" in a practical sense only to refer to the individual elements analyzed with reference to a phylogeny (= event-pairs here). A glossary of other terms used in this paper is given in Table 3.

We have found that the goal of deriving the timing shifts of single events can be achieved through the en bloc comparison of all event-pair synapomorphies along a given branch in a phylogeny. In describing our protocol for this, which we refer to as event-pair cracking, we make reference to the simple hypothetical data in Table 1 and Figure 1. For simplicity, our hypothetical example uses standard (012) event-pair coding, weighting all transformations equally. However, it is easily adaptable to other coding schemes and assumptions about character transformations (e.g., ordered characters, Sankoff step-matrices).

### *Identifying Significant Movement*

Event-pair cracking is based on the principle that unless the shifts are very small, events that have moved during evolution will have changed their timing relationships with several other events. Thus, our method operates generally by identifying events that occur in an above-average number of event-pair synapomorphies for a given branch. The direction and magnitude of the shift

TABLE 3. Definitions of terms as used in this study.

| Term | Definition |
|---|---|
| Advance | A timing shift of a given event to an earlier position in the developmental sequence |
| Character | The element that is mapped onto the phylogenetic tree (following Smith, 1997); herein, an event-pair |
| Column-event | Events found along the columns of an event-pair matrix (e.g., Table 2); by convention, they are what the positions of event-pairs are compared to |
| Delay | A timing shift of a given event to a later position in the developmental sequence |
| Developmental event ("event") | Any heritable morphological, molecular, or physiological transformation during ontogeny |
| Developmental sequence | A set of events in the ontogeny of an organism listed in chronological order |
| Event-pair | A summary of the relative timing between any two developmental events (= "sequence unit" of Velhagen, 1997); by convention, it refers to the position of the row-event relative to that of the column-event |
| Event-pair synapomorphy | The character state transformation in an event-pair reconstructed along a branch; it represents a change in relative timing between two events |
| Hitchhiker | An event that has only apparently moved because it is well-positioned in relation to several actively moving events; it shows no movement relative to nonmoving events |
| Row-event | Events found along the rows of an event-pair matrix (e.g., Table 2); by convention, event-pairs and event-pair cracking are taken from the perspective of the row-events |
| TAC | Total absolute change of an event inferred from event-pair cracking; it represents the maximal change possible for the event given the data |
| TRC | Total relative change of an event inferred from event-pair cracking; it represents the actual change for an event, including information about polarity ("direction") |
| Timing shift | Shorthand summary of the heterochronic movement of any single event as inferred from the event-pair synapomorphies in which it is involved |
| Twin | An event that is inferred as having actively moved, but only in relation to a single nonmoving event; a special kind of hitchhiker |

for these events (i.e., whether the event has moved earlier or later and how many other events it has moved relative to, respectively) is then determined. The procedure is slightly conservative. Shifts will be identified only if they are relatively large and coherent (although this can be adjusted) and if the number of event-pair synapomorphies is sufficient to establish a background for comparison.

Event-pair cracking is a three-step process. In the first step, the total relative change for each event along a given branch is calculated. In the second step, events are filtered to retain only those showing the largest relative change. In the third step, the relative changes of selected events from step 2 are corrected for changes involving other selected events. Altogether, these steps serve to distinguish actively moving events from those that are only apparently moving. We now describe each step in greater detail.

*Step 1: Determining relative change.*—First, the magnitude and polarity of the change associated with each event-pair synapomorphy are determined from the transformation in character state (Table 4). The magnitude is the cost of the transformation according to the weighting scheme used. For polarity, decreases in the character state (i.e., from 2 to 0 or 1, or from 1 to 0, reflecting earlier shifts of the row-event relative to the column-event) are defined as negative. Increases in state, reflecting later shifts, are defined as positive. In other words, the "relative change" associated with a synapomorphy is obtained by subtracting the primitive state of the event-pair transformation from the derived state and applying any weighting schemes. The absolute change associated with a synapomorphy is simply the cost of the transformation (i.e., magnitude), regardless of polarity. In Table 4, all synapomorphies involve decreases in the character state. Therefore, because all transformations are equally weighted, all relative changes are −1. Similarly, all absolute changes are 1.

Next, the constituent events of each event-pair synapomorphy are determined (Table 4). This serves as a prelude to calculating the total relative change (TRC) in each event across all synapomorphies for a given

TABLE 4. Determining the polarity and magnitude of the change associated with each event-pair synapomorphy reconstructed for the hypothetical example in Figure 1. Transformations to a lower character state (which represent the row-event shifting earlier in the sequence relative to the column-event) are arbitrarily given a negative sign; transformations to a higher state are given a positive sign. All transformations are assumed to be of equal weight in this example. The constituent events of each character are made with reference to the event-pair matrix in Table 2, with characters being numbered going across rows.

| Character | Character state transformation | Relative change | Constituent events | |
|---|---|---|---|---|
| | | | Row | Column |
| 1 | 2 → 0 | −1 | B | A |
| 4 | 2 → 0 | −1 | D | A |
| 7 | 2 → 0 | −1 | E | A |
| 11 | 1 → 0 | −1 | F | A |
| 16 | 2 → 0 | −1 | G | A |
| 17 | 2 → 0 | −1 | G | B |
| 18 | 2 → 0 | −1 | G | C |
| 20 | 2 → 0 | −1 | G | E |
| 29 | 2 → 0 | −1 | I | A |
| 30 | 2 → 1 | −1 | I | B |
| 31 | 2 → 0 | −1 | I | C |
| 36 | 2 → 0 | −1 | I | H |

branch. First, the sum of relative changes of which each event is a part is calculated. This is initially done separately for cases when the event is a row-event versus a column-event in an event-pair synapomorphy (Table 5). The TRC of an event is then calculated as its summed row-event relative change minus its summed column-event relative change. This standardizes TRC from a single perspective, that of an event being a row-event. For instance, event B occurs in three synapomorphies in Table 4, once as a row-event (character 1) and twice as a column-event (characters 17 and 30). The relative change for each synapomorphy is −1. Therefore, event B has a TRC of 1, which is equal to its relative change of −1 as a row-event minus its relative change of −2 (i.e., −1 + −1) as a column-event. Similarly, the total absolute change (TAC) for each event is determined, which is simply the weighted number of transformations an event was involved in.

The sign of the TRC indicates the overall direction in which the event is inferred to have moved—negative for early, positive for later. A TRC of 0 indicates the event shows no net change along the branch being examined. One reason for a TRC of 0 is that the event did not move in relation to any other events (i.e., did not appear in any synapomorphic event-pairs), in which case the event will also have a TAC of 0. Another reason would be that the event was involved in an equal number of early and late shifts (e.g., event E, with a TAC of 2) and therefore is only apparently moving in relation to events that actively have moved.

TABLE 5. Determining the total relative change (TRC) of each event in the hypothetical example in Figure 1. The TRC of an event is obtained by subtracting its total relative change as a column-event from that as a row-event (see Table 4). All values account for any differential weighting of character state transformations (not employed in this example). The TRC of an event may be less than its total absolute change (TAC; e.g., events B and E), indicating that all inferred movements have not been in the same direction.

| Event | Relative change | | | Absolute change (TAC) |
|---|---|---|---|---|
| | As row-event | As column-event | Total (TRC) | |
| A | 0 | −6 | 6 | 6 |
| B | −1 | −2 | 1 | 3 |
| C | 0 | −2 | 2 | 2 |
| D | −1 | 0 | −1 | 1 |
| E | −1 | −1 | 0 | 2 |
| F | −1 | 0 | −1 | 1 |
| G | −4 | 0 | −4 | 4 |
| H | 0 | −1 | 1 | 1 |
| I | −4 | 0 | −4 | 4 |

TABLE 6. Initial identification of events that have actively moved. The absolute value of the total relative movement (TRC) of selected events is greater than the median of this value across all events (median = 1). Event C is selected despite not having actively moved (see Fig. 1).

| Selected | | Rejected | |
|---|---|---|---|
| Events | Absolute value of TRC | Events | Absolute value of TRC |
| A | 6 | B | 1 |
| C | 2 | D | 1 |
| G | 4 | E | 0 |
| I | 4 | F | 1 |
| | | H | 1 |

*Step 2: Identifying actively moving events.*—Actively moving events will show a large and mostly coherent pattern of change (reflected by a large value for TRC ). In contrast, the remaining events should show little or no change (nonmovers) or largely inconsistent changes (apparent movers); both cases are characterized by TRCs closer to 0. Therefore, step 2 seeks to filter out actively moving events by retaining only those events for which the absolute values of their TRCs are greater than some threshold. In our example (Fig. 1; Table 6), the threshold is the median of the absolute values of the TRCs of all events. This minimizes the influence of any outlier events (i.e., ones that have changed position by a great amount or very little, if at all) and will select as many as 50% of the events, depending on the distribution of TRC scores.

The choice of threshold is admittedly arbitrary. However, in this respect it does not differ from normal statistical practice, where a rejection level of $\alpha = 0.05$ is used *by convention*, rather than for any fundamental biological principle. Moreover, the use of other threshold values allows the investigator to set the stringency of the selection process, similar to adjusting the rejection level in statistical procedures. Potential methods include using other measures of central tendency (e.g., arithmetic or geometric means), transforming the TRC values (e.g., log, square, or square root transformations), or discarding events with a TAC of 0 from the calculation. One could even use a rejection level of 0.05 by selecting only those events having a TRC 1.645 standard deviations greater than the mean of all TRCs. A sensitivity analysis is also possible by using thresholds of increasing stringency (e.g., the mean +0.5, 1.0, 1.5, ... standard deviations).

In our hypothetical example, events A, C, and G are "selected" (Table 6). However, it can be seen from Figure 1 that event C has not actually moved. Instead, C has effectively "hitchhiked" onto the movements of two other events that did move (G and I, which have been successfully identified) by being positioned advantageously at the conjunction of their movements. Note also that event B was not selected despite its being involved in more synapomorphic event-pairs than was event C (three versus two). This resulted because B was involved in three movements that were not coherent, which leads us to suspect that it is only apparently moving in relation to events that actually are moving.

*Step 3: Accounting for other actively moving events.*—The final step identifies events like C that appear to move because of their position in the developmental sequence relative to events that have actively moved. This step operates by recalculating both TRC and TAC of selected events (i.e., those that passed step 2) such that changes involving other selected events are excluded (Table 7). In other words, we wish to characterize the movement of selected events relative to nonselected events only. For instance,

TABLE 7. Correcting for other actively moving events. Total relative change (TRC) and absolute change (TAC) are recalculated by excluding event-pairs involving other selected events (i.e., events A, C, G, and I). Event C is successfully identified as having only apparently moved in relation to events A, G, and I ("hitchhiking").

| | | Relative change excluding other selected events | | Adjusted | | | |
|---|---|---|---|---|---|---|---|
| Event | TRC | As row-event | As column-event | TRC | TAC | *J* | Status |
| A | 6 | 0 | 4 | 4 | 4 | 1 | Selected |
| C | 2 | 0 | 0 | 0 | 0 | [a] | Hitchhiker |
| G | −4 | −2 | 0 | −2 | 2 | −1 | Selected |
| I | −4 | −2 | 0 | −2 | 2 | −1 | Selected |

[a] Undefined.

step 1 determined that event A has a TRC of 6, deriving from synapomorphies involving characters 1, 4, 7, 11, 16, and 29 (Table 4). However, two of these synapomorphies are with other selected events, G (character 16) and I (character 29). When we correct for this, the adjusted TRC of event A decreases to 4 (Table 7). A similar adjustment is made for TAC.

The adjusted TRC and TAC values of an event are necessarily less than or equal to their analogous values from step 1. Hitchhikers like C are identified by having an adjusted TAC of 0; all their movement is apparent and derives solely from other, actively moving events. Events that are moving together as a unit would not be identified as hitchhikers because they would still be showing movement relative to other nonmoving events. Occasionally, adjusted TAC scores will indicate a selected event that has moved relative to one nonselected event only ("twins"). This event should be rejected as a special kind of hitchhiker. Although the twins have moved relative to one another, the selection of one of them derives from its apparent movement relative to other actively moving events. Like the nonselected twin, it would not have been selected without this apparent movement.

Having thus identified which events have actively changed position in the developmental sequence, one can characterize the heterochronic shifts in terms of their direction and magnitude relative to other nonmoving events in the developmental sequence. Direction and magnitude are provided by the adjusted TRC scores from step 3. When character state transformations are not equally weighted, some idea of the magnitude is also provided by the number of nonmoving events the actively moving event has moved relative to.

### *Coherence of Movement*

The coherence of movement for a moving event is quantified by the quotient *J*, which indicates whether all the timing shifts of an event have been in the same direction. It can therefore highlight noise in event movements. $J_x$ is calculated by dividing the adjusted TRC of event X by its adjusted TAC. Because $J$ uses the adjusted change values from step 3, it characterizes movement relative to nonmoving events only. Event-pairs involving other events selected in step 2 are excluded from the calculation.

$J$ has a range of $-1$ to 1, negative values indicating early movement (advances) and positive values indicating later movement (delays). If all the shifts of a particular event are in the same direction, the adjusted TRC equals the adjusted TAC and $J = \pm 1$. Conversely, if the shifts are an equal mixture of early and late, the adjusted TRC will be 0, and $J = 0$. In practice, we have found that actively moving events show high coherence in their movement; $J$ is almost invariably $\pm 1$. Events that are only apparently moving show less coherence, whereas hitchhikers have an undefined $J$ value because their adjusted TAC is 0. Thus, the $J$ quotient can highlight those rare nonmoving events that have erroneously passed both rounds of selection (i.e., steps 2 and 3). Actively moving events are inferred to be those passing both rounds of selection and possessing a $J$ coefficient of $\pm 1$.

### *Reconstructions of Ancestral Developmental Sequences*

As with standard phylogenetic analyses, the ancestral states at any internal node of the tree can be reconstructed by using parsimony. The event-pair scores thus inferred can be decomposed to generate an ancestral developmental sequence (Velhagen, 1997). This is useful both to examine the ontogenetic changes correlated with major evolutionary transitions (e.g., the transition to land in tetrapods) and to serve as the basis of testable hypotheses regarding the developmental sequences of species not included in the analysis.

An initial, consensus sequence is obtained by a method similar to that of Velhagen (1997). Events are given points according to their relative position in each of the ancestral event-pairs reconstructed for an internal node of which they are a part. If an event is the earlier one in the event-pair, it receives $-1$. If it is the later event, it receives $+1$. Events that are simultaneous or are missing from the event-pair receive 0. The points accumulated by each event are then totaled. Ranking the events in ascending order according to the number of points they have produces the ancestral sequence.

However, the developmental sequence so inferred may not be self-consistent in the

sense that specific event-pairs may conflict with the consensus position of its constituent events. For instance, although the event-pair for any two events X and Y indicates that X occurs before Y, the consensus of all event-pairs (i.e., the ancestral sequence) may indicate the reverse. These conflicts can arise because the ancestral sequence is derived indirectly by consideration of event-pairs, rather than from the developmental sequences of the terminal taxa themselves. Homoplasy, equivocal reconstructions, and missing data will probably exacerbate this effect. We find it therefore instructive to calculate the earliest and latest possible positions of each event as an estimate of the variation in its consensus position. To do so, we compare its consensus position in the sequence against that indicated by all its event-pairs individually. For example, say event X is inferred to have a consensus rank of 10 in a hypothetical ancestral sequence. In examining all event-pairs involving X, we find it to be placed consistently with respect to all events of rank 15 or higher (i.e., X occurs before all these events). However, one event-pair states that X occurs after an event (X) with rank 14. Thus, the latest possible position for A is 14.5. Conversely, the earliest possible position for event X, barring other inconsistencies, is 9.5.

The earliest and latest positions can be exaggerated by a few, highly inconsistently placed events that conflict with a large number of other events. We therefore suggest that the positions be recalculated after excluding these highly inconsistent events, which are identified by having more than the median number of inconsistencies for all events. Again, other thresholds are possible, depending on the desired stringency of the analysis. In practice, we think this correction provides a more realistic estimate of the variation in the consensus position of an event. The earliest and latest positions for each event can be indicated simply on an abacus graph through error bars (Fig. 3).

## APPLICATION

We have applied our method elsewhere to reconstruct important shifts in developmental timing in the history of tetrapods (Jeffrey

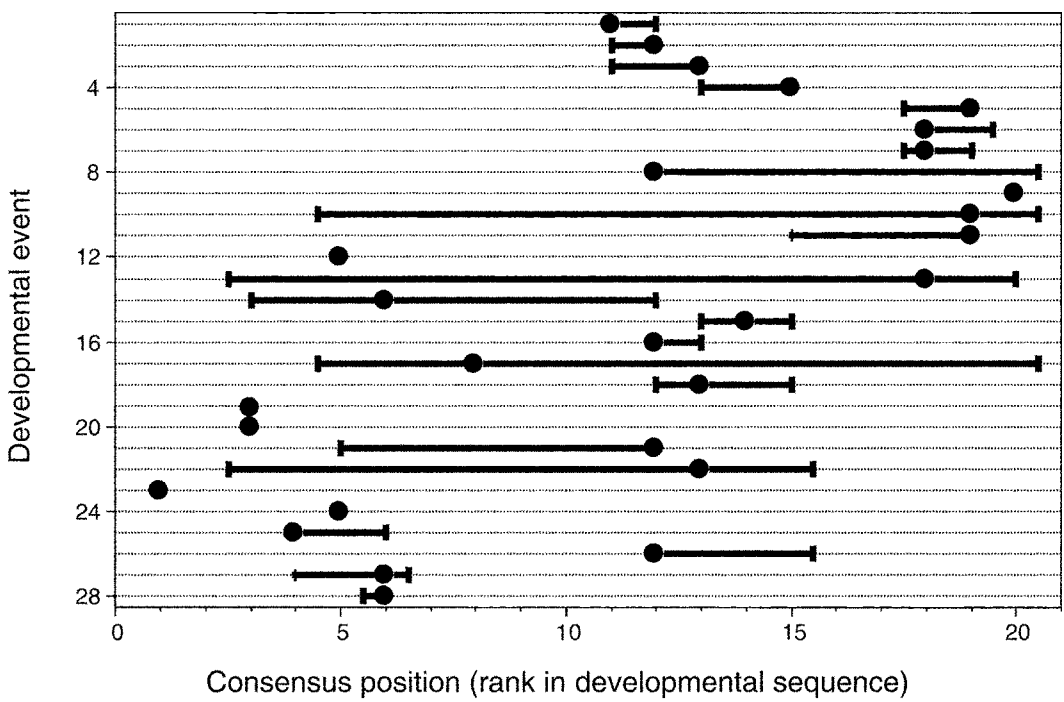


FIGURE 3. The ancestral developmental sequence for Theria (the common ancestor of marsupial and placental mammals), as determined by decomposing the ancestral event-pairs inferred from the phylogeny and timing data of Smith (1996, 1997; data appears in Nunn and Smith, 1998). Reconstructions used ACCTRAN optimization. Because the set of reconstructed event-pairs may not be self-consistent, some events possess multiple equally possible placements (indicated by error bars).

et al., in press). Among other findings, we determined that homeothermy in birds and mammals was not linked to an advance in heart development (contra McCrady, 1938); instead, numerous independent advances in heart development occurred throughout amniotes. We also found that mammals were characterized ancestrally by a delay in several events in eye development, which we hypothesized might relate to the relatively poorly developed visual system of mammals compared with other amniotes.

Here we apply our method to the developmental data of Smith (1996, 1997; as given in Nunn and Smith, 1998) in an attempt to localize the timing differences that she observed between marsupial and placental mammals. In so doing, we have distilled equivalent, homologous information for an outgroup species (the chicken, *Gallus gallus*; Table 8) to polarize the transformations. We acknowledge the problems in using a member of Aves as an outgroup (see Smith, 1997), notably its distance with respect to therian mammals (minimal divergence time of 290 million years before the present [Ma]; Harland et al., 1990; Benton, 1991) and the fact that several of Smith's (1997) events are specific to mammals. In this, Monotremata would represent a better choice of outgroup (Smith, 1997). However, monotreme material is extremely rare and no complete developmental series exists. Furthermore, the long divergence time between monotremes and therians (~150 Ma; Luo et al., 2001) still allows for many derived heterochronic changes to occur along the lineage leading to monotremes. We still feel that the advantages of including an outgroup outweigh the specific disadvantages of using the chicken as the outgroup. Otherwise, we have attempted to duplicate Smith's (1997) analysis wherever possible, including the use of unordered (012) multistate event-pair coding. We focused on changes occurring along the branches that lead to either the marsupial or placental common ancestor.

Our results are highly concordant with those of Smith (1997). We found 12 heterochronic shifts common to all of ACCTRAN, DELTRAN, and MinF character optimizations leading to the marsupial common ancestor; in contrast, there were none for the placental equivalent (Table 9). Of these 12 changes, 11 are also found among the 56 event-pairs identified by Smith (1997) as distinguishing the two therian clades; only the event-pair "layering in cortex is late relative to squamosal ossification" was not identified. Therefore, we suggest that the pattern observed by Smith (1997) in therian mammals results from the movement of only a few developmental events in the lineage leading to marsupials only. These five events consist of an advance in the onset of maxillary ossification and a delay in the onset of several central nervous system events (i.e., evagination of telencephalon, layering in cortex, swelling in thalamic structures, and filling of the primary lens vesicle). Thus, the marsupial pattern of development represents the derived condition among therians. But, perhaps contrary to expectations, it is the development of the central nervous system that is delayed rather than craniofacial development being advanced.

We also uncovered an additional 32 timing changes that were present for only one or two of the character optimizations examined. Because these shifts are equivocal, we do not hold them to be as robust as the 12 above. Nor do they concur as strongly with Smith's (1997) observed pattern. Only 13 of the 26 changes along the marsupial branch and three of the six along the placental branch matched Smith's (1997) observations. In all cases, however, directionality was identical between shifts identified by both Smith (1997) and ourselves.

## Discussion—A Comparison of Methods

Event-pairing represented a breakthrough for the study of sequence heterochrony. However, its full potential has been limited by the lack of a rigorous quantifiable framework in which to analyze the data. Previous attempts to interpret event-paired data had to be done by hand (e.g., Smith, 1997). This is a cumbersome procedure, in part because the number of event-pairs to be examined $[(n^2 - n)/2]$ is much greater than the number of developmental events ($n$) raises the possibility of introducing human error (Nunn and Smith, 1998). Moreover, the method is not quantifiable. Only pairs of events that have changed position during evolution can be identified,

TABLE 8. Developmental timing data for the chicken (*Gallus gallus*) for the 28 events of Smith (1997). Information was obtained from Romanoff (1960) and Bellairs and Osmond (1998) and is given in number of days postfertilization. Two sequences were derived to account for different timing information for skull ossification events (events 1–12). Full definitions of events are found in Smith (1997).

| | Event | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| *Gallus*-1 | 8 | 8 | 8 | 8 | 8 | 10 | 7 | ? | 11 | 8 | 9 | ? | 13 | 11 | ? | ? | ? | ? | ? | ? | ? | ? | 5 | 2.67 | ? | ? | ? | 4 |
| *Gallus*-2 | 12 | 11 | 12 | 12 | 12 | 13 | 10 | ? | 13 | 11 | 12 | ? | 13 | 11 | ? | ? | ? | ? | ? | ? | ? | ? | 5 | 2.67 | ? | ? | ? | 4 |

TABLE 9. Significantly moving events identified from a reanalysis of Smith's (1997) data set (as given in Nunn and Smith, 1998[a]). Inferred changes are common to ACCTRAN, DELTRAN, and MinF optimization criteria. All changes occurred along the branch leading to marsupials; no unequivocal changes occurred during the evolution of placentals. The single change indicated in italics was not observed by Smith (1997). Full definitions of events are given in Smith (1997).

| Event | Direction | Relative to |
|---|---|---|
| First ossification of maxilla | Early | First appearance of tooth buds, differentiation of malleus and incus |
| 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 malleus and incus, craniofacial muscles distinguishable |
| Primary lens cells fill lens vesicle | Late | First ossification of frontal, differentiation of malleus and incus |

[a] The data for events 17 and 18 were accidentally reversed in Table 2 of Nunn and Smith (1998) (K. K. Smith, pers. comm.), which has been accounted for here.

whereas the number of actively moving events giving rise to this pattern may be much more restricted (see worked example above).

Event-pair cracking extends the utility of event-pairing by providing a quantifiable framework that summarizes efficiently the wealth of data points produced by the latter to identify events that have actively changed their developmental timing. In our opinion, it is the only method that possesses the three desirable features we listed for analyzing developmental timing data with a view to elucidating sequence heterochronies.

An alternative to event-pairing is to analyze the ranked developmental data directly, often through the use of rank correlation measures such as Spearman's rank correlation (Mabee and Trendler, 1996; Larsson, 1998) or Kendall's coefficient of concordance (Nunn and Smith, 1998). Nunn and Smith (1998) also used ANOVAs to analyze the ranked developmental data of Smith (1997), were able to identify specific events for which the position (i.e., mean rank) in the developmental sequence differed significantly between marsupial and placental mammals. Altogether, their results were largely in agreement with those of Smith (1997) and of this study.

However, direct analyses of the ranked developmental sequence use overall (phenetic) measures of similarity and thus are not suitable within a phylogenetic framework (Mabee and Trendler, 1996)—even with computer simulation to produce a null distribution that accounts for phylogenetic relatedness (e.g., Nunn and Smith, 1998). For example, because the ANOVA procedure of Nunn and Smith (1998) compares the group means of sister groups, it cannot determine in which lineage any heterochronic changes have occurred, even with the inclusion of an outgroup. To precisely localize the changes, and determine their direction of movement, one must make a post hoc comparison of the ancestral developmental sequences.

Moreover, because these procedures cannot truly localize any changes, they can be misled by strong signal in extreme cases. For instance, in comparing marsupial and placental mammals, consider the situation where the heterochronic changes actually occurred within marsupials after the sister species of the remaining marsupials had diverged (Fig. 4). This sister species possesses an otherwise placental developmental pattern. If the heterochronic change within marsupials is great enough, the group means of marsupials and placentals would still be very different. Thus an ANOVA could falsely infer that the changes *also* occurred between these two groups. Errors such as these can be spotted only if all sister groups on the tree are compared and some method of accounting for duplicate changes in adjacent nodes is devised.

We simulated this scenario by using the data of Nunn and Smith (1998). One species of placental mammal was declared to be a marsupial and the $F$ statistics for each event were recalculated. We did this for each placental mammal species in turn. In all cases, events 23, 25, 27, and 28 still differed

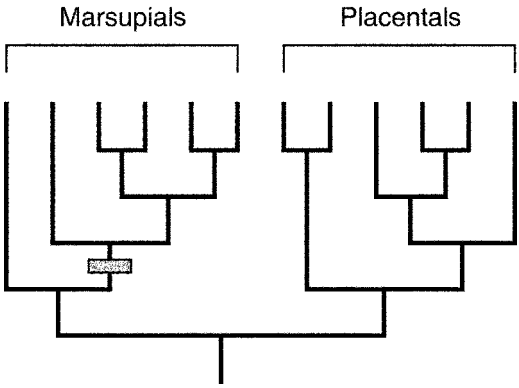


FIGURE 4. Hypothetical example showing how the direct analysis of a ranked developmental sequence can be led astray by skewed distributions. All heterochronic changes have taken place at the point marked with a bar (i.e., within marsupials), with the sister species of the remaining marsupials having a placental developmental pattern.

significantly between "placentals" and "marsupials"; events 1 and 3 differed significantly for five out of the six cases when a placental mammal was designated as a marsupial. Admittedly, the *F* values were lower in all cases, indicating that the differences were not as strong. The differences were also usually significant only when we used critical values uncorrected for phylogenetic relatedness. However, the corrected critical values were highly dependent on the model of evolution assumed for the simulations (Nunn and Smith, 1998). In any case, our example demonstrates the potential for direct analyses of ranked developmental data to be misled by their failure to localize inferred timing changes.

In contrast, event-pair cracking makes explicit use of phylogenetic information, which yields numerous advantages. Event-pair cracking can precisely localize heterochronic changes without additional analyses because it uses the evolutionary transformations inferred to have occurred on specific branches on a phylogeny (itself a rigorous technique). Furthermore, all these transformations, except for those occurring between the ingroup and the outgroup, are already polarized, allowing event-pair cracking to simultaneously infer the direction of movement. Finally, our method is able to distinguish between events that are actively moving from those that are only apparently moving (i.e., twins and hitchhikers). This unique combination of features makes event-pair cracking a potentially valuable technique for the analysis of sequence heterochrony. Further, we believe our method will be useful in the analysis of many forms of comparative dynamic data, both temporal and spatial (e.g., changes in gene order).

## ACKNOWLEDGMENTS

We thank Sharon Gobes, Paula Mabee, Kathleen Smith, and an anonymous reviewer for many constructive suggestions on an earlier version of this paper. J. E. J. was supported by the Leverhulme Trust. O. R. P. B.-E. was supported in part by an NSERC postdoctoral fellowship and by the van der Leeuw Fonds.

## REFERENCES

ALBERCH, P. 1985. Problems with the interpretation of developmental sequences. Syst. Zool. 34:46–58.

BELLAIRS, R., AND M. OSMOND. 1998. The atlas of chick development. Academic Press, San Diego, California.

BENTON, M. J. 1991. Amniote phylogeny. Pages 317–330 *in* Origins of the higher groups of tetrapods: Controversy and consensus (H.-P. Schultze and L. Trueb, eds.). Cornell Univ. Press, Ithaca, New York.

BININDA-EMONDS, O. R. P., J. E. JEFFERY, M. I. COATES, AND M. K. RICHARDSON. In press. From Haeckel to event-pairing: The evolution of developmental sequences. Theory in Biosciences.

DETTLAFF, T. A., AND A. A. DETTLAFF. 1961. On relative dimensionless characteristics of the development duration in embryology. Arch. Biol. (Liege) 72:1–16.

GOULD, S. J. 1977. Ontogeny and phylogeny. Belknap Press, Cambridge, Massachusetts.

GOULD, S. J. 1982. Change in developmental timing as a mechanism of macroevolution. Pages 333–346 *in* Evolution and development (J. T. Bonner, ed.). Springer Verlag, New York.

GRANDCOLAS, P., P. DELEPORTE, L. DESUTTER-GRANDCOLAS, AND C. DAUGERON. 2001. Phylogenetics and ecology: As many characters as possible should be included in the cladistic analysis. Cladistics 17:104–110.

GRIBNAU, A. A. M., AND L. G. M. GEIJSBERTS. 1981. Developmental stages in the Rhesus monkey (*Macaca mulatta*). Springer-Verlag, Berlin.

HALL, B. K., AND T. MIYAKE. 1995. How do embryos measure time? Pages 3–20 *in* Evolutionary change and heterochrony (K. J. McNamara, ed.). Wiley, New York.

HARLAND, W. B., R. L. ARMSTRONG, A. V. COX, L. E. CRAIG, A. G. SMITH, AND D. G. SMITH. 1990. A geologic time scale 1989. Cambridge Univ. Press, Cambridge.

HOLLAND, P. W. 1999. The future of evolutionary developmental biology. Nature 402 (suppl.):41–44.

JEFFREY, J. E., O. R. P. BININDA-EMONDS, M. L. COATES, AND M. K. RICHARDSON. In press. Analyzing evolutionary patterns in vertebrate embryonic development. Evol. Dev.

LARSSON, H. C. E. 1998. A new method for comparing ontogenetic and phylogenetic data and its application to the evolution of the crocodilian secondary palate. Neues Jahrb. Geol. Paläontol., Abh. 210:345–368.

LUO, Z.-X., A. W. CROMPTON, AND A.-L. SUN. 2001. A new mammaliform from the early Jurassic and evolution of mammalian characteristics. Science 292:1535–1540.

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: Percomorpha): Intraspecific variation and interspecific comparisons. J. Morpholol. 227:249–287.

MCCRADY, E., JR. 1938. The embryology of the opossum. The Wistar Institute of Anatomy and Biology, Philadelphia.

MCKINNEY, M. L., AND K. J. MCNAMARA. 1991. Heterochrony: the evolution of ontogeny. Plenum Press, New York.

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.

RAFF, R. A. 1996. The shape of life: Genes, development, and the evolution of animal form. Univ. of Chicago Press, Chicago.

RAFF, R. A. 2000. Evo-devo: The evolution of a new discipline. Nat. Rev. Genet. 1:74–79.

RICHARDSON, M. K. 1995. Heterochrony and the phylotypic period. Dev. Biol. 172:412–421.

RICHARDSON, M. K. 1999. Vertebrate evolution: The developmental origins of adult variation. BioEssays 21:604–613.

RICHARDSON, M. K. 2000. Developmental sequences. Pages 120–122 *in* McGraw-Hill Yearbook of Science and Technology 2001. McGraw-Hill, New York.

ROMANOFF, A. L. 1960. The avian embryo: Structural and functional development. Macmillan, New York.

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. 2001. Heterochrony revisited: The evolution of developmental sequences. Biol. J. Linn. Soc. 73:169–186.

VELHAGEN, W. A. 1997. Analyzing developmental sequences using sequence units. Syst. Biol. 46: 204–210.

WILEY, E. O. 1981. Phylogenetics: The theory and practice of phylogenetic systematics. Wiley-Liss, New York.

*First submitted 17 July 2001; revision submitted 10 December 2001; final acceptance 27 December 2001*
*Associate Editor: Paula Mabee*