EVOLUTION & DEVELOPMENT 6:1, 50–57 (2004)

Comparative and phylogenetic analysis of developmental sequences

Susanne Schulmeister* and Ward C. Wheeler

Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA

*Author for correspondence (e-mail: schulmei@amnh.org)

SUMMARY Event pairing has been proposed for the optimization of developmental sequences (event sequences) on a given phylogenetic hypothesis (cladogram) to determine instances of sequence heterochrony. Here, we show that event pairing is faulty, leading to the optimization of impossible hypothetical ancestors, the underestimation of the lengths of the developmental sequences on the tree, and the proposition of synapomorphies that are not supported by the data. When used for phylogenetic analysis, event pairing can even produce cladograms that are inconsistent with the data. These errors are caused by the fact that event pairing treats dependent features as if they were independent. We present a new method for comparative and phylogenetic analysis of developmental sequences that does not exhibit these errors. Our method applies Search-based character optimization and treats the entire developmental sequence as a single character that is then analyzed by using an edit cost function, which specifies the transformation cost between pairs of observed and unobserved character states, and dynamic programming. In other words, the developmental sequence is directly optimized on the tree. We used event pairing as an edit cost function, but others are possible.

INTRODUCTION

In the discipline of evo-devo (evolutionary developmental biology), developmental observations are put into an evolutionary context. But many comparative studies in developmental biology do not use a phylogenetic framework (Richardson et al. 2001). A reason for this is the lack of methods for the phylogenetic analysis of developmental sequences until recently.

The development or ontogeny of an individual can be described roughly as a long list of events. These events can be morphological or molecular. The time at which a certain developmental event takes place during ontogeny is usually more or less fixed for a given species but can change during the course of evolution. Because comparison of the *absolute* timing of ontogenetic events among different taxa is fraught with difficulties (Raff and Wray 1989; Smith 1997; Bininda-Emonds et al. 2002; Jeffery et al. 2002b, p. 479), the *relative* temporal order of these events has been used instead for comparative analysis of interspecific differences in developmental timing (Smith 1997, 2001). A list of selected developmental events in chronological order (i.e., the order in which they take place in the ontogeny of an individual) is called a developmental sequence. If a change in developmental timing brings about a change in the chronological order of the events, this is called sequence heterochrony. (For discussions of various definitions of heterochrony, see Raff and Wray 1989; Godfrey and Sutherland 1995; Alberch and Blanco 1996; Klingenberg 1998; Gould 2000; Smith 2001, 2002; Kovác 2002; and references therein.) Here we are concerned with methods for the comparative and phylogenetic analysis of developmental sequences to identify instances of sequence heterochrony. In this article, comparative analysis (character mapping, ancestral state reconstruction) refers to the optimization of developmental sequence data on a given phylogeny, whereas phylogenetic analysis of developmental sequences refers to their use in determining the phylogeny itself. Both types of analyses aim to determine the hypothetical developmental sequences of the ancestors of the species included in the analysis to study the evolution of developmental sequences. In addition to this, phylogenetic analysis aims to determine the phylogeny of the studied species.

To avoid confusion, it may be worth mentioning that even though development and ontogeny are synonymous, developmental sequences are quite different from ontogenetic sequences. A developmental sequence (= developmental series = event sequence) is a list of different events in the chronological order in which they happen in the ontogeny (e.g., differentiation of neural plate → eye vesicle formation → formation of lens placode → retina formation → pigmentation of retina → heart formation), whereas an ontogenetic sequence (= transformation sequence) is a description of the development of a single organ or structure throughout the course of the ontogeny, that is, a list of the conditions of one character

found in an individual at different times (e.g., Mabee and Humphries 1993, p. 175). Koenemann and Schram (2002) referred to developmental sequences as "ontogenetic event sequences," further increasing the potential of confusing them with ontogenetic sequences. Velhagen (1997, p. 204) understood the term "developmental sequences" to comprise both types, that is, event sequences and transformation sequences, and Reiss (2002, p. 89) said the same about the term "ontogenetic sequences."

Developmental sequences have three properties that, in combination, make their analysis a unique problem:

1. Dependency. If the relative order of the events is the focus of the analysis, events are not independent of each other because no event can be studied in isolation. For example, if in the sequence ABC event A changes its relative order with C, it cannot do so without also changing its relative order with B, unless B also changes its position in the sequence; these changes are interdependent.
2. Simultaneity. Two or more events can be recorded as happening at the same time, that is, sampling of the ontogeny is not sufficient to completely resolve the temporal order of the events in question (Bininda-Emonds et al. 2002, pp. 302–305).
3. Distance. If it is assumed that the timing of an event changes gradually in the course of evolution, that is, changing from sequence ABC to CAB must proceed through the sequence ACB, then the number of events that have been "passed by" must be taken into account in the analysis. Contrary to gene order, for example, the evolutionary change from ABCD to BCDA would be assigned a higher cost than would be a change from ABCD to BCAD.

These properties have posed a challenge for the development of methods for the analysis of developmental sequences. No methods for their use in phylogenetic analysis have been proposed up to now. In the past, at least a few methods for comparative analysis of developmental sequences have been proposed, but most of these only serve to compare two sequences or two groups of sequences and are not discussed here. (For reviews of these methods, see Nunn and Smith 1998; Smith 2001; Bininda-Emonds et al. 2002.) Event pairing is the most recent proposal and the first method that can be used to examine developmental sequence heterochrony on a cladogram. The principle of analyzing developmental sequences by recoding them in all possible pair-wise combinations of events was developed simultaneously by Smith as "event pairs" (1996, 1997) and by Velhagen as "sequence units" (1995, 1997) with only a minor difference and, less explicitly, by Mabee and Trendler (1996). This principle has been used in a number of studies (Smith 1996, 1997; Velhagen 1997; Blanco and Sanchiz 2000; Chipman et al. 2000; Jeffery et al. 2002a; Maisano 2002; Sánchez-Villagra 2002). Event pairing had been proposed to optimize variation on a given cladogram, but it has also been used as a method for phylogenetic analysis (i.e., for the determination of the phylogenetic hypothesis itself), even though the authors who made these analyses themselves pointed to problems associated with this (Velhagen 1997; Blanco and Sanchiz 2000; Jeffery et al. 2002a; Maisano 2002; Sánchez-Villagra 2002).

Here we demonstrate that event pairing has a serious flaw that can lead to erroneous results. The error is not restricted to the use of the method for phylogenetic analysis but appears in simple character optimization (comparative analysis) as well. We then describe a method for both comparative and phylogenetic analysis of developmental sequences that does not have this error.

EVENT PAIRING

In event pairing, the developmental sequences of several taxa (Fig. 1A) are broken up into all possible pair-wise combinations of events. These are called event pairs and make up the characters of the data matrix (Fig. 1B). In the example in Fig. 1, the developmental sequences contain four events

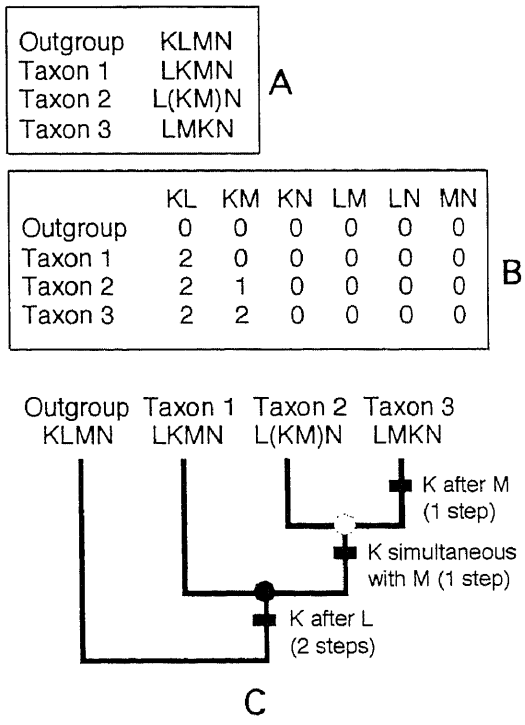

Fig. 1. Analysis of developmental sequences with event pairing. (A) A simple data set of developmental sequences with four events for four taxa. (B) The event-pair data matrix derived from the data set. (C) The results of optimizing the data matrix on a given cladogram. See text.

(Fig. 1A). In one of the sequences, two events, K and M, happen simultaneously in the ontogeny, which is indicated by the parentheses. For sequences with four events, there are six event pairs, that is, six characters in the data matrix (Fig. 1B).

The character states of each character in the data matrix describe the relative order in the developmental sequence of the two events of the pair, that is, whether the first event listed in the event pair comes *before* the other event in the ontogeny/developmental sequence (state 0), *simultaneously* with the other event (state 1), or *after* it (state 2). This is the notation proposed by Smith (1996, p. 71, 2001). This coding implies that if an event X moves from being before Y to being simultaneous with Y it costs one step, whereas going from being before Y to being after Y costs two steps, provided that the character is treated additively (i.e., as an ordered character). An alternative notation, suggested by Velhagen (1995, 1997), is to code event pairs (= sequence units) of simultaneous events as unknown ("?"), because the simultaneity is usually artifactual and obscures the actual temporal order of the two events. However, here we focus on the notation of Smith (1997) and postpone any discussion of the properties of the different notations to a future publication.

The characters of an event-pair data matrix like that in Fig. 1B can be analyzed like other phylogenetic characters, using maximum parsimony, either Fitch parsimony (Fitch 1971) (i.e., characters are treated as unordered/nonadditive) or Wagner parsimony (Kluge and Farris 1969; Farris 1970) (i.e., characters are treated as ordered/additive). If the data matrix (Fig. 1B) is optimized on the cladogram shown in Fig. 1C, the total length of the characters on the cladogram is four steps (ordered analysis). The stem species of the ingroup (closed circle) is hypothesized to have had event K happening after L in the ontogeny, which is a synapomorphy for the ingroup taxa. The most recent common ancestor of taxa 2 and 3 (open circle) is optimized as having had K simultaneous with M in the developmental sequence, which is a synapomorphy for taxa 2 and 3.

This optimization is unambiguous, provided that the characters are treated as additive (ordered) in the analysis. Additive treatment of the event-pair characters must be used if it is assumed that developmental events are delayed or advanced in the ontogeny only gradually in the course of evolution (i.e., through intermediate states), meaning that in the case where a developmental sequence evolved from ABCD to ACBD, one assumes that there must have been an intermediate state A(BC)D in which B and C were more or less simultaneous. If, however, it is assumed that the sequence ABCD could directly evolve into ACBD without going through an intermediate, the characters should be treated as nonadditive (unordered). If the characters from Fig. 1B are treated as nonadditive when optimizing them on the cladogram in Fig. 1C, the optimization of the second character is ambiguous because "K simultaneous with M" could either be an autapomorphy of taxon 2 or a synapomorphy of taxa 2 and 3.

Event pairing can easily accommodate missing or inapplicable events. For example, if event K were not present in taxon 2, a question mark would be coded for the first three characters for taxon 2 in the data matrix (Fig. 1B). Because a transformation from any character to "?" (or vice versa) has zero cost, the length of the characters on the tree would still be four. State 2 of the first character (event pair KL) would still be hypothesized as a synapomorphy of taxa 1, 2, and 3. The optimization of the second character would be ambiguous.

In an ideal case like the one described above, event pairing provides the correct length of the characters on the tree—that is, the length according to the rules postulated by the method—and correctly determines hypothetical ancestral developmental sequences. However, we found this is not always the case. An example is presented in Fig. 2. One of the ingroup species has the same sequence (ABC) as the outgroup taxon. In taxon 2, event A is switched with B and C, whereas in taxon 3, event C is switched with A and B (Fig. 2A). These are two contradicting changes, and the distance between any two of the three sequences is always four steps. Therefore, the length of the developmental sequences on any tree must always be eight steps. Which means that no relationships whatsoever are supported by these sequences.

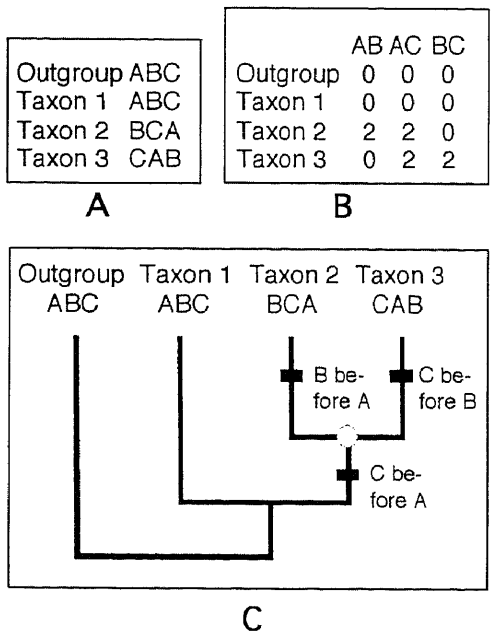

Fig. 2. A simple example demonstrating the flaw in event pairing. (A) The data set. (B) The event-pair data matrix derived from the data set. (C) The results of optimizing the data matrix on a given cladogram. See text.

If these developmental sequences are coded as an event-pair data matrix (Fig. 2B) and this matrix is then optimized on the tree in Fig. 2C, the most recent common ancestor of taxa 2 and 3 (open circle) is hypothesized to have had the character states 020. (This is true for both nonadditive and additive treatment of the characters.) This would mean that in the developmental sequence of this stem species, event A should be before B and event B should come before C and that, at the same time, event C should be *before* A, which is impossible. The reconstruction of the developmental sequence of this ancestor is clearly incorrect. (The same happens if the coding of Velhagen [1997] is used.) Moreover, the use of this impossible hypothetical ancestral sequence leads event pairing to underestimate the length of the developmental sequences on the tree as six steps instead of the eight steps that would result from using only possible ancestral assignments. Also, event pairing supports the sister-group relationship of taxa 2 and 3 with the synapomorphy "event C before A," even though no support for any relationships can actually be found in the original data.

SEARCH-BASED OPTIMIZATION

The cause of the error in event pairing is that dependent features (e.g., the relative order of A and B and the relative order of B and C) get treated as if they were independent. To avoid this, we propose using the entire developmental sequence as one complex character. This method is illustrated in Fig. 3, using the example (of Fig. 2) that caused event pairing to fail. The data set shown in Fig. 3A is translated 1:1 into a data matrix, which is shown in Fig. 3B. The data matrix contains a single character. The three different developmental sequences observed in the four terminal taxa each constitute one state of this character. The character is optimized using Search-based optimization (Wheeler 2003), an algorithm that uses dynamic programming and an edit cost metric among the possible states (Sankoff optimization; Sankoff and Rousseau 1975).

The crux of the method is that the edit cost matrix is not restricted to comparisons among the observed developmental sequences (in this case ABC, BCA, and CAB) but may also include other (perhaps all) possible developmental sequences. This means not only all permutations of the developmental events, but also sequences in which events are missing. In the example in Fig. 3, we excluded the possibility of simultaneous events or deletions (missing events) for simplicity and space limitation (there would be six additional states involving simultaneous events and many more involving missing events). With the exclusion of simultaneities and deletions, there are only six possible permutations for the three events in the developmental sequences. The step matrix in Fig. 3C gives

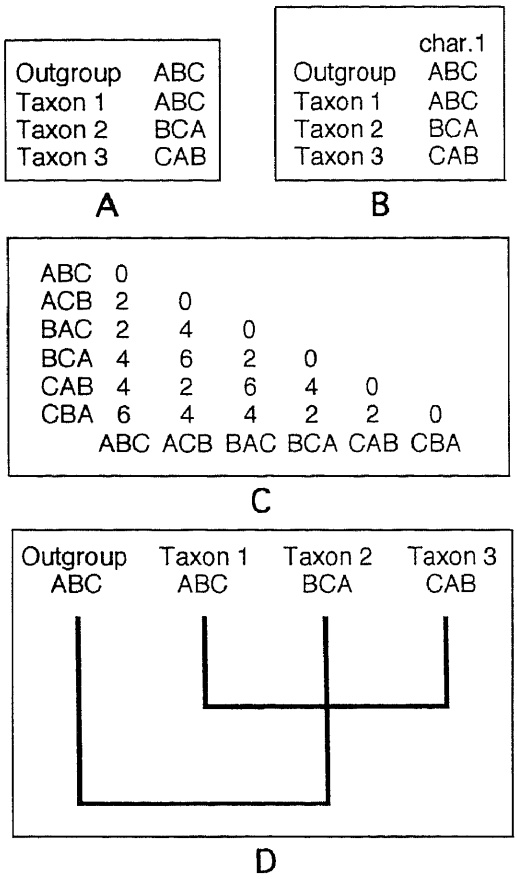

Fig. 3. Analyzing the example from Fig. 2 with Search-based character optimization. (A) The data set. (B) The data matrix with a single character representing the developmental sequence. (C) The corresponding step matrix (Sankoff matrix) with all possible states for the character (if the possibility of simultaneities is excluded). (D) The cladogram resulting from the analysis of the character in B with the step matrix in C. It is unresolved and has a length of eight steps.

the cost involved to transform one character state into another, for all combinations of the six character states.

The cost assigned to the transformations from one character state to another is calculated according to an edit cost function that is specified before the analysis. In the present example, we used event pairing as the edit cost function, which means that the transformation cost between two developmental sequences is determined by coding them in event pairs (as in Fig. 2B); the cost between two sequences is the distance between their event-pair character states. We used the notation of Smith (1997): state 0 = before, state 1 = simultaneous, and state 2 = after. This implies that a switch of two neighboring events costs two steps. If an event "skips" over two other events, for example, from ABCDE to ACDBE, this would cost four steps, and so on. Analyzing the data matrix of Fig. 3B with the step matrix in Fig. 3C using

dynamic programming (Sankoff and Rousseau 1975) leads to the unresolved tree in Fig. 3D with a length of eight steps. As explained above, this is the correct result for the given data, because no relationships are implied by them.

Many other edit cost functions could be specified, for example, ones that are based on sorting or based on genomic break-point analysis. It is important to note that the edit cost function introduces assumptions into the analysis. One should be well aware of these implicit assumptions, which differ among edit cost functions. The choice of the edit cost function can greatly influence the result of the analysis. However, we postpone the discussion of different edit cost functions and their ramifications to a future publication, because here we are focusing on the general principle of the analysis of developmental sequences.

Missing and inapplicable events are unproblematic with our method. The event that is missing or unknown in a particular taxon is simply left out of the developmental sequence. Intraspecific variation (polymorphisms) could also be taken into account, but this has not been implemented yet. However, the possibility of polymorphisms must be excluded from hypothetical ancestral sequences because they could cause non-metricity. As noted above, developmental events can be morphological or molecular in nature, which means that our method can be used not only for analyzing morphological developmental data but also chronological sequences of expression of genes.

The method outlined above is an application of Search-based character optimization that is described in more detail by Wheeler (2003). Search-based character optimization is an extension of Fixed States optimization (Wheeler 1999). The method outlined above has been implemented in the existing computer program POY (Wheeler et al. 2002). POY allows other data, such as morphological characters, DNA sequences, and gene order data, to be analyzed simultaneously with the developmental sequences.

AN EXAMPLE

A reanalysis of the Velhagen (1997) data was performed. The data set contains six species of thamnophiine snakes, for which five events in the ossification of skull bones had been recorded by Velhagen (1997). We performed an analysis of these data using the method presented here. First, the program "gendevoseqs" was used to create a list of all 541 developmental sequences that are possible if the possibility of missing events is disregarded. This is a reasonable assumption because the six species in the data set each have all five events. Using this list, the 541 × 541 edit cost matrix was generated by the program "gen2edit." (Binaries and source code of the two programs, which were written by W. C. Wheeler, are available at ftp.amnh.org/pub/molecular/devoseqs.) The edit costs (transformation costs) in the matrix were calculated by using the event-pairing methodology of Smith (1996). Search-based optimization (Wheeler 2003) was used to search for a topology and those of the 541 states that would minimize cladogram length (Wheeler et al. 2002). The program did an exhaustive state search, trying all combinations of the 541 states for all four ancestral nodes in the 70 topologies examined (wagner build+TBR); this took 48 seconds on a personal computer with an 800-MHz PIII processor. In this case, a cladogram of 17 steps was returned, but each of the internal (and several of the terminal) branches has a potential length of zero; hence, all internal branches are collapsed, meaning that the cladogram is actually a bush. This is because most ancestor-descendant branches share possible states under equally parsimonious optimizations.

Conventional event-pairing analysis with the coding of Smith (1996) yielded 10 event-pair characters and a single most parsimonious cladogram of 16 steps. The reason why our analysis yielded a tree that is one step longer than that obtained with conventional event pairing is that the latter gave an ancestral state reconstruction that postulates an impossible ancestor (in one ancestor, event B is supposed to be simultaneous with M as well as S, but in the same ancestor event S is supposed to be before M). This demonstrates that the problem does not only occur in hypothetical constructed examples but also with real data.

DISCUSSION

Event pairing and dependence

We showed above and in Fig. 2 that the use of event pairing in comparative analysis of developmental sequences can lead to the generation of impossible ancestral sequences, underestimation of the length of the characters on the tree, and erroneous synapomorphy schemes. These problems are caused by the disruption of the necessary relationships of the individual events. A linear sequence of events is broken up into individual parts that are treated as if they had no relation to each other. Breaking up the sequences with the events A, B, and C into the pairs AB, AC, and BC enables the analysis to take a nonmetric "short-cut" via a logically impossible hypothetical ancestor.

Event pairing was developed originally only for comparative analysis—the optimization of developmental sequences on a given phylogeny. Bininda-Emonds et al. (2002, p. 297) suggested that event pairing "may also yield data that can be used in phylogeny reconstruction," and Velhagen (1997), Blanco and Sanchiz (2000), Jeffery et al. (2002a), Maisano (2002), and Sánchez-Villagra (2002) actually used it for phylogenetic analyses. However, if used for this purpose, event pairing can produce all the errors mentioned above for its use in comparative analysis and, in addition, can even lead

to the preference of trees for which there is no evidence in the original data. If the data matrix in Fig. 2B is subjected to a cladistic analysis and the topology is rooted on the outgroup, the resulting cladogram looks like the one in Fig. 2C and has a length of six steps. However, because the data themselves do not contain any phylogenetic information, the tree should be unresolved and have a length of eight steps, as the one determined by our method (Fig. 3D).

Bininda-Emonds et al. (2002, p. 314) discussed that "event-pairing involves two forms of non-independence," which is (a) ontogenetic dependence (meaning that a certain combination of events have to occur in a certain order during ontogeny) and (b) coding dependence. Bininda-Emonds et al. stated further that "[n]either form of non-independence is detrimental when heterochrony data are mapped onto an existing phylogeny." As we showed above, this statement is incorrect. Bininda-Emonds et al. (2002, pp. 314–316) went on to say that "both are problematical when heterochrony data are used to infer a phylogeny." In their further discussion of this statement, they were, however, not concerned with the technical errors discussed here. They merely worried about the questions (a) whether the concerted "movement" of two or more ontogenetically dependent events should and could somehow be counted as if only one event moved and (b) whether a large "movement" of one event (skipping several other events) should count more than a small "movement" (skipping only one neighboring event) and if yes, whether the cost should increase linearly (as it does in event pairing). These two questions concern the implicit assumptions that the analytical methods make about the evolution of development and are not discussed here. Koenemann and Schram (2002) also noted that "ontogenetic events are characterized by both a collective and linear type of dependence and, in this, violate the criterion of independence," but they also did not notice that breaking this dependence up causes event pairing to be logically inconsistent.

Velhagen (1997, p. 209), however, noticed a "logical interdependence," meaning that "the units of a sequence should not contradict each other." But he dismissed the problem by saying that "[e]ven in cases where the character states of units" (= event-pairs) "are determined independently of each other (such as when ancestral sequences are inferred . . .), I have not yet encountered a case of true contradiction . . . ." This is interesting, considering that exactly this kind of contradiction occurs with his data set: If his own sequence-unit data matrix as presented in his Table 3 (Velhagen 1997) is optimized on the tree {({({*N. taxispilota T. radix*} *T. proximus*) *S. occipitomaculata*} *N. sipedon*) *S. dekayi*}, the most recent common ancestor of the first three species is postulated to have had the states 2111112122, which means that event M is before B and B is before S and that, at the same time, S is supposed to be before M, which is clearly impossible. This does not mean, however, that event pairing is entirely useless for the analysis of developmental sequence data. It can be used as an edit cost function to determine the distance between pairs of developmental sequences.

Search-based character optimization

To replace event pairing as an optimization method for developmental sequences, we propose the use of Search-based character optimization, in which the entire developmental sequence is used as a single character. The character is optimized by using a step matrix that contains hypothetical in addition to observed developmental sequences. The transformation costs specified by the step matrix are calculated according to an edit cost function that is specifically tailored for the purpose of analyzing developmental sequences. We used event pairing as an edit cost function, but other functions are possible.

Our method avoids the problem that causes the failure of event pairing by treating the entire developmental sequence as one complex character, thus treating dependent relationships as dependent relationships instead of as independent characters. In event pairing, the word "character" is used "in a practical sense only to refer to the individual elements analysed with reference to a phylogeny (= event-pairs here)" (Jeffery et al. 2002b, p. 481), whereas in our method, the analyzed character is identical with the observed character, the developmental sequence (cf. Fig. 3, A and B). Because the developmental sequence is not recoded in any way, at no point in the analysis can impossible ancestors be created.

It was mentioned above that Search-based optimization may search among all possible hypothetical developmental sequences. The program "gendevoseqs" can be used to create a list of all possible states that can then be fed into POY (Wheeler et al. 2002). However, this is feasible only with a very small number of events. Various heuristics can be used to lessen the computational load. In the extreme, the analysis could be restricted to include only observed developmental sequences. But that should hardly ever be necessary. If the full set of events has been recorded for each of the terminal taxa, it is reasonable to assume that all events had also been present in all ancestors (at least as a heuristic), so that the possibility of missing events can be excluded when creating the list of hypothetical ancestral states with "gendevoseqs." The only other heuristic that is currently implemented in "gendevoseqs" samples possible sequences from the universe of all possible sequences at random (i.e., with a uniform probability distribution). It is desirable to develop heuristics in which the set of possible ancestral sequences is selected in a directed fashion, for example so that those sequences would be selected that are intermediate between pairs of observed states.

Our method can be used for both comparative and phylogenetic analysis of developmental sequences. Search-based

character optimization (Wheeler 2003) was originally developed for the optimization of DNA sequences and gene order data. It is an extension of Fixed States optimization (Wheeler 1999). Both methods use a DNA sequence fragment as a character, but in Fixed States the step matrix contains only observed DNA sequences.

Historical background of our method

Sankoff and Rousseau (1975) described a method that used differential transformation (edit) costs for maximum parsimony analysis of multistate characters. The edit costs were calculated according to an edit cost function and specified in an edit cost matrix, which was later termed a step matrix. This technique is an application of dynamic programming and is now often referred to as Sankoff optimization.

Mabee and Humphries (1993) suggested to code an entire transformation sequence of a structure as a single character and analyze it with Sankoff optimization. They also suggested that in addition to observed transformation sequences, unobserved ones could be included in the step matrix. Mabee and Humphries (1993) did not develop an algorithm to do this but simply coded the step matrix by hand with all observed and unobserved states (there were only 31 in total). Contrary to event sequences, the conditions in the transformation sequence are always in the same relative order and cannot be recorded as being simultaneous, so that the edit cost function simply involved counting the number of conditions that were different between two transformation sequences (e.g., the distance between "abc" and "bd" is three steps). Hence, their problem is much simpler than ours.

The idea of using the entire developmental sequence as a single character is not new. Velhagen (1997, p. 209) mentioned that "it may be argued that entire sequences should be treated as single characters" and that step matrices could be used "to quantify the evolutionary steps among different possibilities," which corresponds to the basic principle of the method proposed here. However, he argued that in practice "it can be difficult to derive a single number to represent the simultaneous (and interdependent) changes in order among several events, particularly when sequences are not fully resolved." This means that Velhagen (1997) did not realize that the method he proposed (which is equivalent to event pairing) could be used for exactly this purpose: determining the number of evolutionary steps between pairs of developmental sequences. Maisano (2002, p. 280) noted that "there are two extremes as to how sequence data can be treated in phylogenetic analyses" and that "one extreme is to code the entire ontogenetic sequence as a different character state (Mabee and Humphries 1993)." However, she did not use this approach, because it "would be unsatisfactory in the present study because it is based on the assumption that the entire sequence of postnatal development behaves as a single character. This cannot be the case, because the events in these sequences change their relative order from one species to another." As we have shown, the opposite is true. The fact that the relative order of the events changes makes it necessary to analyze the sequence as a whole.

Obviously, the analysis of developmental sequences is a problem similar to that of analyzing gene order data (genomic rearrangement data), and Wheeler (2003, p. 355) mentioned that Search-based optimization could be applied to gene order data. Currently, break-point analysis is used to analyze gene order data. In a future article we will discuss the use of break-point analysis for the optimization of developmental sequences.

Biological constraints

Here we showed that event pairing can lead to the postulation of impossible hypothetical ancestral developmental sequences. By this, we were exclusively referring to logically impossible sequences (e.g., if the character states of the event-pair characters imply that in the developmental sequence A should come before B and B before C and that C should come before A, which is obviously impossible in a linear sequence). It should be noted, however, that developmental sequences that are logically possible could still be impossible for biological reasons. Generative developmental constraints can limit two or more events to a certain chronological order (e.g., Richardson and Chipman 2003 and references therein). For example, pigmentation of the retina can only happen after the formation of the retina.

In principle, our method can take these biological constraints into account, even though this has not yet been implemented in the software. However, this can easily be done by running the list of possible developmental sequences generated by "gendevoseqs" through a filter to remove those developmental sequences that are thought to be biologically impossible.

SUMMARY AND CONCLUSIONS

Here we show that event pairing cannot be used as a method for comparative or phylogenetic analysis of developmental sequences. The reason for this is that event pairing treats temporally dependent relations as independent characters. We have demonstrated that Search-based character optimization can be used for both comparative and phylogenetic analysis of developmental sequences. It treats the entire sequence as one phylogenetic character, thus avoiding the disruption of the temporal relationships of the individual ontogenetic events. Though Search-based character optimization is in principle much more direct than event pairing, it is also much more computationally intense.

In the context of this new approach, event pairing can be used as one of several possible edit cost functions. Future work should focus on the development of other edit cost functions and discuss the assumptions that the edit cost functions imply about the evolution of developmental sequences. Developmental biologists should consider the appropriateness of these assumption for various types of data. Future work should also be concerned with the development of better heuristics for the selection of hypothetical ancestral states (developmental sequences) to be used in the analysis.

The application of Search-based character optimization to the analysis of developmental sequence heterochrony serves to show the generality of the method. We envision an even broader application of the method in the future.

Acknowledgments

We thank Olaf Bininda-Emonds for bringing the problem of analyzing developmental sequences to the attention of the first author. We also thank Mark Wilkinson and Pablo Goloboff for reviewing the manuscript.

REFERENCES

Alberch, P., and Blanco, M. J. 1996. Evolutionary patterns in ontogenetic transformation: from laws to regularities. *Int. J. Dev. Biol.* 40: 845–858.

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

Blanco, M. J., and Sanchiz, B. 2000. Evolutionary mechanisms of rib loss in anurans: a comparative developmental approach. *J. Morphol.* 244: 57–67.

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

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

Fitch, W. M. 1971. Towards defining the course of evolution: minimal change for a specific tree topology. *Syst. Zool.* 20: 406–416.

Godfrey, L. R., and Sutherland, M. R. 1995. What's growth got to do with it? Process and product in the evolution of ontogeny. *J. Human Evol.* 29: 405–431.

Gould, S. J. 2000. Of coiled oysters and big brains: how to rescue the terminology of heterochrony, now gone astray. *Evol. Dev.* 2: 241–248.

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

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

Klingenberg, C. P. 1998. Heterochrony and allometry: the analysis of evolutionary change in ontogeny. *Biol. Rev.* 73: 79–123.

Kluge, A. G., and Farris, J. S. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18: 1–32.

Koenemann, S., and Schram, F. 2002. The limitations of ontogenetic data in phylogenetic analyses. *Contr. Zool.* 71 (1/3): 47–65.

Kovác, V. 2002. Synchrony and heterochrony in ontogeny (of fish). *J. Theor. Biol.* 217: 499–507.

Mabee, P. M., and Humphries, J. 1993. Coding polymorphic data: examples from allozymes and ontogeny. *Syst. Biol.* 42: 166–181.

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

Maisano, J. 2002. The potential utility of postnatal skeletal developmental patterns in squamate phylogenetics. *Zool. J. Linn. Soc.* 136: 277–313.

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

Raff, R. A., and Wray, G. A. 1989. Heterochrony: developmental mechanisms and evolutionary results. *J. Evol. Biol.* 2: 409–434.

Reiss, J. O. 2002. The phylogeny of amphibian metamorphosis. *Zoology* 105: 85–96.

Richardson, M. K., and Chipman, A. D. 2003. Developmental constraints in a comparative framework: a test case using variations in phalanx number during amniote evolution. *J. Exp. Zool. Mol. Dev. Evol.* 296: 8–22.

Richardson, M. K., Jeffery, J. E., Coates, M. I., and Bininda-Emonds, O. R. P. 2001. Comparative methods in developmental biology. *Zoology* 104: 278–283.

Sánchez-Villagra, M. R. 2002. Comparative patterns of postcranial ontogeny in therian mammals: an analysis of relative timing of ossification events. *J. Exp. Zool. Mol. Dev. Evol.* 294: 264–273.

Sankoff, D., and Rousseau, P. 1975. Locating the vertices of a Steiner tree in arbitrary space. *Math. Progr.* 9: 240–246.

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.

Smith, K. K. 2002. Sequence heterochrony and the evolution of development. *J. Morph.* 252: 82–97.

Velhagen, W. A. 1995. A comparative study of cranial development in the thamnophiine snakes (Serpentes: Colubridae). Ph.D. dissertation, Duke University, Durham, North Carolina.

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

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

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

Wheeler, W. C., Gladstein, D. S., and De Laet, J. 2002. POY. Version 3.0. Source code and binaries available at ftp.amnh.org/pub/molecular/poy. Documentation by D. Janies and W. Wheeler.