E-Article

Ontogeny Tends to Recapitulate Phylogeny in Digital Organisms

Jeff Clune,[1,*] Robert T. Pennock,[2] Charles Ofria,[1] and Richard E. Lenski[3]

1. Department of Computer Science and Engineering, Michigan State University, East Lansing, Michigan 48824; 2. Department of Philosophy, Department of Computer Science and Engineering, and Lyman Briggs College, Michigan State University, East Lansing, Michigan 48824; 3. Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824

Submitted January 11, 2012; Accepted April 24, 2012; Electronically published July 16, 2012

Dryad data: http://dx.doi.org/10.5061/dryad.9k6k0.

abstract: Biologists have long debated whether ontogeny recapitulates phylogeny and, if so, why. Two plausible explanations are that (i) changes to early developmental stages are selected against because they tend to disrupt later development and (ii) simpler structures often precede more complex ones in both ontogeny and phylogeny if the former serve as building blocks for the latter. It is difficult to test these hypotheses experimentally in natural systems, so we used a computational system that exhibits evolutionary dynamics. We observed that ontogeny does indeed recapitulate phylogeny; traits that arose earlier in a lineage's history also tended to be expressed earlier in the development of individuals. The relative complexity of traits contributed substantially to this correlation, but a significant tendency toward recapitulation remained even after accounting for trait complexity. This additional effect provides evidence that selection against developmental disruption also contributed to the conservation of early stages in development.

Keywords: ontogeny, phylogeny, evolution of development, experimental evolution, complex traits, recapitulation.

Introduction

> [T]he adult differs from its embryo, owing to variations supervening at a not early age, and being inherited at a corresponding age. This process, whilst it leaves the embryo almost unaltered, continually adds, in the course of successive generations, more and more difference to the adult. Thus the embryo comes to be left as a sort of picture, preserved by nature, of the ancient and less modified condition of each animal. This view may be true, and yet it may never be capable of full proof. Darwin (1859, p. 358)

Much ink has been spilled discussing the biogenetic law proposed by Ernst Haeckel (1866) and its claim that "ontogeny is the short and rapid recapitulation of phylogeny" (translation from p. 6 of McKinney and McNamara 1991; Gould 1977). In fact, the first recapitulation theory was proposed by J. F. Meckel (1811), but Haeckel brought discussions of the idea to the forefront of biology (Sander and Schmidt-Ott 2004). Haeckel's version of this theory, which proposes that organisms rapidly progress through the adult stages of their ancestors before reaching their own adult form, has been rejected for some time. Many biologists instead favor Karl Ernst von Baer's proposition (1828) that organisms start their lives morphologically similar and become increasingly different with age (Gould 1977). Even Charles Darwin (1859) entertained these ideas, as in the quotation above, although he never developed a formal theory of recapitulation (Mayr 1994).

To explain the similarities between ontogenies (developmental changes within an individual's lifetime) and phylogenies (evolutionary changes across generations), early proponents of recapitulation posited that two forces were at work. The first was "terminal addition," in which newly evolved traits appear toward the end of the organism's life, and the second was "condensation" (or "anticipation"), in which later traits evolve toward an earlier onset (Gould 1977). These explanations are incomplete, however, because they do not explain why evolution would operate in this way. Supposed causal explanations existed under a Lamarckian paradigm, in which it was thought that those tasks that were repeated the most (such as those done routinely by adults) would be more strongly absorbed into the organism's inherited constitution and that their age of onset would thereby be shifted earlier in development (Gould 1977). The fall of Lamarckism invalidated this causal explanation. Thus, recapitulation, which was once described as a law of biology and at the forefront of evolutionary

* Corresponding author. Present address: Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14853; jeffclune@cornell.edu.

Am. Nat. 2012. Vol. 180, pp. E54–E63. 0003-0147/2012/18003-53577$15.00.
DOI: 10.1086/666984

inquiry, fell out of favor and faded into the background (Gould 1977).

However, as Mayr (1994, p. 227) emphasized, "in spite of the disrepute into which Haeckel's claims had fallen ... every embryologist knew that there was a valid aspect to the claim of recapitulation." The sentiment is still widely held today, and the idea that ontogeny recapitulates phylogeny in some form has its modern proponents (Lovejoy 2000). There are detractors as well (Richardson et al. 1997), and genetic findings have fueled the debate even as they have provided new data (Arthur 2002). Nevertheless, the current consensus seems to be that recapitulation is a general trend of evolution, at least in the form of terminal addition being the most common way that new traits are added to ontogenies (Ekstig 1994). At a minimum, the fact that the debate has continued for so long lends credence to Mayr's view that there is at least some validity to recapitulation (Mayr 1994). What is missing is an accepted causal explanation or set of causal explanations for recapitulation.

Gould attempted to provide an explanation based on heterochrony, which refers to the delay or acceleration of some developmental events relative to others (Gould 1977). However, as Mayr (1994) pointed out, this model might provide a proximate explanation for recapitulation, but it does not explain why those heterochronic changes that cause recapitulation would be selectively favored. Indeed, Gould recognized that heterochrony could have the opposite effect, as in the case of paedomorphosis, whereby adults express traits associated with juvenile states of their ancestors (Gould 1984). Mayr also noted that the maintenance of unnecessary ontogenetic stages cannot be explained by heterochrony (Mayr 1994). In fact, a major reason that early biologists were interested in recapitulation was because of those features retained in organisms that seem to serve no purpose, such as pharyngeal pouches (vestigial gill structures) in human embryos (Gould 1977). In an effort to explain why biologists did not reject Haeckel's seemingly ridiculous idea that organisms pass through the adult forms of their ancestors in favor of von Baer's more plausible view that organisms become increasingly different and more elaborate with age, Mayr (1994, p. 227) wrote:

> [von Baer's theory] was in frequent conflict with the facts and did not explain at all the more remarkable cases of recapitulation. The gill arch stage simply was not a simpler, more homogenous, and general version of the subsequent differentiation. Why should the notochord be considered a simpler stage of development when it is later entirely reabsorbed? In fact, just about every example of recapitulation fails to agree with [von Baer's] claim that it was a simpler and more general version of the subsequent differentiation.

To put these observations into a modern Darwinian framework, Mayr (1994) hypothesized that ontogenetic changes to an organism would typically disrupt its later development and that earlier changes would tend to be more disruptive, thereby providing a selective advantage to those organisms that retain ancestral developmental features, especially those expressed early in life. Similar arguments were made by others, including Gould (1984), and were made as far back as the late 1800s (Kleinenberg 1886; Levinton 1988). An example of such developmental disruption is the process of induction, in which part of the embryo produces a chemical signal that initiates developmental processes in other parts of the embryo: mutations that disrupt such induction can mean that important downstream developmental processes never begin (Horder 2008). As McKinney and McNamara (1991, p. 335) colorfully put it, earlier stages are conserved because "evolution is not only restricted to 'tinkering' with preexisting ontogenies, but must do so 'while the engine is running' (keeping the ontogeny viable at all times). Most additions will occur at the end of ontogeny (minimizing interference with later development)." There is no generally accepted name for this hypothesized selective mechanism that prevents the disruption of later developmental processes. In this article, we will call it the "developmental disruption force." Evidence from modern developmental biology that gene regulatory networks have been highly conserved through evolutionary time (Davidson and Erwin 2006) is consistent with the hypothesis that changing early developmental stages disrupts later development. Because the regulatory networks that govern development are complex and interconnected, they are highly sensitive to perturbations (Davidson 2001). The developmental disruption force can cause newly evolved features to be expressed later in life, meaning that the sequence of developmental stages (ontogeny) will tend to recapitulate the order in which traits evolved (phylogeny). The developmental disruption force thus offers a plausible selection-based explanation for the often-dismissed claim that ontogenic order may provide some information about phylogenetic order.

Fritz Müller (1864) provided a different explanation for the conservation of early developmental stages, one that also emphasizes the role of selection (Gould 1977). He argued that if a function were beneficial from the time it first develops, then the earlier it appears the longer the benefit would accrue. By this logic, selection will preferentially favor beneficial traits with earlier onset over those with later onset; this general explanation includes the specific life-history pressures to mature and reproduce earlier

(Williams 1957; Emlen 1984; Stearns 1992). Moreover, those beneficial traits that arose early in evolutionary history would have had more time for selection to move their development forward to earlier stages, thereby generating a tendency for ontogeny to recapitulate phylogeny. We will call this selective mechanism "Müller's force."

The differences between Müller's force and the developmental disruption force are subtle but important. With Müller's force, the advantage that comes from conserving an early developmental stage is that a beneficial function is expressed for a longer period. With the developmental disruption force, the benefit does not derive from the conserved developmental stage itself, but rather comes from leaving that stage unmodified so as not to disrupt later beneficial processes that depend upon it. In other words, the developmental disruption force selects against mutations that have "domino effects" that disrupt later developmental processes. Note also that Müller's force would not preserve early traits that provide no direct benefit, such as those that are strictly vestigial. By contrast, the developmental disruption force could preserve vestigial traits that provide no direct benefit, if modifying or eliminating them would adversely affect beneficial traits in later stages of development.

We now propose a third factor that could produce a tendency for ontogeny to recapitulate phylogeny. This third factor, which we call the "complexity correlation effect," is not a selective force per se, but rather a correlation that might result if simpler structures and functions are commonly used as building blocks for more complex structures and functions. In that case, simpler structures and functions may precede more complex ones in both ontogeny and phylogeny owing to the fact that complexity tends to increase over both time scales. This issue is related to Gould's assertion that complexity appears to increase over evolutionary time, not because increased complexity is always favored, but rather because early states invariably had low complexity, whereas modern ones have both low and high complexity (Gould 1996). However, to our knowledge, Gould did not make any connection between this point and the tendency of ontogeny to recapitulate phylogeny.

In this article, we investigate whether these mechanisms generate some tendency for ontogenies to recapitulate phylogenies, all other things being equal. Other forces can cause exceptions, of course, which makes it unlikely for any such recapitulation to be perfect. It is known, for example, that intermediate stages of development in many phyla, called phylotypic stages, are more conserved than earlier stages (Raff 1996; Galis and Metz 2001; Galis et al. 2002). In the "Discussion," we interpret such phylotypic stages in light of the results of this article.

The contributions of the previously described mechanisms to the relationship between ontogeny and phylogeny are difficult to quantify and tease apart in biological systems. However, they can be studied with computational systems that exhibit evolutionary dynamics. One such study found evidence of Müller's force being important in evolving neural networks, but the system explicitly excluded the possibility of the developmental disruption force (Nolfi and Parisi 1995). Other computational studies have documented instances of heterochronic change, but they did not address whether these changes tended to conserve early developmental stages (Cangelosi 1999; Matos et al. 2005). We are unaware of any experiments that have specifically examined whether ontogenies contain information about phylogenies because of either the developmental disruption force or the complexity correlation effect.

In this article, we use an evolving computational system to study whether ontogeny tends to recapitulate phylogeny and, if so, why that tendency exists. We employ a system in which digital organisms self-replicate, mutate, compete for resources, and evolve. This system allows us to control our experiment in such a way that we can eliminate Müller's force by making the performance of a given function be rewarded only once, irrespective of the time at which it is performed within an organism's life. We can also separately quantify the complexity correlation effect and the developmental disruption force by comparing the ordinal relationship between events during the development of individuals and during evolution along the lineages leading to those individuals. Moreover, we can independently evolve many replicate ontogenies and thereby statistically quantify these effects. Such analyses would be extremely difficult, if not impossible, in a biological system. Because this approach is unfamiliar to most biologists, we provide an extensive overview of the experimental system and of earlier findings that bear on our study in the following section.

Method and Previous Results

Overview of the Experimental Design

Lenski et al. (2003) used the Avida system to investigate the evolutionary origins of complex features. In one experiment, 50 populations of digital organisms evolved in environments that rewarded individuals that performed any of nine different one- and two-input logic functions (table 1). The organisms evolved to perform many of these functions, ranging from five to all nine functions, across the 50 runs (i.e., replicates). The authors also traced the precise line of descent in each run from the common ancestor to the most abundant genotype in the final population. In our current study, we use data from these same

Table 1: Logic functions that were rewarded when performed by the digital organisms

Logic function name	Logic operation
NOT	~A, ~B
NAND	~(A and B)
AND	A and B
OR_NOT	(A or ~B), (~A or B)
OR	A or B
AND_NOT	(A and ~B), (~A and B)
NOR	~A and ~B
XOR	(A and ~B) or (~A and B)
EQU	(A and B) or (~A and ~B)

Note: This table lists the logic functions that, when performed by the digital organisms, were rewarded in the experiments reported by Lenski et al. (2003). The data from those experiments are further analyzed in our article. The letters A and B represent 32-bit binary strings. The tilde character indicates bitwise negation (e.g., the negation of the binary string "0011...0" is "1100...1"). The result of each logic operation is another 32-bit binary string, where the appropriate logic operation has been performed on the *i*th bit in either A or B for NOT or on the *i*th bits in A and B for all other functions.

50 runs to ask whether ontogeny tends to recapitulate phylogeny in this system. Specifically, we compare the historical order in which the functions evolved in each of the 50 lineages with the order in which those tasks are expressed during the development of the final organism in each lineage, and we ask whether these orders are correlated.

Overview of the Avida System

The Avida platform provides a virtual world in which populations of digital organisms evolve by random mutation, genetic drift, and natural selection acting on phenotypic differences among individuals (Lenski et al. 2003; Ofria and Wilke 2004). At the start of a typical Avida run, this world is seeded with digital organisms that can self-replicate but cannot perform any other functions. Each organism has a genome that consists of a sequence of instructions, and the execution of these instructions corresponds to the development of an individual organism. During typical development, various operations are performed that change the state of an organism's data containers; these stacks, registers, and memory spaces are, in essence, analogous to parts of an organism's body. The genome is also copied during development, and the resulting offspring is placed in the population. That copy process is imperfect, however, and leads to random mutations that produce the phenotypic variation upon which natural selection acts. Most mutations are deleterious or neutral, as in biological systems, but some mutations are beneficial and thereby allow adaptation to occur (Lenski et al. 1999; Ofria et al. 2002). In the course of development, an organism may perform one or more logic functions and, if so, the organism obtains additional energy. Digital organisms require energy to execute their genomic programs and thereby replicate, and those organisms that obtain more energy, use that energy more efficiently, or both have higher expected fitness (Lenski et al. 2003; Ofria and Wilke 2004). An individual may fail to reproduce if a deleterious mutation renders it unable to replicate or if it is eliminated at random; population size is typically held constant, so that each reproductive event leads to the loss of a randomly chosen individual.

As Dennett (2002, p. E83) has emphasized, "evolution will occur whenever and wherever three conditions are met: replication, variation (mutation), and differential fitness (competition)." The Avida system fulfills all of these conditions, and it is thus a tractable model for investigating the general properties of evolving systems (Lenski et al. 1999, 2003; Adami et al. 2000; Wilke et al. 2001; Chow et al. 2004; Goings et al. 2004; Misevic et al. 2006; Pennock 2007; Clune et al. 2008, 2010).

Specific Methods in Previous Research on the Evolution of Complex Traits

Here we summarize the specific methods used in the original experiments, which are described fully by Lenski et al. (2003). Fifty experimental runs each began with 3,600 identical ancestors that could replicate but could not perform any logic functions. When an offspring was produced, it was placed locally in the population and replaced one of the existing Avidians, such that the population size was constant over time. All 50 populations evolved in identical environments, but each of the 50 runs was seeded with a different random number, which led to different sequences of stochastic events, including mutations. Each run terminated after the population as a whole had executed a fixed number of genomic instructions. That number would have allowed more than 15,000 generations of the ancestor, but because development time (i.e., the number of instructions executed by an organism) was free to evolve, the resulting number of generations varied. Organisms obtained the energy required to execute their genomic instructions at a rate proportional to the product of their genome length and computational merit, scaled by the total energy demand of the population. Computational merit is the product of the rewards for performing various Boolean logic operations, as described below.

Mutations occurred stochastically as the Avidians copied their genomes in the course of reproduction. There were 26 different instructions in the genetic code, and one of them was present at each position in the genome. The ancestor had a genome length of 50 instructions, whereas

evolved organisms had genomes with variable lengths. Mutations included substitutions (at a constant rate of 0.0025 per instruction copied) as well as single-instruction deletions and insertions (both at constant rates of 0.05 per genome replication). The total mutation rate was thus 0.225 mutations per genome in the ancestor. Mutations that altered an organism's genome may or may not have changed that organism's fitness, because Avida has no explicit function that specifies replication rate as a function of genome sequence. Instead, organisms execute the programs encoded by their genomes and generate phenotypes that may or may not obtain extra energy and produce offspring.

Avidians compete for the energy needed to execute instructions. The execution of these instructions enables an organism to replicate itself and perform logic functions that may yield more energy. To perform these functions, organisms must input one or two 32-bit strings (numbers) and then manipulate them internally such that another string is output that precisely matches the number expected if they have performed one of nine distinct logic operations (table 1). For example, successful performance of the NOT function requires flipping each bit in a single numerical string; the AND function requires inputting two strings and outputting a string with a 1 at each bit position where both input strings have 1 and with a 0 at all other positions; and EQU (equals) requires outputting a string that has a 1 where each bit position has the same value in the two input number strings and a 0 if the two values at that position differ. The successful performance of each function requires the coordinated execution of multiple instructions, with some functions being more complex than others.

Given the Avida instruction set, NOT is one of the simpler functions, whereas the EQU function is the most complex to perform and probably the most difficult to evolve. The energy reward for an Avidian that performed a function increased exponentially from twofold for NOT to 32-fold for EQU, on the basis of the complexity of the function, as judged by the minimum number of nand (not and) operations required to perform that function in a handwritten program.

Organisms received rewards for as many different functions as they performed, but they did not receive any additional reward for performing the same function multiple times. Importantly, the order in which an individual performed the functions during its life did not matter, because the rewards were tallied only after successful replication, at which time both the parent and its offspring received the energy earned by the parent, and the state of the parent organism was reset, thereby producing two "daughter cells." The effect of Müller's force was thus eliminated, because an organism did not receive more energy if it performed a function earlier rather than later in its life history.

Key Results from Previous Research on the Evolution of Complex Traits

As noted above, the successful performance of logic functions required a digital organism to manipulate numbers via a series of steps to produce the rewarded outputs, with more steps required for complex functions than for simpler functions. As a consequence, Avidians typically evolved the ability to perform some of the simpler functions before they evolved the ability to perform the most complex functions. Moreover, simpler functions typically served as building blocks for more complex ones. Thus, EQU evolved in 23 of 50 runs in an experiment in which all nine functions were rewarded; however, EQU never evolved in the 50 otherwise identical runs in which EQU was rewarded but no other logic functions that could serve as building blocks were rewarded (Lenski et al. 2003).

In addition to serving as evolutionary building blocks, the instructions used to perform these simpler functions may also serve as intermediate stages during the development of individual organisms. That is, reaching an internal state that would allow an individual to perform the EQU function, for example, requires a complex series of earlier states, some of which could be useful for the performance of other rewarded tasks. Thus, a mutation that disrupts an instruction necessary to achieve one of these early states may alter subsequent development in such a way that the organism loses later rewards that it would otherwise have earned. In other words, a single mutation can affect multiple traits, including ones expressed both early and late in development.

Quantifying the Similarity Between Ontogeny and Phylogeny

A total of 373 logic functions evolved in the 50 runs, for an average of 7.46 derived functions per run. We recorded a development vector, defined as the ontogenic order in which functions were expressed during genome execution, for the most abundant organism (genotype) at the end of each run. We also recorded an evolution vector, defined as the order in which functions evolved along the line of descent leading to each of these organisms. We quantified the similarity of these two vectors for every run, as described below.

To compute a similarity score, we first counted the number of pairs of the rewarded logic functions that occurred in either the same or opposite orders in the evolution and development vectors. If they were in the same order (e.g., if NOT preceded EQU in both vectors), then the count

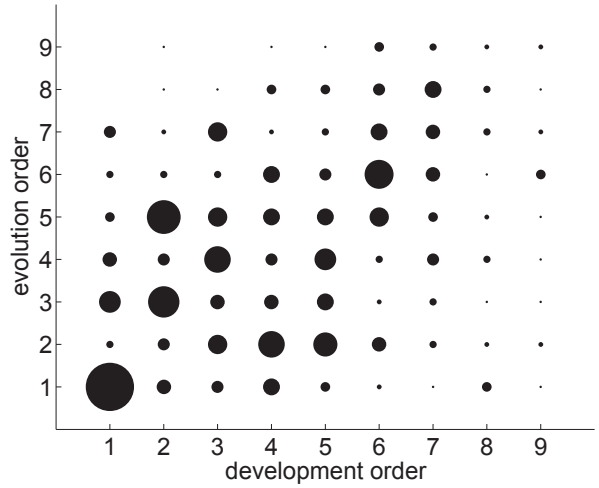

Figure 1: The order that logic functions evolved along the lineage leading to 1 of 50 independently derived digital organisms (evolution order) is plotted against the order in which the functions are expressed developmentally within the lifetime of the organism (development order). For example, if the first function performed during an organism's lifetime (developmental order = 1) is also the first to have evolved on the lineage leading to that organism (evolutionary order = 1), then it would be included in the circle at the lower left. All 373 functions that evolved in the 50 evolutionary runs are shown in this figure. The size of each circle represents the number of times that each combination occurred, with an empty cell indicating 0 and the largest circle indicating the most occurrences, which was 20.

was increased by one. If they were in the opposite order (e.g., if NOT preceded EQU in one vector but followed it in the other), then the count was decreased by one. We then calculated an ontogeny recapitulates phylogeny (ORP) score by dividing the total count by the maximum count possible for the organism at hand. Thus, the maximum possible ORP score is 1, which would result if an organism performed all of its functions in the same order in which they had evolved in its ancestors.

This proportional scoring system allows for vectors of different lengths; the number of functions performed by the final organisms ranged from five to nine. In some cases, functions evolved that were later lost, and sometimes these functions reevolved. In calculating the ORP score, a function entered the evolution vector the first time that it evolved, even if it was later lost. If the lost function did not reevolve and thus did not occur in the development vector, then it was ignored and had no impact on the ORP score. Ties occurred very rarely when two functions evolved in the same generation; these cases also had no effect on the ORP score. In those cases in which the final organism performed some function multiple times in its development, the first performance of that function determined its order in the development vector.

If there were no correlation between ontogeny and phylogeny, then the average ORP score would be 0. A significant tendency toward positive scores would mean that the development and evolution vectors were correlated, such that ontogeny indeed tends to recapitulate phylogeny. Negative scores are theoretically possible and indeed sometimes occurred; an excess of negative scores would imply anticorrelation between the two vectors.

New Results

Ontogeny Recapitulates Phylogeny

There is a positive correlation between the order that logic functions evolved in 50 independent lineages and the order in which those functions were expressed during the development of the derived organisms (fig. 1). The mean ORP score is 0.270, with a standard error of 0.047 and a range of −0.5 to 1.0 (fig. 2). The mean score is significantly greater than 0 based on a one-tailed *t*-test (t_s = 5.799,

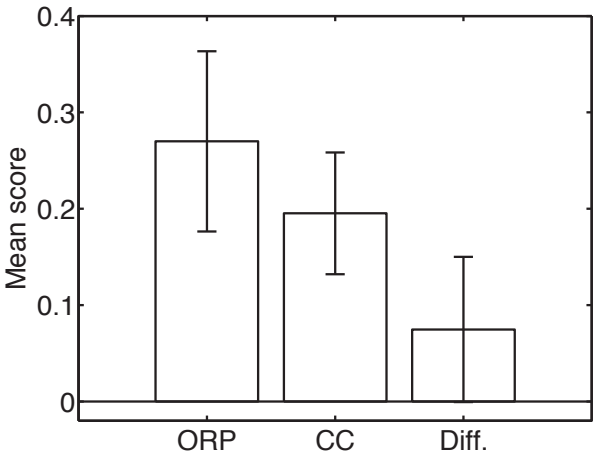

Figure 2: Mean scores and 95% confidence intervals for three metrics describing the relationship between ontogeny and phylogeny. The ORP score measures the overall tendency of ontogeny to recapitulate phylogeny; the expected value is 0 under the null hypothesis of no relationship between the order in which traits evolved and the order in which they are expressed within an individual's lifetime. The maximum score is 1 when the order in which traits are expressed precisely matches the order in which they evolved. The CC score measures the contribution of trait complexity to the ORP score. Simpler traits tended to both evolve sooner and be expressed earlier than more complex traits, and the contribution of this complexity correlation to the ORP score was calculated as described in the text. The Diff. score measures the difference between the ORP and CC scores. This metric reflects the tendency of ontogeny to recapitulate phylogeny beyond the effects of trait complexity on the order of events in development and evolution. Although the confidence interval for the mean Diff. score includes 0, a one-tailed test indicates that the average Diff. score is significantly greater than 0.

$P < .0001$). These data thus support the conclusion that there is a strong tendency for ontogeny to recapitulate phylogeny in this experimental system.

Complexity of Traits Contributes to the Similarity of Ontogeny and Phylogeny

Complex traits are often built upon and may even require the prior existence of simpler traits. Complexity may thus increase during both evolution and development, thereby generating a correlation in the ordering of events across these distinct timescales that reflects that dependency. Thus, any similarity between ontogeny and phylogeny that can be explained by this complexity correlation effect need not be attributed to the developmental disruption force, which is driven by selection against mutations that disrupt both early and late developmental stages.

In the Avida system, simpler functions require the coordinated execution of fewer instructions, and they are thus more likely to arise by random mutation early in a population's history than are more complex functions. The performance of complex functions, such as EQU, typically involves intermediate steps that are themselves solutions to simpler functions, such as NAND. An organism must still output the intermediate calculation to obtain the benefit of performing the simpler function, but doing so while the intermediate state is readily available is usually developmentally simpler (and more efficient) than recomputing and outputting the same solution at a later stage. However, organisms could compute and output the answer to a complex function without outputting any intermediate calculations, meaning that the most complex task could be the first task in the development vector. The execution of instructions can also reset the developmental state of an organism, meaning that all permutations of the developmental vector are possible, although some are less likely to evolve, because it tends to be more efficient to do simpler tasks first and output their answers before computing more complex tasks.

We can quantify this complexity correlation effect, because it should exist between any pair of evolution and development vectors generated using the same ancestral organism and selective environment, even when the evolution and development vectors come from different runs. By contrast, if the developmental disruption force were important, then the evolution vector for a given run should contain information about the development vector for that same run above and beyond the complexity correlation effect.

To estimate the contribution of the complexity correlation to the ORP score, we computed a second score, called CC. We began by comparing the development vector for each run to the evolution vectors from the other runs in the same manner as used to calculate the ORP score; only the functions present in both vectors were used in the calculations. We then took the mean of these values for each run and the grand mean CC score across the 50 runs. A statistical complication arises because the evolution vector for each run is used in calculating the values for all of the other runs. To solve this problem, we employed Tukey's jackknife method (Sokal and Rohlf 1995) by excluding each run one at a time and computing 50 so-called pseudovalues that reflect each independent run's contribution to the overall CC score.

The grand mean CC score for the 50 runs is 0.195 (fig. 2), with a standard error of 0.031 and a range from −0.318 to 0.587 based on the pseudovalues. The mean is significantly greater than 0 according to a one-tailed *t*-test ($t_s = 6.215$, $P < .0001$). The complexity correlation effect thus contributes to the observed tendency of ontogeny to recapitulate phylogeny in this system. Based on the ratio of the mean CC and ORP scores, the complexity correlation effect explains approximately 72% of the overall relationship.

Developmental Disruption Is Also Important to the Similarity of Ontogeny and Phylogeny

The finding that the complexity correlation effect accounts for the majority of the relationship between ontogeny and phylogeny does not mean that other factors are unimportant. We can estimate the effect of other factors from the difference between the ORP and CC scores; that quantity reflects the extent to which scores are higher when development and evolution vectors are from the same run rather than different runs. The mean difference was 0.075 (fig. 2), with a standard error of 0.038 and a range of −0.474 to 0.877. The mean difference is significantly greater than zero using a one-tailed paired *t*-test ($t_s = 1.992$, $P = .026$).

Thus, some additional factor causes the development of a particular digital organism to recapitulate its own evolutionary history, above and beyond the effect explained by the tendency of simpler functions to evolve and be expressed earlier in development than more complex functions. As discussed earlier, Müller's force could not have occurred in the experiments analyzed here, because the performance of any single function by an organism was rewarded only once and because the time within an organism's life when functions were performed did not affect fitness. Therefore, the significant difference between the ORP and CC scores indicates that the developmental disruption force was operative and responsible for approximately 28% of the overall similarity between ontogeny and phylogeny in this system.

Discussion

Evolution experiments using digital organisms (i.e., computer programs that replicate, mutate, compete, and evolve) offer the opportunity to test evolutionary hypotheses with control, replication, and precision that Darwin could not have envisioned. We have taken advantage of this approach to examine an issue that was contentious in his day and that remains unsettled, which is whether and why ontogeny seems to recapitulate phylogeny. This system allowed us to determine precisely when along an evolutionary lineage a set of discrete traits first evolved and when during an individual's life each of these traits was first expressed. Moreover, we could obtain these data for 50 replicate lineages that had independently evolved in the same environment, starting from the same ancestor that lacked the derived traits of interest. We could then quantify the extent to which ontogeny and phylogeny covaried to an extent greater than would be expected by chance. Indeed, there was a strong tendency for ontogeny to recapitulate phylogeny.

In this system, two distinct factors contributed to the similar ordering of events on the developmental and evolutionary timescales. One factor reflects differences between traits with respect to their complexity. Some traits are simpler to perform than others, and moreover simple traits are often used as building blocks for more complex ones. As a consequence, simple traits tend to precede more complex traits during both evolution and development, thereby generating correspondence in the ordering of events, which we call the complexity correlation effect. To our knowledge, this explanation has not previously been proposed to explain the tendency of ontogeny to recapitulate phylogeny. It seems plausible to us that many systems with both developmental and evolutionary timescales would show this correlation. However, there may not be explicit criteria to order the complexity of most traits in biological systems. Also, the absence of statistical replication, in which the same pairs of traits repeatedly evolve, would make it difficult to test this hypothesis in biological systems. By contrast, in the digital realm, we have both a priori estimates of the relative complexity of the various traits and empirical data on the ordering of traits from independently evolved lineages. This latter information allowed us to test the complexity correlation effect by quantifying the similarity in the order in which traits are expressed developmentally in each lineage with the order that the same traits evolved in other lineages. This analysis showed that over half of the similarity in the ontogenetic and phylogenetic order of events reflected the complexity correlation effect in this system.

This same analysis also showed, however, that not all the tendency of ontogeny to recapitulate phylogeny was explained by the complexity correlation effect. In fact, the similarity between the order in which traits are expressed in a given lineage and the order in which they evolved in the same lineage was significantly greater than the corresponding similarity across independently evolved lineages. Müller's force, which refers to selection for the cumulative benefit of a continuously expressed trait, cannot explain this difference, because the performance of each logic function was rewarded only once during a digital organism's life. Instead, we attribute this difference to the developmental disruption force. That is, selection tends to preserve early developmental stages more than later stages, because early acting mutations often disrupt traits expressed both early and late in life and therefore are more costly, on average.

These results reveal a selective pressure that causes a tendency for earlier developmental stages to be conserved, all else being equal. In nature, of course, all else is not equal, and exceptions occur. In the case of recapitulation, notable examples of such exceptions are phylotypic stages, which are more conserved than both earlier and later developmental stages. Raff (1996) hypothesized that phylotypic stages are conserved because they have the most pleiotropy (i.e., the developmental disruption force is strongest during this stage). In earlier stages, there is less pleiotropy, because the complexity of the embryo is low; the gradients providing positional information are global and simple, which allows substantial developmental flexibility. In later stages, development is compartmentalized in modules, which limits pleotropic effects. In the intermediate phylotypic stage, complexity is high, yet compartmentalization is incomplete and intermodule interactions are pervasive, which leads to high pleiotropy (Raff 1996). Experiments have supported this theory by showing that induced changes to phylotypic stages have larger, more deleterious, and more widespread effects than changes to both earlier and later stages (Galis and Metz 2001; Galis et al. 2002). The greater conservation of the phylotypic stages is in conceptual agreement with our finding that the developmental disruption force causes selection to maintain early developmental stages. The simplicity of the earliest developmental stages, however, sometimes leads to exceptions to the rule that the developmental disruption force is stronger in earlier stages of development (Horder 2008, p. 170). Indeed, the reason so much attention has been paid to the fact that phylotypic stages are more conserved than earlier stages is because the general expectation is that the developmental disruption force is stronger for earlier stages (Raff 1996).

Our research with digital organisms leaves open the question of which potential causes of recapitulation are most important in nature. Whatever the case may be, we have provided evidence that two distinct forces can cause

ontogeny to recapitulate phylogeny, which we believe is a useful step towards clarifying the debate that has raged on this subject from the days of Meckel, von Baer, and Darwin through Mayr and Gould to the present.

Acknowledgments

We thank S. Benson-Amram, A. Hintze, A. McAdam, P. McKinley, D. Misevic, and the members of the digital evolution laboratory at Michigan State University for valuable discussions. We also thank the two reviewers for their insightful comments. This research was supported in part by the BEACON Center for the Study of Evolution in Action (National Science Foundation Cooperative Agreement DBI-093945) and a National Science Foundation Postdoctoral Fellowship (DBI-1003220 to J.C.).

Literature Cited

Adami, C., C. Ofria, and T. C. Collier. 2000. Evolution of biological complexity. Proceedings of the National Academy of Sciences of the USA 97:4463–4468.

Arthur, W. 2002. The emerging conceptual framework of evolutionary developmental biology. Nature 415:757–764.

Cangelosi, A. 1999. Heterochrony and adaptation in developing neural networks. Proceedings of the Genetic and Evolutionary Computation Conference 2:1241–1248.

Chow, S. S., C. O. Wilke, C. Ofria, R. E. Lenski, and C. Adami. 2004. Adaptive radiation from resource competition in digital organisms. Science 305:84–86.

Clune, J., H. Goldsby, C. Ofria, and R. T. Pennock. 2010. Selective pressures for accurate altruism targeting: evidence from digital evolution for difficult-to-test aspects of inclusive fitness theory. Proceedings of the Royal Society B: Biological Sciences 278:666–674.

Clune, J., D. Misevic, C. Ofria, R. E. Lenski, S. F. Elena, and R. Sanjuán. 2008. Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. PLoS Computational Biology 4:e1000187.

Darwin, C. 1859. On the origin of species by means of natural selection. J. Murray, London.

Davidson, E. H. 2001. Genomic regulatory systems: development and evolution. Academic Press, San Diego.

Davidson, E. H., and D. H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. Science 311:796–800.

Dennett, D. C. 2002. The new replicators. Pages E83–E92 *in* M. Pagel, ed. The encyclopedia of evolution. Oxford University Press, Oxford, UK.

Ekstig, B. 1994. Condensation of developmental stages and evolution. BioScience 44:158–164.

Emlen, J. M. 1984. Population biology. Macmillan, New York.

Galis, F., T. J. M. van Dooren, and J. A. J. Metz. 2002. Conservation of the segmented germband stage: robustness or pleiotropy? Trends in Genetics 18:504–509.

Galis, F. and J. A. J. Metz. 2001. Testing the vulnerability of the phylotypic stage: on modularity and evolutionary conservation. Journal of Experimental Zoology 291:195–204.

Goings, S., J. Clune, C. Ofria, and R. T. Pennock. 2004. Kin selection: the rise and fall of kin-cheaters. Pp. 303–308 *in* J. Pollack, M. Bedau, P. Husbands, T. Ikegami, and R. Watson, eds. Artificial Life: Proceedings of the 9th International Conference on Artificial Life. MIT Press, Cambridge, MA.

Gould, S. J. 1977. Ontogeny and phylogeny. Belknap, Cambridge, MA.

———. 1984. Relationship of individual and group change. Human Development 27:233–239.

———. 1996. Full house. Three Rivers, New York.

Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proceedings of the Royal Society B: Biological Sciences 205:581–598.

Haeckel, E. 1866. Generelle Morphologie der Organismen. G. Reimer, Berlin.

Horder, T. 2008. A history of evo-devo in Britain. Theoretical ideals confront with biological complexity. Annals of the History and Philosophy of Biology 13:101–174.

Lenski, R. E., C. Ofria, T. C. Collier, and C. Adami. 1999. Genome complexity, robustness and genetic interactions in digital organisms. Nature 400:661–664.

Lenski, R. E., C. Ofria, R. T. Pennock, and C. Adami. 2003. The evolutionary origin of complex features. Nature 423:139–144.

Levinton, J. S. 1988. Genetics, paleontology, and macroevolution. Cambridge University Press, New York.

Lovejoy, N. R. 2000. Reinterpreting recapitulation: systematics of needlefishes and their allies (Teleostei: Beloniformes). Evolution 54:1349–1362.

Matos, A., R. Suzuki, and T. Arita. 2005. Evolution of development and heterochrony in artificial neural networks. Proceedings of the 10th International Symposium on Artificial Life and Robotics: 635–638.

Mayr, E. 1994. Recapitulation reinterpreted: the somatic program. Quarterly Review of Biology 69:223–232.

McKinney, M. L., and K. McNamara. 1991. Heterochrony: the evolution of ontogeny. Plenum, New York.

Meckel, J. 1811. Entwurf einer Darstellung der zwischen den Embryonalzustanden der hoheren Tiere und den permanenten der niederen stattfindenden Parallele. Beiträge zur vergleichenden Anatomie 2:1–60.

Misevic, D., C. Ofria, and R. E. Lenski. 2006. Sexual reproduction reshapes the genetic architecture of digital organisms. Proceedings of the Royal Society B: Biological Sciences 273:457–464.

Müller, F. 1869. Facts and arguments for Darwin. Translated by W. S. Dallas. J. Murray, London.

Nolfi, S., and D. Parisi. 1995. Evolving artificial neural networks that develop in time. Pp. 353–367 *in* F. Morán, A. Murenu, J. J. Merelu, and P. Chacin, eds. Advances in Artificial Life. Springer, Berlin.

Ofria, C., C. Adami, and T. C. Collier. 2002. Design of evolvable computer languages. IEEE Transactions on Evolutionary Computation 6:420–424.

Ofria, C., and C. O. Wilke. 2004. Avida: a software platform for research in computational evolutionary biology. Artificial Life 10: 191–229.

Pennock, R. T. 2007. Models, simulations, instantiations and evidence: the case of digital evolution. Journal of Experimental and Theoretical Artificial Intelligence 19:29–42.

Raff, R. A. 1996. The shape of life: genes, development, and the evolution of animal form. University of Chicago Press, Chicago.

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.

Sander, K., and U. Schmidt-Ott. 2004. Evo-devo aspects of classical and molecular data in a historical perspective. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 302:69–91.

Sokal, R., and F. J. Rohlf. 1995. Biometry. Freeman, San Francisco.

Stearns, S. C. 1992. The evolution of life histories. Oxford University Press, Oxford.

Wilke, C. O., J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412:331–333.

Williams, G. C. 1957. Pleiotropy, natural selection, and the evolution of senescence. Evolution 11:398–411.

von Baer, K. 1828. Entwicklungsgeschichte der Thiere: Beobachtungen und Reflexionen. Bornträger, Königsberg.

Associate Editor: Stephen B. Heard
Editor: Mark A. McPeek