

## TEST OF VON BAER'S LAW OF THE CONSERVATION OF EARLY DEVELOPMENT

STEVEN POE

Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico 87131  
E-mail: anolis@unm.edu

*Abstract.*—One of the oldest and most pervasive ideas in comparative embryology is the perceived evolutionary conservation of early ontogeny relative to late ontogeny. Karl Von Baer first noted the similarity of early ontogeny across taxa, and Ernst Haeckel and Charles Darwin gave evolutionary interpretation to this phenomenon. In spite of a resurgence of interest in comparative embryology and the development of mechanistic explanations for Von Baer's law, the pattern itself has been largely untested. Here, I use statistical phylogenetic approaches to show that Von Baer's law is an unnecessarily complex explanation of the patterns of ontogenetic timing in several clades of vertebrates. Von Baer's law suggests a positive correlation between ontogenetic time and amount of evolutionary change. I compare ranked position in ontogeny to frequency of evolutionary change in rank for developmental events and find that these measures are not correlated, thus failing to support Von Baer's model. An alternative model that postulates that small changes in ontogenetic rank are evolutionarily easier than large changes is tentatively supported.

*Key words.*—Development, heterochrony, ontogeny, vertebrates, Von Baer.

Received May 4, 2006. Accepted August 7, 2006.

Karl Von Baer (1828) observed that anatomical characteristics of early ontogeny tend to be shared across several taxa, whereas characteristics of later ontogeny tend to be specific to particular taxa. Darwin (1859), Haeckel (1874), and others (e.g., Wimsatt 1986; Arthur 1997) formulated this phenomenon, known as Von Baer's law, in evolutionary terms, suggesting a model in which early ontogeny is less susceptible to evolutionary change. Although this "law" has been criticized as an overgeneralization (views summarized in Raff 1996)—evolutionary change appears possible at virtually any stage in ontogeny (e.g., Wray and McClay 1989)—the lability of late ontogeny relative to early ontogeny remains accepted as a general trend (e.g., Arthur 2004). The concept has never been tested quantitatively on a broad scale. Here, I compare rank in ontogeny to frequency of change in rank of developmental events to show that this model is not supported in several vertebrate clades.

Early ontogeny is expected to be more evolutionarily conserved than late ontogeny due to the temporal and spatial nonindependence of ontogenetic events (e.g., Arthur 2004 p.132). Changes in early ontogeny may cause changes in later ontogeny, but the converse occurrence seems extremely unlikely. This asymmetry is a simple consequence of temporal causality: a current event may or may not alter some future (i.e., later ontogenetic) event, but cannot alter the past (i.e., earlier ontogeny) (Hume 1739). Laid on top of this temporal relationship is an apparently greater integration of early ontogeny. A mutation in a later developmental pathway may affect morphology mostly independently of other events or pathways (a change in the number of digits does not affect brain chemistry), but early mutations have the potential to effect changes in many systems. If most changes in ontogeny—like other mutations—are deleterious, then a single early change may be deleterious for many future ontogenetic events, including critical events such as the formation of a brain.

Comparative tests of ontogenies have been elusive because such tests require comparability across species, and many events of ontogeny differ between species. The context of a ranked sequence of developmental events surmounts this dif-

ficulty (Smith 2001) and allows comparison of degree of evolutionary conservation between species across ontogenetic time. Examples of events are ossification of a bone or appearance of a somite. Recent papers (Bininda-Emonds et al. 2003; Poe and Wake 2004) have used this context to test general models for how development evolves across species.

In the context adopted here, evolutionary change is quantified in terms of changes in rank of ontogenetic events (Smith 1996; Mabee and Trendler 1996). Von Baer's law suggests a temporal aspect to constraint: that changes in rank should occur less frequently in early ontogeny. The operationalization of Von Baer's model used in this paper assumes that temporal position in ontogeny and frequency of evolutionary change should be correlated: if early ontogeny is more constrained, changes in rank are expected to occur more frequently as ontogeny progresses. An ontogenetically atemporal model postulates that change is more frequent between more adjacent ranks, and the simplest unconstrained model allows all types of changes to be equally probable. Poe and Wake (2004) called these the early conservation (EC), adjacency (AJ), and unconstrained (UC) models, respectively. Here I use 13 datasets of vertebrates to test whether developmental events tend to change rank less frequently early in ontogeny (EC), whether adjacent changes in rank are more common than larger changes (AJ), or whether all magnitudes of change in rank are equally probable (UC).

### METHODS

#### *Datasets*

Datasets of ranked ontogenetic events were compiled from the literature to satisfy the following criteria: at least three taxa, at least five events, maximum of 26 events (the limit for MacClade), no missing data, and no taxa with all missing data or all ranks tied. In some cases events or taxa were omitted from datasets in order to satisfy these criteria. Ten datasets were found to meet these criteria. One of these (Jeffery et al. 2002) was analyzed as three clades and another (Nunn and Smith 1998) was analyzed as two clades (see below), for a total of 13 datasets analyzed (See Appendix

available online only at <http://dx.doi.org/10.1554/06-266.1.s1>). Phylogenetic estimates were obtained from the original papers.

For each dataset, events were ranked from a (first) to z (or earlier, depending on number of events) according to relative appearance in ontogeny for each taxon and entered into MacClade (Maddison and Maddison 1997) as unordered characters. Simultaneously occurring ranks were assigned multiple states. For example, if two events first occur at stage six in some taxon after three other events (ranks a, b, c) have already appeared, both these events are assigned both state d and state e.

#### *Test of the Early Conservation Model*

A test of relative lability of ontogenetic stages requires a quantitative measure of position in ontogeny. I measure position in ontogeny as position of a one-step change in an event's rank. That is, changes between first (a) and second (b) are earliest, between second (b) and third (c) are next, and so on. Only one-step changes (a → b, b → a, b → c, etc.) were used because it is not clear how best to summarize position in ontogeny for events that undergo large changes. For example, a change in rank between b and x occurs in both early and late ontogeny, yet summary measures of this change such as mean position ( $(b + x)/2 = m$ ) will tend to mark it as occurring in intermediate ontogeny. An alternative analysis using only "independent" changes (ab and cd but not bc) found identical results to using all one-step changes in the largest dataset (Jeffery et al. 2002; results not shown).

A test of the EC model requires a quantitative measure of evolutionary lability. Changes in rank were mapped onto the phylogenetic tree using Fitch optimization. I measure evolutionary lability at each position as (Poe and Wake 2004):

$$L = \frac{\text{[number of one-step evolutionary changes in rank]} - \text{[number of evolutionary stases in rank]}}{\text{[number of one-step evolutionary changes in rank]} + \text{[number of evolutionary stases in rank]}} \quad (1)$$

This measure is summed across all branches of the phylogeny for each ontogenetic position. For example, if three changes occur from rank a to rank b, two changes occur from rank b to rank a, rank a does not change on 10 branches, and rank b does not change on 11 branches,  $L$  is calculated for ontogenetic position ab as

$$(3 + 2 - 11 - 10)/(3 + 2 + 11 + 10) = -0.615. \quad (2)$$

For calculation of  $L$  (and testing of the AJ model; see below), average number of changes over all optimizations was used. This summary was used rather than the number of unambiguous changes in order to increase power for the smaller datasets, which have few unambiguous changes. I tested whether average number of changes is a suitable proxy for number of unambiguous changes by comparing the number of changes of each type under both summaries using Kendall's  $\tau$  and the largest dataset (Jeffery et al. 2002). These summaries were strongly correlated both for one-step changes at different stages in ontogeny ( $P < 0.0001$ ; used to calculate  $L$ ) and for changes of different amounts ( $P < 0.0001$ ; used

to compare AJ and UC models, see below), which suggests that average number of steps is a reasonable measure for these tests.

The equation  $L$  employs both the number of changes and the number of stases rather than just raw numbers of changes because uncorrected summaries are affected by the number of states of each type, and ties in rank (see Appendix available online) cause some ranks to be more common than others. For example, the number of states and the number of changes away from that state are correlated in the largest dataset (Jeffery et al. 2002;  $P < 0.01$ , Kendall's  $\tau$ ).

I tested for a relationship between ontogenetic position and  $L$  in each dataset using Kendall's  $\tau$ . A significant positive relationship is expected under the EC model.

Bininda-Emonds et al. (2003) pioneered the approach of comparing position in ontogeny to variance in ontogenetic rank. However, their approach has been shown to be statistically biased (appendix 1 in Poe and Wake 2004).

#### *Test of the Unconstrained Model versus the Adjacency Model*

I tested the AJ model against the UC model by comparing the number of steps change in rank to the scaled average frequency of that number on the phylogeny. Scaling was accomplished by dividing the average number of changes of a given type (one-step, two-step, etc.) by the number of ways for that type to occur. For example, in a dataset with 26 ranks (a → z), there are 50 ways for a one-step change to occur (e.g., c → d), 48 ways for a two-step change to occur (e.g., c → e), and two ways for a 25-step change to occur (a → z or z → a). Thus, the number of one-step changes was divided by 50, the number of two-step changes was divided by 48, and so on. This scaling is convenient because expected values under this scaling for each kind of change are 1.0 under an UC model in which ranks are randomized (see below). Using unscaled values had no effect on results (data not shown).

I used Kendall's  $\tau$  to assess the significance of a potential association between number of steps change and frequency of change. A significant negative relationship indicates rejection of the UC model.

## RESULTS

None of the datasets showed a significant positive relationship between evolutionary lability and position in ontogeny (Table 1), indicating no support for the EC model. Figure 1A shows detailed results for dataset 1; Figure 1B shows qualitative results for all datasets.

Eleven of 13 datasets show a significant negative relationship between amount of change in ontogenetic rank and scaled average number of evolutionary changes, indicating rejection of the UC model in favor of the AJ model (Fig. 2; Table 1).

## DISCUSSION

### *No Support for the Early Conservation Model*

The results do not support Von Baer's EC model of the conservation of early ontogeny. Although Von Baer's model has been advocated based on qualitative observational (e.g.,

TABLE 1. Summary of datasets analyzed in this paper. None of the datasets shows a statistically significant positive relationship for the test of Von Baer's law (one dataset [Nunn and Smith 1998] shows a significant negative relationship). All datasets show a significant negative relationship for the test of the adjacency (AJ) versus unconstrained (UC) models at  $P < 0.05$  without Bonferroni correction; 11 of 13 are significant at  $P < 0.05$  under Bonferroni correction.

| Dataset | Reference               | Data type              | Taxon             | $P$ , test of Von Baer's law | $P$ , test of AJ vs. UC model |
|---------|-------------------------|------------------------|-------------------|------------------------------|-------------------------------|
| 1       | Alberch et al. 1979     | ossification           | salamanders       | 0.58                         | <0.0001                       |
| 2       | Davies 1989             | ossification           | frogs             | 0.20                         | 0.0005                        |
| 3       | Jeffery et al. 2002     | bone, nerve, muscle    | amphibians        | 0.17                         | <0.0001                       |
| 4       | Jeffery et al. 2002     | bone, nerve, muscle    | mammals           | 0.45                         | <0.0001                       |
| 5       | Jeffery et al. 2002     | bone, nerve, muscle    | diapsids          | 0.48                         | <0.0001                       |
| 6       | Mabee and Trendler 1996 | ossification           | fish              | 0.76                         | <0.0001                       |
| 7       | Maisano 2002            | braincase ossification | lizards           | 0.80                         | 0.04                          |
| 8       | Maisano 2002            | carpal ossification    | lizards           | 0.30                         | 0.0018                        |
| 9       | Maisano 2002            | limb ossification      | lizards           | 0.59                         | <0.0001                       |
| 10      | Nunn and Smith 1998     | bone, nerve, muscle    | placental mammals | 0.07                         | <0.0001                       |
| 11      | Nunn and Smith 1998     | bone, nerve, muscle    | marsupial mammals | 0.0025                       | <0.0001                       |
| 12      | Strauss 1990            | ossification           | fish              | 0.63                         | 0.0005                        |
| 13      | Velhagen 1997           | ossification           | snakes            | 1.0                          | 0.04                          |

von Baer 1828) and theoretical (e.g., Wimsatt 1986) grounds, this model has not been tested quantitatively on several datasets until now. More work is needed on models of change in rank of anatomical developmental events and models that incorporate other aspects of development, but present evidence suggests that Von Baer's law does not accurately depict the evolution of development. Although counter to established dogma, this result should not be too surprising considering the body of work demonstrating that virtually any stage of ontogeny is susceptible to evolutionary change. Probably the best-known example is the extreme variation in blastula size and structure in vertebrates (Von Baer 1828; Haeckel 1874; Elinson 1987), but similarly great variations have been

demonstrated at other stages (Raff 1996). These differences often have been treated as noteworthy exceptions (e.g., Arthur 1997), but perhaps they represent the rule rather than the exception.

It could be argued that the alternative (AJ) model discussed here that fits the data better than the EC model is simpler than EC, and thus no mechanistic explanation of these results is necessary. That is, the finding that there is nothing special about early development may suggest that an explanation of ontogenetic diversity among species need not invoke special properties of ontogeny. However, the theoretical arguments of many authors in favor of EC are compelling (e.g., Wimsatt's [1986] developmental lock model), and Von Baer's law

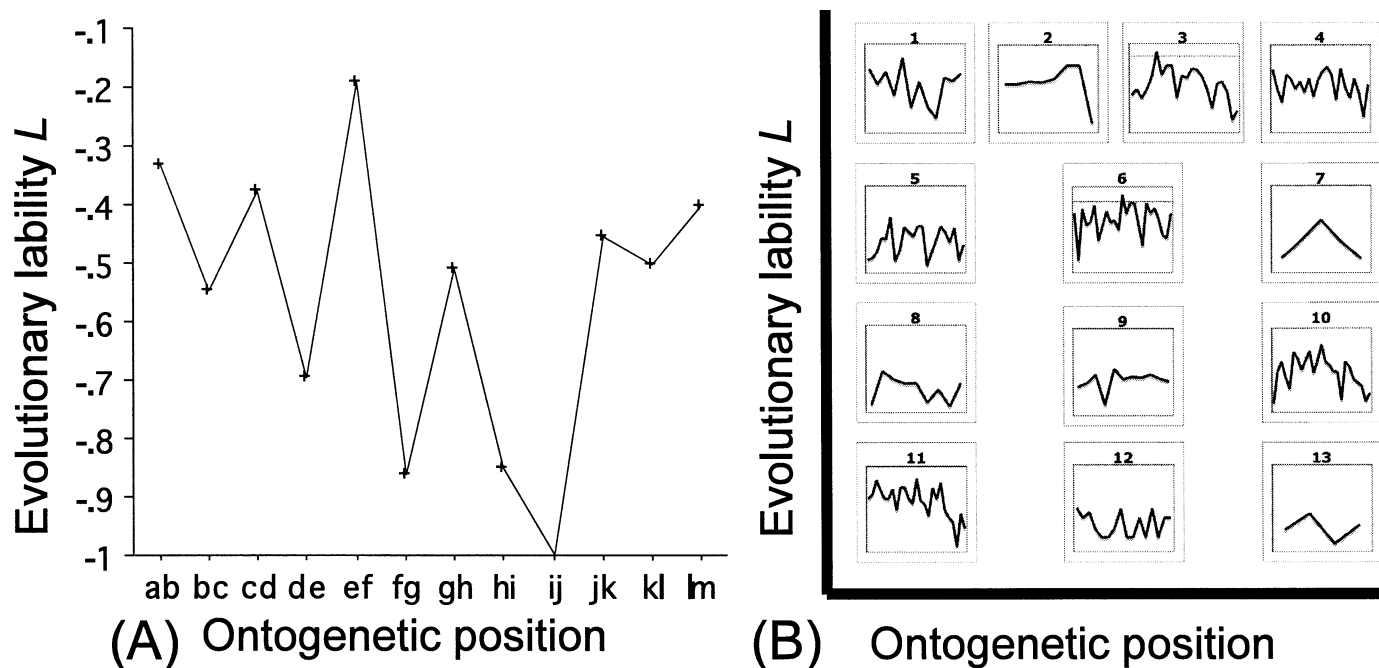


FIG. 1. Results of test of Von Baer's law comparing position in ontogeny to evolutionary liability in (A) dataset 1 (Alberch et al. 1979) and (B) datasets 1–13 (Table 1; see Appendix, available online). Graphs show that Von Baer's early conservation model is not supported—early ontogeny is not more evolutionarily conserved than late ontogeny—because there is no relationship between ontogenetic position (horizontal axes) and amount of evolutionary change  $L$  (vertical axes).

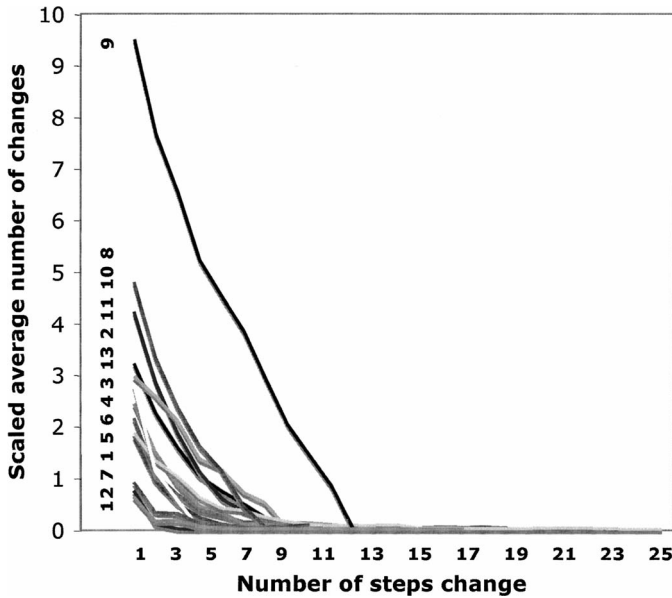


FIG. 2. Results of comparison of adjacency and unconstrained models. Downward slopes of lines indicate that small changes in rank occur more frequently than large changes, thus supporting the adjacency model. Numbers refer to datasets in Table 1.

may turn out to be supported under other tests or other types of data. Given these concerns, why would early development *not* be especially conserved in the cases tested here? Factors that may contribute to the possible constancy of evolutionary lability over ontogeny are an unrecognized resiliency of later ontogeny (as evidenced by, e.g., variation in larval stages of amphibians and echinoderms), the constancy of selective forces over ontogenetic time (as would be predicted by orthodox neodarwinian theory), and the modularity of development (creating relative independence of tissues at all stages of ontogeny).

*Rejection of the Unconstrained Model:  
What Model Is Preferred?*

The UC model was rejected in favor of the AJ model. However, this rejection does not lead to the conclusion that the changes tallied in the AJ model can be interpreted literally. The presence of, for example, five one-step changes does not mean that five independent instances of small evolutionary change have occurred. This interpretation is inappropriate because of the possibility of multiple changes and the nonindependence of ranks. Figure 3 shows an example where a single large evolutionary change is recorded as one large change and four smaller changes.

Although cases such as Figure 3 indicate that the counts of change in rank should not be interpreted as numbers of independent evolutionary events, they do not invalidate the rejection of the UC model or the use of this test. This can be seen by generating data according to a UC model and performing the test on these data. One formulation of the UC model is randomization of ranks. The expectation under the randomized UC model can be demonstrated analytically for simple cases. For example, for two taxa and three events, there are six potential sequences: 123, 321, 132, 213, 231,

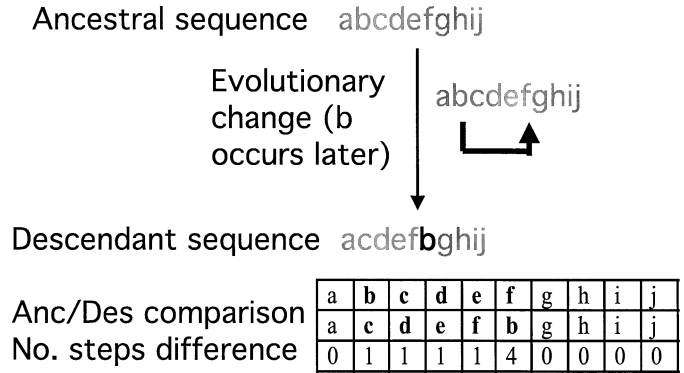


FIG. 3. Example case in which a single large change in ontogenetic rank is recorded as one five-step change and four one-step changes by test of the adjacency model used in this paper.

312. Under the randomized UC model, each of these sequences is equally likely, and we can get the expected values for scaled number of changes by summing over all possible combinations. For all 15 possible pairwise comparisons between taxa there are a total of 24 one-step differences and 12 two-step differences (e.g., the comparison of 123 and 231 shows two one-step differences and one two-step difference). There are four ways to differ by one step in rank (1 → 2, 2 → 1, 2 → 3, 3 → 2) and two ways to differ by two steps (1 → 3, 3 → 1). Scaling by the number of possibilities, 24/4 = 12/2. Thus, there is no relationship between number of steps change and scaled average number of changes under this model. Put another way, if all sequences are equally, likely then all types of changes are also equally likely once allowance is made for the number of ways to change. This result of no relationship (i.e., a flat regression line) is different from the strong negative relationship (i.e., downward sloping regression line) observed in the empirical datasets (Fig. 2).

This extreme version of the UC model—that is, complete randomization of ranks—is clearly rejected, but also is unrealistic for species that are not extremely distantly related. More realistic versions of UC would take time of divergence into account. For example, one could generate data such that events are equally likely to “move” large and small “distances” in ontogeny, with probability of movement for each event proportional to phylogenetic branch length. One could compare this UC model to an AJ model in which smaller

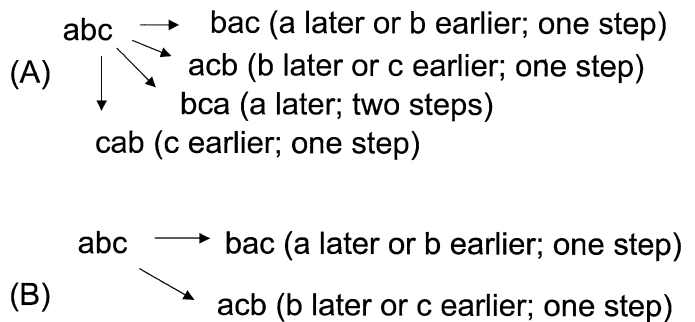


FIG. 4. Alternative time-dependent models for change in ontogenetic rank for ancestor with three ranked events. (A) All changes are possible in time  $t = 1$ . (B) Only one-step changes are possible in time  $t = 1$ .

TABLE 2. Probability of a descendant possessing each possible ranking at time  $t$  under UC-time model if ancestor has rank 123.

| Ranking        | Probability of rankings |         |         |         |         |         |         |         |         |
|----------------|-------------------------|---------|---------|---------|---------|---------|---------|---------|---------|
|                | $t = 0$                 | $t = 1$ | $t = 2$ | $t = 3$ | $t = 4$ | $t = 5$ | $t = 6$ | $t = 7$ | $t = 8$ |
| 123 (ancestor) | 1.000                   | 0       | 0.278   | 0.120   | 0.190   | 0.155   | 0.172   | 0.164   | 0.168   |
| 132            | 0                       | 0.333   | 0.111   | 0.194   | 0.154   | 0.173   | 0.164   | 0.168   | 0.166   |
| 213            | 0                       | 0.333   | 0.111   | 0.194   | 0.154   | 0.173   | 0.164   | 0.168   | 0.166   |
| 231            | 0                       | 0.167   | 0.139   | 0.181   | 0.158   | 0.171   | 0.164   | 0.168   | 0.166   |
| 321            | 0                       | 0       | 0.222   | 0.130   | 0.185   | 0.157   | 0.172   | 0.164   | 0.168   |
| 312            | 0                       | 0.167   | 0.139   | 0.181   | 0.158   | 0.171   | 0.164   | 0.168   | 0.166   |

changes in rank are more probable than larger changes. As a first attempt at such a comparison, I determined the exact probabilities of particular rankings of three events for a descendant given an ancestral sequence and two simple time-dependent models. These probabilities allowed calculation of the expected values for the AJ/UC test statistic of scaled average number of changes used in this paper. A model that I call "UC-time" allows a single event to shift position to any other position (Fig. 4A) in time  $t$ . Alternatively, the "AJ-time" model allows a single event to make only one-position changes (Fig. 4B) in time  $t$ . At each time  $t = 1, 2, 3, 4, 5, 6, 7, 8$  a change occurs with probability 1.0 according to these rules. Tables 2 and 3 show the probabilities of each possible ranking as a function of time  $t$  under these models, and Table 4 shows the expected percentage of one and two-step differences for these models.

Both sequences begin with a higher probability of one-step differences after one time unit and then converge quickly to complete randomization (still more realistic models would incorporate some probability of stasis; this would not affect the relevant points here). However, the intermediate stages differ between these two models in their comparison with the ancestral sequence in terms of the AJ/UC test statistic. In the UC-time model, two-step differences are more probable than one-step differences in some stages after scaling by number of ways to change. In the AJ-time model, two-step changes are never more probable than one-step changes. The results from data generated according to this simple AJ-time model thus conform slightly better to the pattern seen in the empirical datasets, in which one-step differences are much more frequent than two- or multistep differences (Fig. 2).

It is possible to compare these analytical results to empirical results by performing pairwise comparisons of sets of three events. I pared the *Xantusia* braincase dataset (Maisano 2002) down to the three events that vary in position across the five xantusid taxa and compared sequences between all 10 possible pairs of taxa. Average number of one-step dif-

ferences was 1.6 (89%); average number of two-step differences was 0.2 (11%). There were no cases where two-step differences were more common than one-step differences (i.e., 123 compared with 321). These data conform best to either the AJ-time model at  $t = 1$ , the UC-time model at  $t = 1$ , or the AJ-time model at intermediate stages.

These simulation, analytical, and empirical sampling exercises suggest that the randomized UC model can be rejected. The AJ-time model is slightly favored over the UC-time model, but much more work needs to be done on these and similar models.

#### Future Directions

The problems associated with using developmental ranks to test evolutionary hypotheses have been detailed both generally and with reference to the hypotheses of this paper (Smith 2001; Jeffery et al. 2002; Poe and Wake 2004) and will not be repeated here. Instead, I will focus on the prospects for additional development of tests like the ones used here and on a limitation of the sample of datasets analyzed.

Within the framework of analyzing hypotheses in terms of ontogenetic ranks, there is a need for development of stochastic models for character change on phylogenetic trees. Models like those in Figure 4 could be expanded to accommodate multiple species and events. Alternatively, existing models of character change (e.g., Garland et al. 1993) could be adapted to this context. That is, sets of quantitative characters could be evolved stochastically along branches according to, for example, a Brownian motion model and the resulting values ranked. Expected rankings resulting from either of these types of models could be used to evaluate empirical rankings probabilistically in a likelihood context.

With respect to testing Von Baer's EC model, an important component missing from the tests of this study is some allowance for differences in morphology during ontogeny. A great advantage of analyzing ranks of ontogenetic events is

TABLE 3. Probability of a descendant possessing each possible ranking at time  $t$  under AJ-time model if ancestor has rank 123.

| Ranking        | Probability of rankings |         |         |         |         |         |         |         |         |
|----------------|-------------------------|---------|---------|---------|---------|---------|---------|---------|---------|
|                | $t = 0$                 | $t = 1$ | $t = 2$ | $t = 3$ | $t = 4$ | $t = 5$ | $t = 6$ | $t = 7$ | $t = 8$ |
| 123 (ancestor) | 1.000                   | 0       | 0.500   | 0       | 0.375   | 0       | 0.344   | 0       | 0.336   |
| 132            | 0                       | 0.500   | 0       | 0.375   | 0       | 0.344   | 0       | 0.336   | 0       |
| 213            | 0                       | 0.500   | 0       | 0.375   | 0       | 0.344   | 0       | 0.336   | 0       |
| 231            | 0                       | 0       | 0.250   | 0       | 0.313   | 0       | 0.328   | 0       | 0.332   |
| 321            | 0                       | 0       | 0       | 0.250   | 0       | 0.313   | 0       | 0.328   | 0       |
| 312            | 0                       | 0       | 0.250   | 0       | 0.313   | 0       | 0.328   | 0       | 0.332   |

TABLE 4. Expected values for percentage of scaled (unscaled) one- or two-step differences from ancestor 123.

|               |            | $t = 0$ | $t = 1$   | $t = 2$ | $t = 3$ | $t = 4$ | $t = 5$ | $t = 6$ | $t = 7$ | $t = 8$ |
|---------------|------------|---------|-----------|---------|---------|---------|---------|---------|---------|---------|
| Randomization | % one-step | 0       | 50 (67)   | 50 (67) | 50 (67) | 50 (67) | 50 (67) | 50 (67) | 50 (67) | 50 (67) |
|               | % two-step | 0       | 50 (33)   | 50 (33) | 50 (33) | 50 (33) | 50 (33) | 50 (33) | 50 (33) | 50 (33) |
| UC-time       | % one-step | 0       | 75 (86)   | 39 (56) | 56 (71) | 47 (64) | 52 (68) | 49 (66) | 50 (67) | 50 (67) |
|               | % two-step | 0       | 25 (14)   | 61 (44) | 44 (29) | 53 (36) | 48 (32) | 51 (34) | 50 (33) | 50 (33) |
| AJ-time       | % one-step | 0       | 100 (100) | 50 (67) | 60 (75) | 50 (67) | 53 (69) | 50 (67) | 50 (67) | 50 (67) |
|               | % two-step | 0       | 0 (0)     | 50 (33) | 40 (25) | 50 (33) | 47 (31) | 50 (33) | 50 (33) | 50 (33) |

the comparability of the data across taxa, but this comparability is gained at a cost to generality. Von Baer formulated his law with reference to similarities and differences in morphology. The heterochronic differences evaluated in this study are only one kind of morphological difference. More general models will incorporate presence and absence of structures rather than just the timing of appearance of shared structures. It is not obvious how such models should be formulated. It seems that some measure of similarity of structure would be needed.

Probably the biggest limitation in the data used here is that all datasets include events taken only from relatively late in ontogeny. This may be a problem for some conceptions of Von Baer's model (e.g., those that identify constraint as occurring only during a particular period), but is not serious for the interpretation of that model used here (see also Poe and Wake 2004). I have operationalized Von Baer's model in a simple, general, testable form relating progression in ontogeny to a monotonic increase in evolutionary lability. Under this interpretation, any time slice of ontogeny is testable for Von Baer's model, including early, late, or any combination of these. This operationalization accords well with the most detailed qualitative articulation of Von Baer's model (Arthur 2004) but is not likely to satisfy all comparative embryologists. Regardless, though, data on ontogenetic ranks from much earlier in ontogeny would constitute a fuller test of Von Baer's model even using the test of this paper. Hopefully, such data will be forthcoming soon.

Finally, if the tests of this paper are to find some use they should be evaluated for their statistical properties. For example, I used arbitrary cut-offs of three taxa and five events in selecting datasets; a more rigorous approach would incorporate quantitative criteria regarding statistical power. Also, additional model-based evaluations like those detailed above (e.g., Fig. 4) would be worthwhile. That is, given some model of developmental change, what are the expectations for the tests of this paper?

### Conclusions

Von Baer's law is one of the oldest and most pervasive ideas in the evolution of development, and there are compelling theoretical reasons why early ontogeny is expected to be conserved. However, the many exceptions to this rule and the failure to detect a general trend in this study suggest that the alternative possibility—that no such trend exists—also should be entertained. Much further work is needed to determine what alternative model(s) should be favored.

### ACKNOWLEDGMENTS

I thank M. Wake, T. Lowrey, and anonymous reviewers for comments that substantially improved the paper.

### LITERATURE CITED

- Alberch, P., S. J. Gould, G. F. Oster, and D. B. Wake. 1979. Size and shape in ontogeny and phylogeny. *Paleobiology* 5:296–317.
- Arthur, W. 1997. The origin of animal body plans: a study in evolutionary developmental biology. Cambridge Univ. Press, Cambridge, U.K.
- . 2004. Biased embryos and evolution. Cambridge Univ. Press, Cambridge, U.K.
- 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. *Proc. R. Soc. Lond. B* 270:314–346.
- Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for existence. John Murray, London.
- Davies, M. 1989. Ontogeny of bone and the role of heterochrony in the myobatrachine genera *Uperoleia*, *Crinia*, and *Pseudophryne* (Anura: Leptodactylidae: Myobatrachinae). *J. Morphol.* 240:269–300.
- Elinson, R. P. 1987. Changes in developmental patterns: embryos of amphibians with large eggs. Pp. 1–21 in R. A. Raff and E. C. Raff, eds. *Development as an evolutionary process*. Alan Liss, New York.
- Garland, T., Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. *Syst. Biol.* 42:265–292.
- Haeckel, E. 1874. *Anthropogenie oder Entwicklungsgeschichte des Menschen*. Engelmann, Leipzig.
- Hume, D. 1739. *A treatise of human nature*. Reprinted 1978. Oxford Univ. Press, Oxford, U.K.
- Jeffery, J. E., M. K. Richardson, M. I. Coates, and O. R. P. Bininda-Emonds. 2002. Analyzing developmental sequences within a phylogenetic framework. *Syst. Biol.* 51:478–491.
- Mabee, P., and T. Trendler. 1996. Development of the cranium and paired fins in *Betta splendens* (Teleostei: Percomorpha): intra-specific variation and interspecific comparisons. *J. Morphol.* 227:249–287.
- Maddison, W. P., and D. R. Maddison. 1997. *MacClade*. Ver. 3.07. Sinauer, Sunderland, MA.
- Maisano, J. A. 2002. Postnatal skeletal ontogeny in five xantusiids (Squamata: Scleroglossa). *J. Morphol.* 254:1–38.
- 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.
- Poe, S., and M. H. Wake. 2004. Quantitative tests of general models for the evolution of development. *Am. Nat.* 164:415–422.
- Raff, R. A. 1996. *The shape of life: genes, development, and animal form*. Univ. of Chicago Press, Chicago.
- Smith, K. K. 1996. Integration of craniofacial structures during development in mammals. *Am. Zool.* 36:70–79.
- . 2001. Heterochrony revisited: the evolution of developmental sequences. *Biol. J. Linn. Soc.* 73:169–186.
- Strauss, R. E. 1990. Heterochronic variation in the developmental

- timing of cranial ossifications in poeciliid fishes (Cyprinodontiformes). *Evolution* 44:1558–1567.
- Vellhagen, W. A. 1997. Analyzing developmental sequences using sequence units. *Syst. Biol.* 46:204–210.
- Von Baer, K. E. 1828. *Über Entwicklungsgeschichte der Thiere: Beobachtung und Reflexion*. Vol. 1. Bornträger, Königsberg, Russia.
- Wimsatt, W. C. 1986. Developmental constraints, generative entrenchment, and the innate-acquired distinction. Pp. 185–208 in P. Bechtel ed. *Integrating scientific disciplines*. Martinus-Nijhoff, Dordrecht, The Netherlands.
- Wray, G. A., and D. R. McClay. 1989. Molecular heterochronies and heterotopies in early echinoid development. *Evolution* 43: 803–813.

Corresponding Editor: D. Pfennig