Notes and Comments

Quantitative Tests of General Models for the Evolution of Development

Steven Poe[1],[*] and Marvalee H. Wake[2],[†]

1. Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico 87131;
2. Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, California 94720-3140

ABSTRACT: Comparative developmental biologists have proposed models to describe patterns of conserved features in vertebrate ontogeny. The hourglass model suggests evolutionary change is most difficult at an intermediate "phylotypic" stage, the adaptive penetrance model suggests change is easiest at an intermediate stage, and the early conservation model suggests change is easier later in ontogeny. Although versions of some of these models have been discussed since the nineteenth century, quantitative approaches have been proposed only recently. Here we present quantitative phylogenetic approaches to evaluating trends in the evolution of ontogeny. We apply these approaches to the proposed models and demonstrate that an existing approach to assessing these models is biased. We show that the hourglass, adaptive penetrance, and early conservation models are unnecessarily complex explanations of the patterns observed in developmental event data for 14 species of vertebrates. Rather, a simpler model that postulates that evolutionary change is easier between ontogenetically adjacent events is adequate.

Keywords: development, evolution, hourglass model, ontogeny, vertebrates.

Recent interest in integrating development and evolution has focused attention on general theories and models for the ways that development evolves. Three models have been proposed to describe the evolution of ontogeny, primarily for vertebrates. According to the early conservation

* Corresponding author; e-mail: anolis@unm.edu.

† E-mail: mhwake@socrates.berkeley.edu.

(EC) model (von Baer 1828; Haeckel 1874; Balfour 1875; summarized in Gould 1977; Arthur 1997; Richardson 1999), evolutionary change occurs more frequently in late development whereas early development is relatively conserved (fig. 1A). Haeckel (1874) postulated that evolutionary change was by terminal addition of features; more recently, developmentalists have construed that the genetic basis for heterochrony, as well as for novel features, is a consequence of relatively later phenotypic expression of relatively earlier (temporally and evolutionarily) genetic modifications (see discussion in Richardson 1999). In the hourglass (HG) model, early and late development are relatively labile, and an intermediate, "phylotypic" stage is most conserved (fig. 1B). This pattern was first observed by von Baer (1828) and Haeckel (1874), and the concept has recently been developed further by several others (e.g., Ballard 1981; Elinson 1987; Slack et al. 1993; Duboule 1994; Collins 1995; Raff 1996). The hourglass is characterized mechanistically by the early variation in features of yolk mass, egg size, and blastulation/gastrulation features, followed by the "constrained" period during which general features of body plans develop, and then by the divergence produced as phenotypic differences emerge. Currently, several biologists interested in that relationship have questioned the existence of a phylotypic stage (e.g., Richardson 1995, 1999; Hall 1997; Richardson et al. 1997, 1998), citing incongruent definitions of the phylotype and limited evidence for stage-based similarity. It remains, however, that "developmental stages are temporal clusters of morphologic character states" (Richardson 1999, p. 609), though the phylotypic stage may be the entire organogenetic period, characterized by a progression during which homologous features emerge. A third model is offered by Richardson (1999, fig. 6), the adaptive penetrance (AP) pattern. Richardson suggests that development is constrained early by there being few positional fields expressed in eggs and that modifications may be amplified but are not likely to produce adaptive adult variation. However, during the phylotypic period, there are many positional fields, and small modifications are amplified that
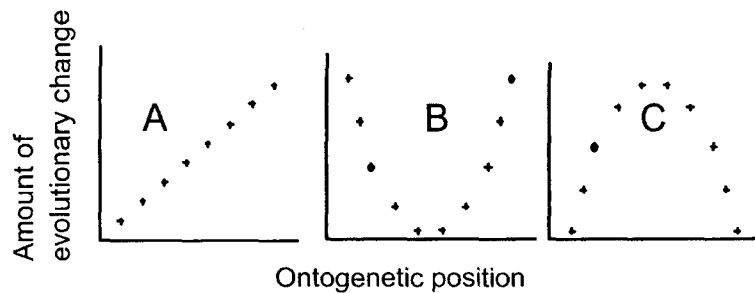
Figure 1: Expected patterns of relationship between ontogeny and evolutionary lability according to (A) early conservation model, (B) hourglass model, and (C) adaptive penetrance model.

are subject to natural selection that may produce adaptive adult variation. Finally, late modifications may have adaptive significance but are not amplified, so the pattern is constrained again (fig. 1C).

Bininda-Emonds et al. (2003) present a useful discussion of these models and point out that the differing conceptions of them (especially HG) have hampered rigorous testing. They distinguished between pattern- and process-based interpretations and adopted a pattern-based view of the HG model that recognized different periods of "phenotypic diversity." The perspective that we adopt in this note is most similar to this pattern-based view. However, we prefer to discuss these models in terms of differences in evolutionary lability rather than diversity in phenotype. This distinction becomes important when more than two species are studied and phylogenetic information is taken into account. For example, great phenotypic divergence in development could be due to either great evolutionary lability of development or great phylogenetic divergence between compared species (see app. A in the online edition of the *American Naturalist*). Verbal and descriptive arguments have been offered in favor of and opposition to each of these models, but only recently (Bininda-Emonds et al. 2003) have quantitative tests been proposed.

Smith (2001) drew from earlier work (e.g., Smith 1996) to discuss the quantification of several concepts in the evolution of development in terms of heterochrony of "sequence events." Smith (1996) and Jeffery et al. (2002) have advocated "event-pairing" approaches to analyzing development wherein each developmental event is categorized as occurring before, simultaneously with, or after each other event, and the resulting data are used to infer evolutionary changes in ontogenetic rank (see also Mabee and Trendler 1996; Velhagen 1997). This class of approaches appears useful for identifying shifts in the timing of particular characters (e.g., Smith [1996] found differences in timing of the central nervous system relative to muscular and skeletal characters in comparisons of mar-

supial and placental mammals) but may be less useful for testing the general theories that are the focus of this note. Approaches have been suggested that examine overall development for trends, such as Nunn and Smith's (1998) use of multiple correlation to assess the conservation of development across taxa. However, because this approach does not recognize temporal differences, it is not suitable for testing the theories of concern in this note. Schlosser's (2001) graphical method may be useful for examining temporal trends in the evolution of development. However, that approach does not incorporate quantitative metrics or significance tests, so it may be more suitable for exploratory work. Bininda-Emonds et al. (2003) presented a quantitative approach to assessing some of the models analyzed here, but we demonstrate below that their approach is biased and thus unable to distinguish these models (app. A). Below we present new, unbiased, quantitative phylogenetic approaches designed specifically to test general theories for how development evolves. Although our tests are novel, we adopt the general framework of the above authors in considering development in terms of a ranked sequence of events.

The concept of sequence events in heterochrony suggests testable predictions for general models of the evolution of development. In the context that we present, evolutionary lability manifests as changes in the relative timing of developmental events. The EC model predicts changes in timing will be less frequent earlier in ontogeny, the HG model predicts changes will be least frequent at an intermediate stage, and the AP model predicts that changes will be most frequent at an intermediate stage. Simpler models also make predictions for heterochrony. A model we call the adjacency (AJ) model predicts that small changes in ontogenetic position are more likely than large changes. Unlike the EC, AP, and HG models, the AJ model does not hypothesize a temporal aspect to ontogenetic constraint. Probably the simplest heterochronic model is one where there are no constraints on change.

This unconstrained (UC) model allows all types of change to be equally probable such that there is no phylogenetic inertia of ontogenetic position. That is, an event that occurs early in ontogeny in a given species, for example, is equally likely to occur in late or early ontogeny in even a closely related species. Here we present the first unbiased statistical assessment of the EC, HG, and AP models and show that none of them is required to explain the patterns observed in developmental event data for 14 species of vertebrates. Rather, the simple AJ model is adequate.

Data

The data we consider are the relative ranks of 41 developmental events distributed across ontogeny in 14 vertebrate species. J. Jeffery kindly provided the data used in Jeffery et al. (2002). (See Jeffery et al. [2002] for a description of events and of the rationale for use of these particular events.) Examples of events include appearance of the first somite and formation of the first aortic arch. Perhaps most importantly for our purposes, these events were selected "without a priori consideration of whether they would show heterochrony" (Jeffery et al. 2002, p. 295). We coded the data for consideration in a phylogenetic context using the event as the character and the rank of occurrence of the event as the character state. We combined states such that ranks 1 and 2 were assigned state (rank) a, ranks 3 and 4 were assigned state b, and so forth, because the phylogenetic program MacClade, which we used to map character transformations, can handle only up to 26 character states. Thus, our coding included 20 ranks (a–t) for each taxon. The data matrix is shown in appendix B in the online edition of the *American Naturalist*. For all analyses, we assumed the tree shown in figure 2 of Jeffery et al. (2002).

Methods

Test for Adjacency Model

We first wanted to test whether large changes in ontogenetic position are less evolutionarily frequent than small changes (the AJ model). We compared this model to a model where change was unconstrained (UC). Given enough evolutionary time (or if evolutionary change occurs fast enough), data generated by these models is expected to converge to the predictions of UC.

We can distinguish these models operationally by mapping the ranks of the developmental events on the phylogenetic tree using MacClade (Maddison and Maddison 1992) and examining how often different classes of change in rank occur using MacClade's "Trace all changes" function. We tested the AJ model against the UC model by

comparing the number of steps change to the frequency of occurrence of that number using a Spearman's signed-ranks test. We scaled the frequency of occurrence of change by the number of opportunities (although this scaling had no effect on results). For example, there are 38 ways to have a change of rank of one step ($a \rightarrow b$, $b \rightarrow a$, $b \rightarrow c$) but only 10 ways to have a change of 15 steps ($a \rightarrow p$, $p \rightarrow a$, $b \rightarrow q$, etc.). Thus, the number of one-step changes was divided by 38, the number of 15-step changes was divided by 10, and so on. The AJ model predicts that smaller changes in rank (i.e., changes of fewer steps) will occur more often than larger changes in rank, whereas the UC model predicts that larger changes in rank will occur just as frequently as smaller changes.

Tests for Hourglass, Adaptive Penetrance, and Early Conservation Models

The HG, AP, and EC models make predictions about the relative temporal occurrence of heterochrony (fig. 1). In order to distinguish these models, we would like to compare amount of evolutionary change to position in ontogeny. In order to make this comparison, we need some quantification of ontogenetic position and some measure of evolutionary lability. Selection of these metrics is complicated by two factors.

First, our data are in the form of ranks, which have peculiar statistical and (in the current case) biological properties. Different ranks facilitate different types of changes. For example, an event ranked a (first) in an ancestor has one way to change by one step (to b) in a descendant, one way to change by two steps, and so on. An event ranked p has two ways to change one step (to q or o), two ways to change two steps, and so on. If there are differences in the evolutionary lability of these types of changes (regardless of temporal position), then comparisons of raw amounts of change (or stasis) will be biased by the positional tendencies of the events. For example, if change is more likely between adjacent ranks (see "Results"), events occurring early and late in ontogeny will be biased toward infrequent change simply because they have fewer opportunities for "easy" change (see also Bininda-Emonds et al. 2003). In order to correct for this bias, we examined only one-step changes in rank (i.e., $a \rightarrow b$, $g \rightarrow f$) in our comparison of these models. The potential downside to analyzing only one-step changes is a decrease in power due to omitting all other types of changes. This decrease appears to be small, however, because a large percentage of changes were of one step (see "Results"). Our measure of ontogenetic position, then, was simply the rank of the one-step changes that occurred. That is, changes between a and b were ranked one, changes between b and c were ranked two, and so on. (We also

analyzed the data using only "independent" changes, e.g., *a* to *b* and *c* to *d* but not *b* to *c*. This had no effect on results).

A second difficulty is that measures of lability that involve counting evolutionary changes or stases are likely to be sensitive to the number of states available. For example, if there are many more *a* states than *p* states because of ties, change to (or stasis at) *a* is likely to occur more often than change to *p* regardless of ontogenetic position just by chance. This phenomenon in fact occurs in the current data set, where the numbers of changes, stases, and states for a given rank are strongly intercorrelated (data not shown). This means that raw measures of evolutionary lability—such as the number of unambiguous evolutionary changes per ontogenetic rank—are biased measures, and some scaling should be enacted to account for this bias. Under the assumption that this bias affects changes and stases equally, an appropriate scaling would be to compare number of changes (i.e., between *a* and *b*) to number of stases (i.e., number of branches on which there is no change in *a* plus number of branches on which there is no change in *b*), scaled by the total possible opportunities for change and stasis. The fundamental unit here is the branch of the evolutionary tree—either change or stasis in rank can occur in each event on each branch, and both change and stasis are important for the models evaluated here.

The measure of evolutionary lability that we adopt is

$$L = \frac{\text{no. unambiguous changes} - \text{no. stases}}{\text{no. unambiguous changes} + \text{no. stases}}.$$

The formula increases as change becomes increasingly common. It is positive when change is more frequent than stasis and negative when stasis is more common, and it reaches its maximum of 1.0 when change occurs on every branch (which will not occur under a parsimony criterion) and its minimum of -1.0 when change does not occur.

We compared ontogenetic rank with our measure of evolutionary lability and analyzed the resulting patterns. The EC model is supported if lability increases monotonically with ontogenetic rank. We tested this model using a Spearman's signed-ranks test. The HG and EC models make slightly more ambiguous quantitative predictions. Although these models make clear conceptual predictions for the approximate relative degree of lability across ontogeny—for example, high on the ends and low in the middle for the HG model—a precise operational depiction of these predictions is not obvious. We chose to model these predictions as a parabola, humped in the case of the AP model, valley-shaped for the HG model. We assess this model using polynomial regression analysis, with rank as the independent variable and $\ln [2 + L]$ as the dependent variable. We tested whether the addition of a second-order term (which is responsible for the parabola shape) significantly increases the fit of the regression model over a simple first order model. If the squared term significantly increases fit, then either the AP model (if the squared term is positive) or the HG model (if the term is negative) is supported.

Example Calculations

Figure 2 shows a worked example of the calculations involved in testing the different models. The example evaluates three ontogenetic events (fig. 2A–2C) as exemplars from a larger data set. Ontogenetic ranks for the three events are listed at the tips of the tree for each of the four species in this example. For instance, the event in figure 2A occurs second (*b*) in the ontogenies of two species, first (*a*) in one species, and third (*c*) in one species. Figure 2A–2C shows the distribution of changes and stases for each event under the Fitch (1971) parsimony criterion, and figure 2D summarizes the number of changes and stases of each type for the three events.

Figure 2E shows the comparison undertaken in testing the AJ model against the UC model. In this limited sample, one-step changes in rank are most common ($n = 3$), as would be expected under the AJ model.

Figure 2F shows the results for the test of EC, HG, and AP models. Lability *L* of the earliest ($a \leftrightarrow b$) rank is calculated as

$$L = \frac{\text{no. } a \leftrightarrow b \text{ changes} - \text{no. } a \text{ stases} - \text{no. } b \text{ stases}}{\text{no. } a \leftrightarrow b \text{ changes} + \text{no. } a \text{ stases} + \text{no. } b \text{ stases}}$$

$$= \frac{1 - 4 - 4}{1 + 4 + 4}$$

$$= -0.78.$$

Lability of the other ranks is calculated the same way. Lability at rank *cd* is undefined because there are no changes between *c* and *d* and no stases of *c* or of *d*. Lability at rank *de* is -1.0 (maximum stasis) because there are no changes between states *d* and *e* but stasis at *e* (although not at *d*) occurs. As there appears to be neither an increase in lability during ontogeny nor a humped or valley-like pattern to these data, none of the more complex models is supported in this limited sample.

Results

Table 1 summarizes data for the test comparing the AJ and UC models using the Jeffery et al. (2002) data. The AJ model is supported. The number of evolutionary
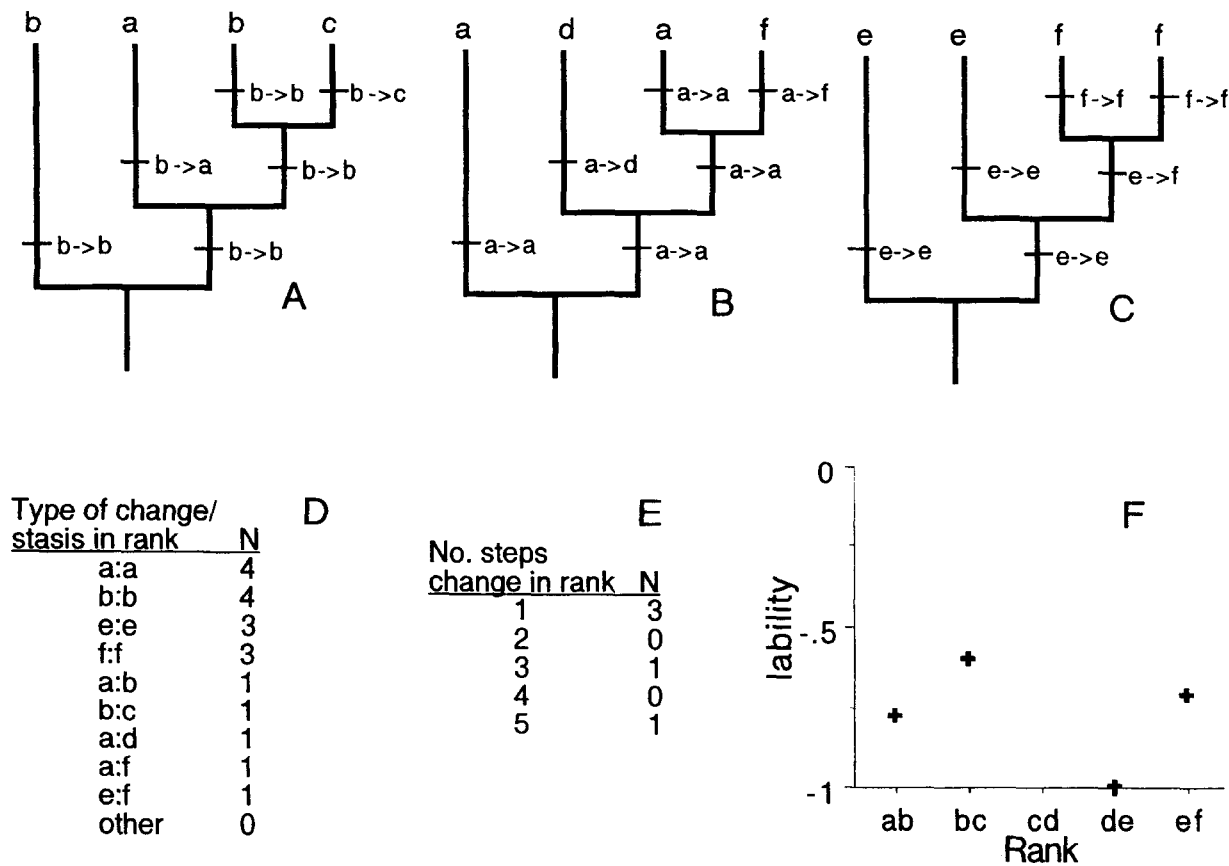
Figure 2: Example calculations for evaluation of models for the evolution of development

changes decreases monotonically with the amount of change in ontogenetic rank ($P = .008$; Spearman's signed-ranks test).

None of the more complex models is required to explain the data (fig. 3). A Spearman's signed-ranks test comparing ontogenetic position and evolutionary lability is not significant ($P = .40$), which suggests that the EC model is not supported. In the polynomial regression analysis, addition of a squared term does not significantly increase the fit of the model ($P = .93$ for squared coefficient in second-order polynomial regression; R^2 increases from .0986 to .0992 when a squared term is added to the linear regression equation). This result suggests that neither the HG nor the AP models offers a significantly improved fit to the data.

Discussion

Interpretation of Patterns

The lack of support for the EC, HG, and AP models suggests that heterochronic changes in development show no

temporal trend of evolutionary conservation. Case studies have shown that evolutionary changes can occur at virtually any stage of development (e.g., Wray and McClay 1989; Mabee 1993). These studies of restricted groups of organisms and isolated periods of development showed that changes may occur over a wide range of developmental time but left open the possibility that some overall trend exists for developmental conservation in evolution. Our quantitative analyses suggest that no such trend in event heterochrony exists and support the qualitative studies that have questioned the existence of especially constrained stages in development such as a "phylotypic" stage (e.g., Richardson et al. 1997).

What does this finding mean in terms of a process-based explanation? It could be argued that no explanation is necessary; that is, the lack of some pattern of temporal constraint in heterochrony supports the notion that there is nothing special about the evolution of development. Perhaps the processes that operate throughout development (e.g., timing of gene function and the interaction of gene products) provide a sufficient template for both stasis

Table 1: Number of unambiguous evolutionary changes for different numbers of steps change in ontogenetic rank

No. steps change in ontogenetic rank	No. unambiguous evolutionary changes	No. unambiguous evolutionary changes/ no. opportunities
1	40	1.05
2	20	.56
3	8	.24
4	7	.22
5	6	.20
6	2	.07
7	1	.04
6–19	0	0

and change that they overshadow mechanistic (e.g., tissue induction) differences between stages of development. Or, alternatively, perhaps the mechanistic differences between periods of development have been overstated. The AJ model that we favor suggests some constancy in evolutionary lability throughout development regardless of whether selective processes and developmental integration are temporally diverse or constant.

Finally, we note the possibility that the perception of patterns such as the "hourglass" in the evolution of development reflects human constructs. That is, even if there is some agreed-on pattern, it is possible that the explanation for such a pattern is uninteresting biologically. For example, if the perceived conservation of early development is a consequence of our categorization of events—for example, the four-cell stage cannot occur before the two-cell stage—rather than of differing selective pressures or greater developmental integration, then there is little to be gained in trying to explain why early development is conserved.

Quantitative Approaches

The methods that we have proposed provide an approach to analysis of ontogenesis that is inclusive of several quantitative and evolutionary facets not fully captured in other approaches. First, the context of sequences of developmental events that we adopt here affords a convenient means to address these hypotheses quantitatively (Bininda-Emonds et al. 2003). Virtually all discussion of the EC, HG, and AP models has been qualitative. Second, by including the effects of stases as well as changes in the evaluation of the patterns of evolution, our lability metric gives more explanatory power to patterns of heterochrony. Third, the tests presented in this note are explicitly phylogenetic. They interpret similarities and differences in ontogeny as changes or stases along the branches of a phylogenetic tree. We consider this phylogenetic aspect

desirable because incorporating phylogeny allows clearer interpretation of results for these evolutionary questions. Nonphylogenetic approaches may be more complicated by the existence of different expectations or null models for closely and distantly related species (app. A; and see Smith 2001, p. 178). Finally, our approaches incorporate significance tests in order to assess these models rigorously. We have formulated these models as precise statistical hypotheses and evaluated them according to established standards of model comparison. Other conceptions of these models (e.g., those incorporating anatomical differences between ontogenies) will require different formulations and different tests.

Limitations of the Approach

We have shown that the traditionally advocated models for the evolution of development are not necessary to explain these data for heterochrony in sequence events for 14 vertebrate species. Rather, the more parsimonious AJ model is adequate. However, we note several limitations to the generality of these conclusions. Many of these limitations are shared with other approaches that utilize sequence events and have been discussed by previous authors (e.g., see Smith 2001; Jeffery et al. 2002). Here we review these limitations with reference to the analyses and conclusions of this note.

First, definition and selection of developmental events for analysis is unavoidably subjective. There is an essentially infinite number of potential events, and all of these events show varying degrees of nonindependence with other events. Many events are constrained to necessarily occur after or before other events (e.g., ossification of a bone cannot occur before formation of mesoderm). Although worthy of consideration as potential sources of bias, these aspects are not necessarily fatal for event-sequence approaches. Regarding the conclusions of this note, subjective selection of events and nonindependence

of events are only likely to affect conclusions if these factors affect large and small changes in rank differently (in the comparison of AJ and UC) or if they affect early and late ontogeny differently (in the evaluation of EC, HG, and AP). We know of no evidence for the former possibility. However, one selection-related factor that may compromise our results is the paucity of events from very early ontogeny. Concepts such as the HG model were originally formulated with such fundamental early events as gastrulation in mind. Very early occurring and complex events such as these may be less amenable to sequence analysis because they are either necessarily similar in timing (gastrulation must occur early) or different in form between distantly related species (gastrulation patterns are a historically important taxonomic character). We consider this factor to be a potential weakness of our study. However, we note that even if the events analyzed in this study are a sample from later ontogeny, one would expect statistically significant trends to emerge from the HG, AP, and EC models. For example, if we are omitting events from the early conserved part of an HG pattern, one would expect a negative slope from regression analysis and a significant value for the Spearman's signed-ranks analysis, and these results do not obtain (fig. 3). Similarly, our formulation of the EC model states that there is a monotonic increase in evolutionary lability with ontogenetic rank. Under this view, any particular slice of ontogeny is legitimately testable for the EC trend.

Another factor that might be considered to compromise our conclusions is the use of parsimony as a character optimization criterion. Model-based methods are generally preferable because they are able to account for differing amounts of evolutionary change along the branches of a phylogenetic tree. Implementing such methods in this case would involve formulating a model for evolutionary change in developmental rank and programming a way to compare changes in types of rank. It might be possible, for example, to produce a null distribution by applying Nunn and Smith's (1998) model to these data (tree, branch lengths, number of events) and tallying the number of evolutionary changes in rank of different degrees (one step, two steps, etc.) and from different periods of development (early, late) for simulated "ontogenies" at the tips of the tree. These distributions could be compared with empirical counts from the analyzed species. Although we do not consider parsimony to be the ideal optimization criterion for these data, we do not expect the use of this criterion to bias our conclusions. As discussed above, the use of this criterion is inappropriate only if it affects different classes of changes (early vs. late, small vs. large) differently, and we see no reason for this to be the case under the Fitch parsimony criterion used in this note.

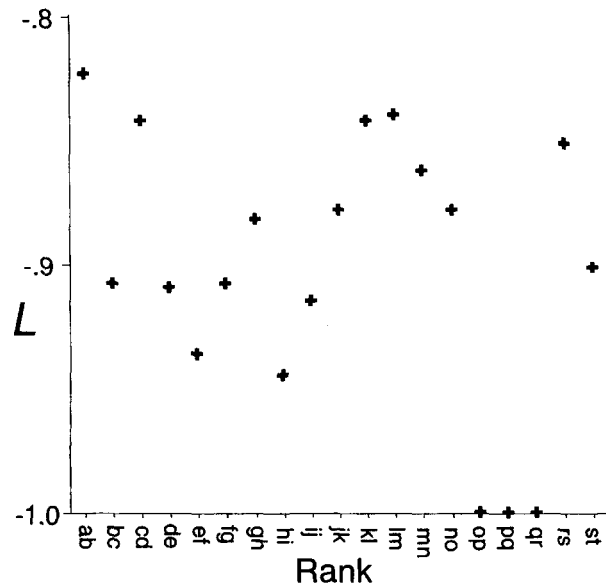Finally, we mention some biological restrictions on the

Figure 3: Comparison of evolutionary lability L to position in ontogeny for 41 developmental events in 14 vertebrate species. The absence of patterns shown in figure 1 indicates that the early conservation, hourglass, and adaptive penetrance models are not supported.

generality of our conclusions. In general, we emphasize that this note does not constitute the final word on the validity of the EC, HG, and AP models for the evolution of development. We have examined these models only in a very narrow context—specifically, in terms of heterochrony in selected morphological sequence events during a particular period of ontogeny in a small number of vertebrate species. These models are unnecessarily complex explanations of the data we analyzed here, but they may turn out to be valid models for, for example, gene expression patterns in closely related insects. We expect that the application of the tests we have presented in this note and of new tests (that perhaps incorporate differences in the events themselves and not just timing of identical events) to diverse types of data for a wide variety of organisms will offer new insights into the applicability of these models. The above reservations aside, we close by emphasizing that there currently exists no quantitative evidence in support of the EC, HG, or AP models. Our simpler AJ model provides an adequate quantitative explanation of developmental pattern as a first approximation.

Acknowledgments

Literature Cited

Arthur, W. 1997. The origin of animal body plans: a study in evolutionary developmental biology. Cambridge University Press, Cambridge.

Balfour, F. M. 1875. A comparison of the early stages in the development of vertebrates. Quarterly Journal of Microscopical Science 15:207–226.

Ballard, W. W. 1981. Morphogenetic movements and fate maps of vertebrates. American Zoologist 21:391–399.

Bininda-Emonds, O. R. P., J. E. Jeffery, and M. K. Richardson. 2003. Inverting the hourglass: quantitative evidence against the phylotypic stage in vertebrate development. Proceedings of the Royal Society of London B 270:314–346.

Collins, P. 1995. Embryology and development. Pages 102–341 *in* P. L. Williams, ed. Gray's anatomy. 38th ed. Churchill Livingstone, London.

Duboule, D. 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate bauplan and the evolution of morphologies through heterochrony. Development (suppl.):135–142.

Elinson, R. P. 1987. Changes in developmental patterns: embryos of amphibians with large eggs. Pages 1–21 *in* R. A. Raff and E. C. Raff, eds. Development as an evolutionary process. Alan Liss, New York.

Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Systematic Zoology 20:406–416.

Gould, S. J. 1977. Ontogeny and phylogeny. Harvard University Press, Cambridge, Mass.

Haeckel, E. 1874. Anthropogenie oder Entwichlungsgeschichte des Menschen. Engelmann, Leipzig.

Hall, B. K. 1997. Phylotypic stage or phantom: is there a highly conserved embryonic stage in vertebrates? Trends in Ecology & Evolution 12:4.

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

Mabee, P. M. 1993. Phylogenetic interpretation of ontogenetic change: sorting out the actual and artifactual in an empirical case study of centrarchid fishes. Zoological Journal of the Linnaean Society 107:175–291.

Mabee, P. M., and T. A. Trendler. 1996. Development of the cranium and paired fins in *Betta splendens* (Teleostei: Percomorpha): intraspecific variation and interspecific comparisons. Journal of Morphology 227:249–287.

Maddison, W. P., and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, Mass.

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

Raff, R. 1996. The shape of life. University of Chicago Press, Chicago.

Richardson, M. K. 1995. Heterochrony and the phylotypic stage. Developmental Biology 172:412–421.

———. 1999. Vertebrate evolution: the developmental origins of adult variation. Bioessays 21:604–613.

Richardson, M. K., J. Hanken, M. L. Gooneratne, C. Pieau, A. Raynaud, L. Selwood, and G. M. Wright. 1997. There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. Anatomy and Embryology 196:91–106.

Richardson, M. K., A. Minelli, M. Coates, and J. Hanken. 1998. Phylotypic stage theory. Trends in Ecology & Evolution 13:158.

Schlosser, G. 2001. Using heterochrony plots to detect dissociated coevolution of characters. Journal of Experimental Zoology 291:282–304.

Slack, J. M., P. W. Holland, and C. F. Graham. 1993. The zootype and the phylotypic stage. Nature 361:490–492.

Smith, K. K. 1996. Integration of craniofacial structures during development in mammals. American Zoologist 36:70–79.

———. 2001. Heterochrony revisited: the evolution of developmental sequences. Biological Journal of the Linnaen Society 73:169–186.

Velhagen, W. A., Jr. 1997. Analyzing developmental sequences using sequence units. Systematic Biology 46:204–210.

von Baer, K. E. 1828. Über Entwicklungsgeschichte der Thiere: Beobbachtung und Reflexion. Vol. 1. Bornträger, Königsberg.

Wray, G. A., and D. R. McClay. 1989. Molecular heterochronies and heterotopies in early echinoid development. Evolution 43:803–813.