© 2020. Published by The Company of Biologists Ltd | Development (2020) 147, dev178004. doi:10.1242/dev.178004

SPOTLIGHT

# Null hypotheses for developmental evolution

**Samuel H. Church[1,*] and Cassandra G. Extavour[1,2,*]**


## ABSTRACT

How much evolutionary change in development do we expect? In this Spotlight, we argue that, as developmental biologists, we are in a prime position to contribute to the definition of a null hypothesis for developmental evolution: in other words, a hypothesis for how much developmental evolution we expect to observe over time. Today, we have access to an unprecedented array of developmental data from across the tree of life. Using these data, we can now consider development in the light of evolution, and vice versa, more deeply than ever before. As we do this, we may need to re-examine previous assumptions that appeared to serve us well when data points were fewer. Specifically, we think it is important to challenge assumptions that change is very rare for all developmental traits, especially if this assumption is used to sustain an erroneous view that evolution always optimizes adaptive traits toward increasing complexity.


**KEY WORDS: Ancestral, Convergence, Derived, Homology, Null hypothesis, Phylogenetic comparative methods**

## Introduction

Modern developmental biology is firmly grounded in evolutionary biology. The principle that all cellular life shares a common ancestor underlies our hypotheses about how the genome interacts with non-genetic parameters to direct cell behavior during development, homeostasis and disease. When we choose model organisms to study development, we often base these choices not only on ease of laboratory maintenance, but also on their evolutionary relationships to other organisms. We do this because we believe that different organisms might teach us different things about the processes we study. Furthermore, we frequently interpret our results by asking to what extent the processes we observe in one organism are likely to be operative or 'conserved' in others. When we perform, for example, a gene ontology (GO) annotation to infer the functions of genes, we are hypothesizing that molecular functions experimentally demonstrated in one organism can predict the untested functions in other organisms, based on an assumption of evolutionary conservation.

At this stage in our field, it is appropriate to ask whether we are robustly applying evolutionary assumptions to the study of development. It has been 125 years since the publication of the first volume of *Archiv für Entwicklungsmechanik der Organismen*, a foundational journal that in some sense marked the rise of an experimental, rather than observational, approach to understanding development (Counce, 1997). A more recent conceptual shift came with the establishment of 20th century 'Evo-Devo': the formalization of the concept that the process of development is a crucial component of the evolutionary process, because it generates phenotypes upon which natural selection may act (Raff, 2012). Increasing numbers of taxa are now emerging as 'new' model organisms, providing both an opportunity and a challenge: how to compare data collected from developmental experiments across organisms, while considering their evolutionary history.

Developmental biologists make these comparisons because we are often interested both in developmental mechanisms, and in the similarities and differences between these mechanisms and their counterparts in other organisms. Evolutionary developmental biologists make this query explicit: we try to determine the evolutionary origins and dynamics of these mechanisms. We frequently approach this by appealing to a hypothesized ancestral state, and use this hypothesis to propose the direction of evolutionary change, distinguishing between 'ancestral' and 'derived' mechanisms. Testing these hypotheses is challenging because the amount of data we have access to, although growing, pales in comparison with the unobserved data from taxa not yet studied. Nevertheless, we believe there are fruitful ways for developmental biology to place its findings in an increasingly robust evolutionary context.

With this Spotlight, our goal is to highlight recent advances in the methodological rigor of evolutionary biology that are highly relevant to the study of development. Specifically, we suggest the approach of defining a null hypothesis – a hypothesis for how much developmental difference we would expect to observe between organisms – by using evolutionary and developmental data to inform our estimates of relevant parameters (Table 1). With this framework, we encourage comparative developmental biologists to interrogate the assumptions we invoke when formulating hypotheses, designing experimental approaches and interpreting results. We believe this approach will help us refine our questions, from 'How are developmental mechanisms different across taxa?' to 'Are developmental mechanisms more or less different than we would expect?'.

## Inferring evolutionary change for developmental characters

Comparisons of developmental characters often share a few qualities that can make evolutionary inferences challenging. These comparisons, e.g. between fruit flies and mice, frequently span hundreds of millions of years in evolution (Jenner, 2006). This is due to the historic choice of taxa to study in developmental biology, as well as an intentional choice of lineages likely to capture variation in developmental features (Jenner and Wills, 2007). In addition, these comparisons may include very few organisms (often fewer than ten taxa, possibly as few as two). This is generally because collecting developmental data is expensive, and requires substantial expertise as well as lab-amenable organisms.

Making an evolutionary inference over deep divergences and using few taxa is a notoriously difficult type of analysis (Freckleton, 2009; Wagner et al., 2000). This is not to say that inferences for developmental data are impossible, rather that care must be taken to avoid common pitfalls and to not over-interpret the results (Swofford and Maddison, 1992). In cases where data points are

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [2]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA.

*Authors for correspondence (church@g.harvard.edu; extavour@oeb.harvard.edu)

S.H.C., 0000-0002-5499-523X; C.G.E., 0000-0003-2922-5855

Development (2020) 147, dev178004. doi:10.1242/dev.178004

**Table 1. Establishing a null hypothesis for trait evolution: how similar or different do we expect traits to appear across taxa?**

| Relevant questions | Source of insight |
|---|---|
| How long since taxa diverged? | Time-calibrated phylogeny |
| How quickly are traits evolving? | • Trait variation across species, mapped onto a phylogeny (estimate of evolutionary rate).<br>• Trait variation within species (e.g. observed in biological replicates in a laboratory assay).<br>• Other data that might provide insight into the likelihood of trait change (e.g. sequence conservation and degree of pleiotropy). |
| How precisely can traits be measured? | Technical replicates in laboratory assays |

few, we are tempted to use the method perceived to be the most simple or intuitive. One commonly used approach is parsimony, a method of inference that, in its most strict application, seeks to minimize the total number of evolutionary changes required to explain extant distributions of traits (Joy et al., 2016). However, there are risks associated with this application, of which we consider two to be the most crucial.

### Parsimony typically provides no measure of confidence for an evolutionary inference

Estimates of parsimony support do not have well-defined statistical uncertainties (Joy et al., 2016). When data points are few, this can lead to an artificial sense of certainty in the ancestral state and the distribution of evolutionary changes. This can be especially prevalent in Evo-Devo comparisons, where the most parsimonious inference of evolutionary history is often regarded as the only possibility. Instead, we suggest that considering the time scale of divergence in question, the number of taxa observed, the number of missing taxa and the uncertainty in observation should all be used to assess the certainty of our interpretation.

### It is not always trivial to apply strict parsimony without violating its underlying assumptions

In its most fundamental form, a strict parsimonious approach seeks to reduce the number of evolutionary transitions, does not favor any direction of transition over another and does not take into consideration the timing of divergence between taxa (i.e. branch lengths) (Joy et al., 2016). When comparing developmental traits in particular, it can be difficult to avoid violating these assumptions (Cunningham et al., 1998).

For example, consider the scenario shown in Fig. 1. Here, we depict a hypothetical study in which the function of a gene has been investigated in two model organisms: the mouse *Mus musculus* and the zebrafish *Danio rerio*. In the mouse, knockdown assays of the gene result in decreased neural activity in the adult brain, and defects in limb bud formation. In the fish, knock down of an ortholog of this gene also decreases adult brain neural activity, but limb formation is unaffected.

With these data, there are two equally parsimonious hypotheses of evolutionary change. One hypothesis is that, in the shared ancestor of the taxa, this gene was involved in neural function, and that, along the lineage leading to *M. musculus*, limb function was gained. Alternatively, in the shared ancestor, this gene may have had a function in both neural activity and limb formation; and along the lineage leading to *D. rerio,* limb function was lost. If we fail to acknowledge both scenarios in our interpretation, we have not correctly applied a strict parsimonious approach, because we have violated one of its fundamental assumptions. As strict parsimony is agnostic about the evolutionary process, it does not favor transitions in one direction over another; gains and losses are considered equally likely. Thus, comparisons of only two taxa with divergent characters can never produce enough information to reliably infer the ancestral state, based on strict parsimony alone and in the absence of other information.

Why might our interpretations fail when we attempt to use strict parsimony to analyze developmental trait evolution? First, we might apply assumptions that are not grounded in evolutionary theory. For example, if we erroneously think the evolutionary process is always a 'ladder toward increasing complexity' or if we ignore non-adaptive contributions to evolutionary change, we may have an unintentional bias toward interpretations that favor trait gains over losses, especially when those gains are on lineages leading to well-studied organisms (Dunn et al., 2015). Similarly, a misrepresentation or misinterpretation of the phylogeny can lead to the incorrect perception that some extant taxa are more closely related to the common ancestor than to other extant descendants (e.g. that certain extant taxa are 'basal' or that certain lineages are 'primitive') (Crisp and Cook, 2005; Grandcolas et al., 2014; Jenner, 2006). Both of these assumptions have roots in the incorrect view of evolution as a universally directional progression of optimization. In reality, it is important to keep in mind that all extant taxa have been evolving for the same amount of time, and that all lineages have experienced both gains and losses of traits (Strathmann and Eernisse, 1994). Thus, their descendants – including favorite laboratory organisms – will inevitably show combinations of both ancestral and derived characters (Jenner, 2006).

Alternatively, we may find it useful to propose hypotheses based on data collected during developmental experiments, but that

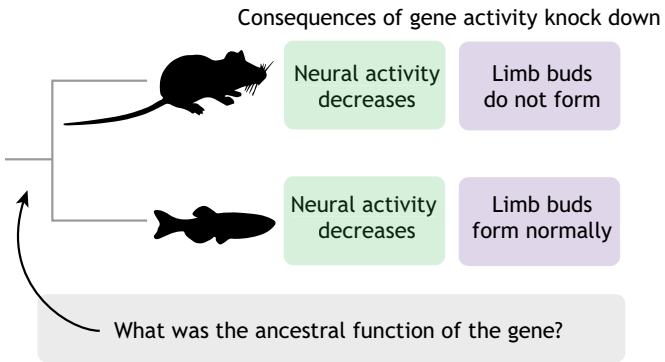


**Fig. 1. Hypothetical comparison of gene function between two vertebrate species.** In this example, a gene of interest has been studied in two model organisms: the mouse *Mus musculus* and the zebrafish *Danio rerio*. There are two equally parsimonious explanations for the results: (1) in the common ancestor, the gene had a neural function, and a new function in specifying the limb bud was acquired on the lineage leading to the mouse; or (2) in the common ancestor, the gene had both a neural and limb bud function, and the latter was lost on the lineage leading to the fish. In the absence of additional information, ancestral states are always ambiguous when comparing only two species with divergent traits.

Development (2020) 147, dev178004. doi:10.1242/dev.178004

deviate from strict parsimony. For example, in the scenario described above, we might have data on the functions of similar genes in other vertebrates, showing that transitions from a neural function to limb bud function are commonly observed, but that transitions from limb bud to neural function rarely occur. If a pattern like this is supported by data, then we could be justified in using a method such as weighted parsimony or a model-based approach that invokes unequal transition frequencies (Schultz et al., 1996).

## Considering rate, reversals and convergence

We can expand the repertoire of approaches we use to compare developmental data across species by incorporating ideas from the field of phylogenetic comparative methods, which has developed a robust literature on models of character evolution (Felsenstein, 1985; Garland and Ives, 2000; Lewis, 2001; Swofford and Maddison, 1992; Tarasov, 2019). These include methods of inference that allow for unequal probabilities of transitions between character states (Cunningham et al., 1998; Swofford and Maddison, 1992), changes in the rate of trait evolution (Rabosky et al., 2014), uncertainty in our observations (Garland and Ives, 2000), missing characters and missing taxa (Maddison, 1993), all of which are common challenges faced by developmental biologists. For discrete characters, many of these methods have their roots in a Markov framework of evolutionary change. Under this framework, the probability of change depends only on the current state, and can be calculated using a matrix that describes the rate at which evolutionary transitions occur (Swofford and Maddison, 1992). Even if we choose not to employ these methods mathematically, we can incorporate the theoretical underpinnings of a transition rate-based approach into our evolutionary interpretations of developmental data.

Consider the example shown in Fig. 2. In this hypothetical scenario, we have generated new data on the expression of a transcription factor during comparable developmental stages across three taxa. In studies of a fly and a beetle (*Drosophila melanogaster* and *Tribolium castaneum*, respectively), both express the gene during stage 1 of development but not during stage 2, whereas another fly (*Clogmia albipunctata*) shows the reverse pattern of expression. However, we are not limited to making inferences based only on our new gene expression data: from previous studies, we also know the following salient details about these organisms and their evolution: (1) the amount of shared evolutionary history that we are considering when comparing these three taxa is ~750 million years (the total sum of all shared branch lengths); (2) these three taxa are a very small sample of the set of all extant insect species, the shared evolutionary history of which sums up to billions of years; and (3) the results suggest that at least one change in expression timing has occurred within our relatively small sample of evolutionary history; there may have been more than one change (under a non-parsimonious hypothesis), but not zero changes.

We might be tempted to hypothesize that the shared expression during stage 1 in *T. castaneum* and *D. melanogaster* represents the ancestral timing of expression in the most recent common ancestor. Under this hypothesis, expression during stage 2 in *C. albipunctata* would be the derived state. However, our data could also be described by multiple alternative scenarios (Fig. 2; Table 2). Deciding whether any of these scenarios is better supported than the others requires us to examine our *a priori* assumptions about the tempo of developmental evolution.

It might be that across the entire insect tree, there was exactly one change in the timing of expression since the common ancestor (Fig. 2, scenario A), and our study happened to sample that single change along the branch leading from the common ancestor of flies

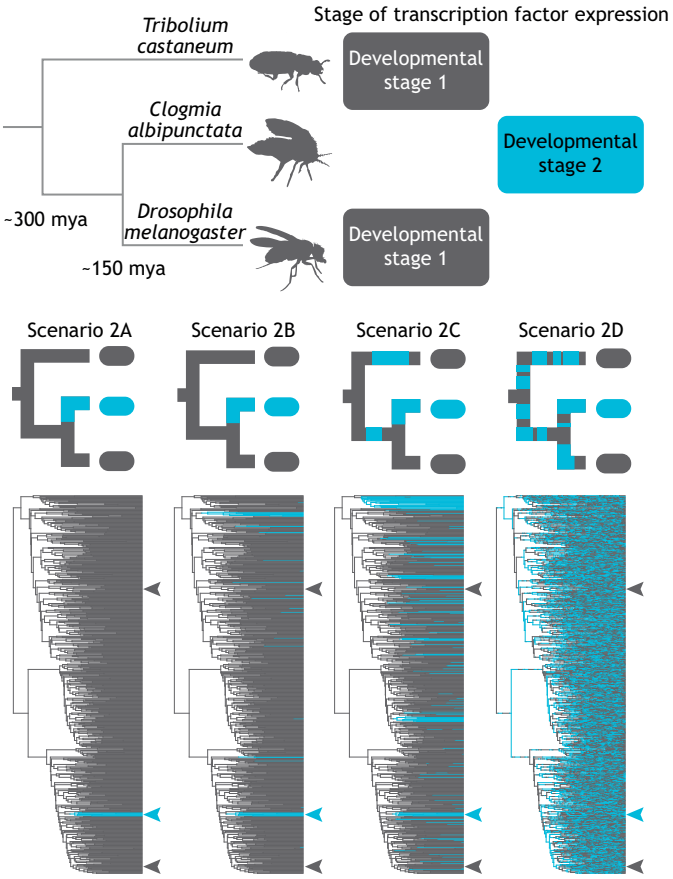


**Fig. 2. Hypothetical comparison of gene expression timing across three insect species.** In this example, the timing of expression of a homologous transcription factor has been studied in three insects: the beetle *Tribolium castaneum*, and the flies *Clogmia albipuncta* and *Drosophila melanogaster*. The observed data for these species could be explained by four different evolutionary scenarios, 2A-D. In these scenarios, we show the evolutionary history of this trait shared by the three taxa sampled here (upper trees with three tips) and the history shared by all extant insects (lower trees with many tips). Each scenario differs in the rate of evolutionary change of the trait of interest (here, the timing of gene expression), from a very low rate of change in 2A to a high rate in 2D (see Table 2). Arrowheads point to sampled taxa. This example assumes a binary model of trait evolution; see section on 'Character state space' for a discussion on the assumptions about non-binary character state space.

to *C. albipunctata*. In this case, the observation of one change in the sample size of 750 million years is not representative of the broader pattern of evolution of this trait. In other words, our study happened to sample a very rare change by chance, even with a sample size that provided a very small window into the total evolutionary history of insects.

Alternatively, we can consider scenarios in which our sample of evolutionary history is representative of the broader distribution (Fig. 2, scenarios B-D). In scenario 2B, one change per 750 million years is approximately representative of the rate of character transitions for the rest of the tree. Given that the full tree contains billions of years of total evolution (summing over all branch lengths), this scenario predicts many more than the single change hypothesized in scenario 2A. We note that, when comparing the three study taxa only, the interpretations proposed by scenarios 2A and 2B are identical (Fig. 2A,B, upper trees), and both are equally parsimonious; it is only when we expand our view to include the

Development (2020) 147, dev178004. doi:10.1242/dev.178004

**Table 2. Comparing the four possible evolutionary scenarios shown in** Fig. 2

| Conditions | Scenario 2A | Scenario 2B | Scenario 2C | Scenario 2D |
|---|---|---|---|---|
| Rate of evolutionary change | Very low | ~1/750 million years | > 1/750 million years | High |
| Is the observed sample of evolutionary history (upper trees) representative of the broader distribution across the entire tree (lower trees)? | No | Yes | Yes | Yes |
| Convergence and reversals | Did not occur | Are infrequent | Are frequent | Are very frequent |
| Ancestral state given the data | Unambiguous | Relatively certain | Relatively uncertain | Completely ambiguous |

unobserved taxa in the full tree of extant insects (Fig. 2A,B, lower trees) that the difference in assumptions underlying these becomes apparent.

In scenario 2C, the rate of character transitions is greater than one change per 750 million years. This scenario predicts that our observed data are explained by more than one change, and therefore includes some instances of either convergence (independent evolution of the same timing of gene expression) or reversal to the ancestral state of pathway activity after evolution to a derived state. This scenario is not strictly parsimonious, but is entirely plausible for many developmental traits.

Finally, in scenario 2D, convergence and reversals occur at such a high rate that they are the rule rather than the exception. Under this scenario *T. castaneum* and *D. melanogaster* share the same character state simply by chance. In both scenarios 2C and 2D, because a large and unknown number of changes have occurred, our ability to infer the ancestral state comes with a considerable amount of uncertainty.

In the absence of additional information beyond our observed expression data, the evolutionary history described above suggests that scenario 2A is less likely than the others. This is because 2A is the only scenario that requires the interpretation that our small sample of evolutionary history (Fig. 2A, upper tree) happens to include a change that is very rare with respect to the entire tree (Fig. 2A, lower tree). In other words, the variation that is present even in the small sample we have suggests that the rate of evolution is greater than the near-zero rate that would be required in scenario 2A. Scenarios 2B-D seem more consistent with the information we have gathered. However, distinguishing between them is difficult because our small sample size cannot tell us whether the rate of transitions between character states (in this case, the timing of expression) is low, high or somewhere in between.

Given the depth of divergence and the small sample, we might despair at the thought of being able to draw any inferences about the ancestral state with these data, because it might seem that almost any possible evolutionary rate is plausible. However, we argue that developmental biology is precisely the discipline that can help us make this distinction. We can contextualize our prediction about the frequency of evolutionary change by considering data collected in laboratory studies of development, including the nature and complexity of the transcriptional regulatory machinery, the degree of genomic sequence conservation across taxa, the degree of pleiotropy and the amount of standing variation in gene expression within a population. These data could help guide our ability to suggest one scenario as being more likely than the others. For example, if we observe that expression timing is variable within a population, is controlled by few regulators and that the sequences of these regulators have little conservation across taxa, then a high rate of evolutionary change in expression timing (scenario 2D) becomes increasingly plausible.

## Character state space

The examples we have considered thus far describe traits as having one of two possible states: either functional during a developmental process or not (Fig. 1); or expressed during one of two developmental stages (Fig. 2). In reality, a character like the timing of gene expression will have many possible states that cannot be captured in a single binary comparison. Given this, we should think critically about how we describe character state space, which is the set of possible transitions between character states that can be realized during evolution (Hoyal Cuthill, 2015; Swofford and Maddison, 1992).

In a transition rate matrix format, the character state space is determined by the number of rows and columns, representing the number of possible evolutionary changes from any single state (e.g. the Mk model) (Lewis, 2001). Thus, for the example in Fig. 2, the full range of evolutionary possibilities would be best described by an expanded matrix that allows for a larger number of transitions, such as expression in earlier or later stages, or the loss of expression of those genes entirely. Considering an expanded set of possible transitions can help us appreciate that if the character space is large enough and the rate of shifts is small but non-zero, it is plausible that the common ancestor could have shown a character state not observed in our extant taxa at all.

Errors in inference can result if we limit ourselves to constructing binary state comparisons based on similarities or differences to a presumed ‘ground state’, especially if that ground state is based on a biased view of animal complexity (Beaulieu et al., 2013). For example, it may be tempting to assign the ‘ground state’ to the organism first studied or most well studied, and to then compare all other organisms to this standard. Such a comparison, however, can mask the hidden differences in unobserved states among non-model organisms, painting a skewed view of evolution that appears to show ‘increasing complexity’ toward well-studied lineages for which we have more observations (Dunn et al., 2015).

We therefore suggest that keeping in mind the likely existence of developmental character states that are thus far unobserved may improve experimental design and data interpretation. Phylogenetic comparative approaches that take into account unobserved or hidden states are one of the most active frontiers of research in evolutionary methods (Beaulieu et al., 2013; Tarasov, 2019; Uyeda et al., 2018). This is particularly relevant to the fields of development and Evo-Devo, as complex character structures involving ordered matrices, hidden states and hierarchical interactions between transitions likely describe the evolution of most developmental mechanisms (Beaulieu et al., 2013; Tarasov, 2019). These comparative approaches seek to reduce the impact on inference of over-simplistic models, where multiple character states are unintentionally reduced to one state, because some states are unobserved at the time of analysis (Beaulieu et al., 2013; Tarasov, 2019). They also seek to resolve epistatic interactions between character states, applicable in cases such as when an evolutionary change resulted in the genomic loss of pathway

Development (2020) 147, dev178004. doi:10.1242/dev.178004

components in one lineage but not in another (Maddison, 1993; Tarasov, 2019). A major objective in the development of these approaches has been to create a framework in which developmental insights, such as the descriptions of gene regulatory networks, can be incorporated mathematically into evolutionary comparisons (Tarasov, 2019).

## Establishing null hypotheses of developmental evolution

The key to distinguishing between the scenarios posed above is answering the question of how much evolutionary change we expect for developmental traits. When taxa diverge from a common ancestor, they will retain some shared ancestral characteristics and evolve to become different in others. Simply observing differences between taxa is not in itself surprising (Jenner, 2006). We suggest that expanding the question from 'Are these traits the same or different between taxa?' to 'Are these traits more or less different between taxa than we would expect?' leads to more interesting and informative results.

Phylogenetic comparative methods answer the latter question by testing findings against a null hypothesis that describes the amount of variation we expect to see by chance (Gittleman and Kot, 1990), thus serving as a contextual comparison for observed variation. Even when data points are too few for formal modeling and simulation, we believe that the conceptual foundations of this approach are applicable to developmental studies.

A good example is found in developmental comparisons that aim to test the homology of structures or cell types across taxa. In these comparisons, evidence of homology can include similar expression patterns of marker genes, or similar regulatory relationships of homologous genes (Wagner et al., 2000). In other words, if the development of these structures is regulated by the same genes working together in the same way, this is sometimes considered evidence that they are derived from the same structure in the common ancestor.

Generating a null hypothesis in this case requires considering what we know about the processes that produced the observed data, including four crucial parameters (see Table 1): (1) the time since our studied taxa diverged from a common ancestor; (2) the frequencies at which our structures or cell types of interest arise *de novo*, and at which they are lost; (3) the frequency of evolutionary change in the expression and function of the genes that regulate structural development; and (4) the uncertainty in determining gene expression or function using laboratory assays.

We find that, for many comparisons of homology, we implicitly assume that all of these parameters are very low. In other words, we assume that the time since divergence has been short enough, the rates of evolution are low enough and the uncertainty in our observations is small enough, such that any observation of shared gene expression is good evidence of homology. Although it may not always be possible to quantify each of these parameters numerically, we should be aware that a default assumption of low parameter values can mislead evolutionary inferences. Instead of this default, we might consider any data we have in hand that suggest this assumption is not valid. If we have reason to believe that the time since divergence is large, rates of structure gain or loss are high, or our assays have a high rate of error, we may conclude that it is very possible to observe convergent expression patterns by chance (Jenner, 2006; Moczek, 2008). Similarly, if we have reason to believe the divergence time is large or the rate of expression change is high, we might consider it increasingly plausible that we would not be able to detect similar gene expression between taxa, even if the extant trait under study were in fact derived from a common ancestral trait (Jenner, 2006).

In some cases, parameters 1 and 2 can both be estimated using data from evolutionary studies. For example, we can estimate the time of divergence (parameter 1) using a time-calibrated phylogeny. Data on the presence and absence across of similar structures or cell types are often available from many more taxa than those we study in our laboratories, even if the homology of these structures has not yet been ascertained. This presents an opportunity to evaluate their phylogenetic distribution. For example, if many closely related species differ in the presence of a structure of interest, this suggests that the rate of gain, loss or convergence of the developmental processes that generate that structure is relatively high (parameter 2).

In terms of parameters 3 and 4, developmental biologists are well positioned to provide the data necessary to estimate these parameters. For example, *in situ* hybridization assays used to assess gene expression rarely yield a single, invariant expression pattern that is identical across all samples. Instead, we are more likely to observe a range of patterns that show some spatial or temporal variation. This likely reflects both the standing variation of gene expression within a population and technical variance in the assay. If the variation due to either one of these is high, the probability of observing any pattern by chance increases. Using these principles to reconsider our assumptions, we can now evaluate comparisons across organisms, not just on the basis of individual experimental outcomes, but against a null hypothesis informed by evolutionary insights and data-driven experience in developmental biology.

## Perspectives

Methods for collecting large amounts of developmental data are now more advanced and applicable to a broader range of organisms than ever before, and previously intractable nodes in the evolutionary tree are gaining resolution. We can enhance progress towards the goal of considering development in the light of evolution, by interrogating the assumptions of the evolutionary models we invoke, asking what data we have to support their use, and explicitly incorporating these assumptions and data into our discussions of evolutionary inference. One way to accomplish this is to apply the rich set of models, approaches and conceptual frameworks that have been developed in the field of phylogenetic comparative methods. These concepts and tools can help us establish robust null expectations of developmental evolution that we can use to contextualize our results. By grounding developmental comparisons in evolutionary methods, we can leverage this era of new data to answer some of our oldest questions.

### Acknowledgements

We thank Seth Donoughe, Bruno de Medeiros and members of the Extavour Lab for discussion of ideas, and the anonymous reviewers for comments that helped to improve the manuscript.

### Funding

The authors' research is supported by the National Science Foundation (IOS-1257217 to C.G.E.) and by the National Science Foundation Graduate Research Fellowship Program (DGE1745303 to S.H.C.)

### References

**Beaulieu, J. M., O'Meara, B. C. and Donoghue, M. J.** (2013). Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Syst. Biol.* **62**, 725-737. doi:10.1093/sysbio/syt034

**Counce, S. J.** (1997). Archives for developmental mechanics. In *Landmarks in Developmental Biology 1883-1924* (ed. W. Roux), pp. 94-108. Springer.

**Crisp, M. D. and Cook, L. G.** (2005). Do early branching lineages signify ancestral traits? *Trends Ecol. Evol.* **20**, 122-128. doi:10.1016/j.tree.2004.11.010

**Cunningham, C. W., Omland, K. E. and Oakley, T. H.** (1998). Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* **13**, 361-366. doi:10.1016/S0169-5347(98)01382-2

Development (2020) 147, dev178004. doi:10.1242/dev.178004

**Dunn, C. W., Leys, S. P. and Haddock, S. H. D.** (2015). The hidden biology of sponges and ctenophores. *Trends Ecol. Evol.* **30**, 282-291. doi:10.1016/j.tree.2015.03.003

**Felsenstein, J.** (1985). Phylogenies and the comparative method. *Am. Nat.* **125**, 1-15. doi:10.1086/284325

**Freckleton, R. P.** (2009). The seven deadly sins of comparative analysis. *J. Evol. Biol.* **22**, 1367-1375. doi:10.1111/j.1420-9101.2009.01757.x

**Garland, T., Jr and Ives, A. R.** (2000). Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.* **155**, 346-364. doi:10.1086/303327

**Gittleman, J. L. and Kot, M.** (1990). Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst. Zool.* **39**, 227-241. doi:10.2307/2992183

**Grandcolas, P., Nattier, R. and Trewick, S.** (2014). Relict species: a relict concept? *Trends Ecol. Evol.* **29**, 655-663. doi:10.1016/j.tree.2014.10.002

**Hoyal Cuthill, J. F.** (2015). The morphological state space revisited: what do phylogenetic patterns in homoplasy tell us about the number of possible character states? *Interface Focus* **5**, 20150049. doi:10.1098/rsfs.2015.0049

**Jenner, R. A.** (2006). Unburdening evo-devo: ancestral attractions, model organisms, and basal baloney. *Dev. Genes Evol.* **216**, 385-394. doi:10.1007/s00427-006-0084-5

**Jenner, R. A. and Wills, M. A.** (2007). The choice of model organisms in evo–devo. *Nat. Rev. Genet.* **8**, 311-314. doi:10.1038/nrg2062

**Joy, J. B., Liang, R. H., McCloskey, R. M., Nguyen, T. and Poon, A. F. Y.** (2016). Ancestral reconstruction. *PLoS Comput. Biol.* **12**, e1004763. doi:10.1371/journal.pcbi.1004763

**Lewis, P. O.** (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913-925. doi:10.1080/106351501753462876

**Maddison, W. P.** (1993). Missing data versus missing characters in phylogenetic analysis. *Syst. Biol.* **42**, 576-581. doi:10.1093/sysbio/42.4.576

**Moczek, A. P.** (2008). On the origins of novelty in development and evolution. *BioEssays* **30**, 432-447. doi:10.1002/bies.20754

**Rabosky, D. L., Donnellan, S. C., Grundler, M. and Lovette, I. J.** (2014). Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Syst. Biol.* **63**, 610-627. doi:10.1093/sysbio/syu025

**Raff, R. A.** (2012). *The Shape of Life: Genes, Development, and the Evolution of Animal form*. University of Chicago Press.

**Schultz, T. R., Cocroft, R. B. and Churchill, G. A.** (1996). The reconstruction of ancestral character states. *Evolution* **50**, 504-511. doi:10.1111/j.1558-5646.1996.tb03863.x

**Strathmann, R. R. and Eernisse, D. J.** (1994). What molecular phylogenies tell us about the evolution of larval forms. *Am. Zool.* **34**, 502-512. doi:10.1093/icb/34.4.502

**Swofford, D. L. and Maddison, W. P.** (1992). Parsimony, character-state reconstructions, and evolutionary inferences. In *Systematics, Historical Ecology, and North American Freshwater Fishes (*ed. R. L. Mayden*)*, pp. 186-223. Stanford University Press.

**Tarasov, S.** (2019). Integration of anatomy ontologies and evo-devo using structured Markov models suggests a new framework for modeling discrete phenotypic traits. *Syst. Biol.* **68**, 698-716. doi:10.1093/sysbio/syz005

**Uyeda, J. C., Zenil-Ferguson, R. and Pennell, M. W.** (2018). Rethinking phylogenetic comparative methods. *Syst. Biol.* **67**, 1091-1109. doi:10.1093/sysbio/syy031

**Wagner, G. P., Chiu, C.-H. and Laubichler, M.** (2000). Developmental evolution as a mechanistic science: the inference from developmental mechanisms to evolutionary processes. *Am. Zool.* **40**, 819-831. doi:10.1093/icb/40.5.819