Examining Community Stability in the Face of Mass Extinction in Communities of Digital Organisms

Tian-tong Luo
Nanjing Agricultural University
  Department of Zoology
  College of Life Sciences
  530317288@qq.com

Lise Heier
  liseheier@lh.net

Zaki Ahmad Khan
Nanjing Agricultural University
  Department of Zoology
  College of Life Sciences
  khanzaki@gmail.com

Faraz Hasan
Aligarh Muslim University
  Department of Computer Science
    and Engineering
  faraz.hasan@live.in

Trond Reitan
University of Oslo
  Centre for Ecological and
    Evolutionary Synthesis
  Department of Biology
  trond.reitan@ibv.uio.no

Abdool S. Yasseen III
University of Toronto
  Dalla Lana School of Public Health
  abdool.s.yasseen@gmail.com

Zi-xuan Xie
Nanjing Agricultural University
  Department of Zoology
  College of Life Sciences
  601818094@qq.com

Jian-long Zhu
Nanjing Agricultural University
  Department of Zoology
  College of Life Sciences
  lijayiben38@126.com

Gabriel Yedid*
Nanjing Agricultural University
  Department of Zoology
  College of Life Sciences
  gyedid02@gmail.com

Abstract   Digital evolution is a computer-based instantiation of Darwinian evolution in which short self-replicating computer programs compete, mutate, and evolve. It is an excellent platform for addressing topics in long-term evolution and paleobiology, such as mass extinction and recovery, with experimental evolutionary approaches. We evolved model communities with ecological interdependence among community members, which were subjected to two principal types of mass extinction: a pulse extinction that killed randomly, and a selective press extinction involving an alteration of the abiotic environment to which the communities had to adapt. These treatments were applied at two different strengths, along with unperturbed control experiments. We examined how stability in the digital communities was affected from the perspectives of division of labor, relative shift in rank abundance, and genealogical connectedness of the community's component ecotypes. Mass extinction that was due to a *Strong Press* treatment was most effective in producing reshaped communities that differed from the pre-treatment ones in all of the measured perspectives; weaker versions of the treatments did not generally produce significant departures from a *Control* treatment; and results for the *Strong Pulse* treatment generally fell between those extremes. The Strong Pulse treatment differed from others in that it produced a slight but detectable shift towards more generalized communities. Compared to Press treatments, Pulse treatments also showed a greater contribution from re-evolved ecological doppelgangers rather than new ecotypes. However, relatively few Control communities showed stability in any of these metrics over the whole course of the experiment, and most did not represent stable states (by some measure of stability) that were disrupted by the extinction treatments. Our results have interesting, broad qualitative parallels with findings from the paleontological record, and show the potential of digital evolution studies to illuminate many aspects of mass extinction and recovery by addressing them in a truly experimental manner.

---

* Corresponding author.

1   Introduction: The Biological and Ecological Interest of Mass Extinction and Recovery

Paleontologists have long recognized that throughout the history of life on Earth, there has been a regular turnover of species (background extinction) punctuated by mass extinctions: infrequent but highly destructive crises that eliminate a sizable fraction of preexisting biological diversity, morphological variety, and ecological structure in geologically rapid episodes [10, 40, 74, 99]. These events are certainly remarkable for their destructive aspects; they are perhaps most notable for their marginalization or outright elimination of previously dominant incumbent taxa. Yet, mass extinctions are also a major creative evolutionary force. Removal of long-standing incumbents creates new opportunities for the subsequent evolution and diversification of surviving clades, often ones that were minor components of the pre-extinction biota. Processes of adaptive radiation and convergent evolution can then lead to new dominant taxa and new ecosystems [7, 72]. Mass extinctions can thus channel evolution in new directions unpredictable from previous background situations, permanently altering the ecological and taxonomic characteristics of biological communities [5, 53, 103].

In spite of the dramatic ecological and evolutionary consequences of mass extinctions and recoveries, they have been difficult to study because of problems posed by the uneven nature of the fossil record. Due to geological control of record quality, many previously studied paleobiologic patterns, including those of extinction and recovery, may be partly or wholly artefactual [43, 56]. Disappearances of taxa from the fossil record, due to geological hiatus or lack of suitable rock formations, may be incorrectly attributed to mass extinction until contrary evidence is discovered [7, 52, 101]. Different ways of adjusting data to compensate for incomplete preservation have resulted in differing conclusions about rates and processes involved in recovery from mass extinctions [4, 58, 66]. Beyond these problems, there is a lack of repeatability and control; one can never know how evolution might have proceeded under different conditions. For example, would a particular biological community have remained "stable" (by some measure) without a mass extinction and recovery? This last issue is worth considering in light of the phenomenon of *coordinated stasis*, where whole paleocommunities purportedly show taxonomic and ecological persistence over geologically significant time periods, interrupted by abrupt mass turnover [15–17, 21, 33, 34, 44, 48, 50, 77, 82, 87, 102]. More generally, we still do not know whether the consequences of different kinds of perturbations (instantaneous versus continuous, random versus selective) are different and repeatable. A deeper understanding of these issues is important not only for comprehending past mass extinction and recovery, but also because current global change represents a continuous, gradual anthropogenic perturbation [20, 73, 75].

Given the aforementioned problems with studying mass extinction and recovery from fossil data, it becomes highly desirable to have an experimental platform with which questions about these phenomena (and other outstanding general problems in long-term evolution) could be addressed without the complications of the geological record. Digital evolution offers some compelling advantages for studying community stability, and mass extinction and recovery more generally. First, and most importantly, it permits a truly manipulative, experimental approach to the problem. Its biggest strength is the ability to "replay life's tape," to use the phrase of Stephen Jay Gould. One may use the capabilities of digital evolution to make changes at specific time points in what would otherwise be identical experimental replicates. For example, how would evolution unfold differently in the presence and absence of an extinction/recovery treatment? One may contrast paleobiological cases, which compare post-extinction and pre-extinction states (which are all one knows in the real world) with counterfactual cases, which compare the known post-extinction state against the corresponding state obtained with different treatments (e.g., no extinction). One may also examine how different kinds of model worlds respond to particular treatments, or vary the number and/or composition of survivors. Second, data may be recorded in far greater detail than is possible with the fossil record, and would be free of the preservational artefacts that complicate quantitative analysis of fossil data. Although not all the details and complexity of the geological record can be re-created with digital evolution, it is highly desirable to have a tractable system that allows one to focus on the key data

sources that are most appropriate for investigating the problem at hand, and to design appropriate and tractable follow-ups. Third, because digital evolution is an actual instantiation of Darwinian evolutionary processes, it allows the problem to be addressed with greater realism of population-dynamic, ecological, and evolutionary factors than in other types of mass extinction simulations (e.g., [95]). As we are interested in both short-term and long-term effects of the perturbations we model (primarily the latter), our consideration goes beyond ecological scales and extends to paleontological time scales corresponding to thousands of post-extinction generations.

2   Methods

2.1   Experimental Platform

We use the digital evolution software *Avida* v.2.4.4 [2, 80] as our experimental platform, allowing a complete record of the course of evolution. An *Avida* world consists of a grid of digital organisms (hereafter referred to as *Avidians*), which in turn consist of a genome of instructions written in an assembly-like, Turing-complete computer language, along with a simple virtual computer chip that executes those instructions. The genome encodes the ability to self-replicate and carry out particular computations. Segments of code responsible for particular functions are analogous to genes. Avidians execute the programs encoded by their genomes, including commands that allow them to copy themselves and divide to produce a daughter organism. Memory for a daughter cell is allocated, and instructions are copied one at a time from parent to daughter. Copying is subject to point mutations, insertions, and deletions. These genomic mutations can indirectly affect an Avidian's ability to self-replicate or perform other computational functions, and their effects depend on their interaction with the rest of the genomic background. Most mutations are deleterious or neutral in their phenotypic effects, but a few are beneficial [61]. Thus, there is a genetic basis for adaptation and speciation. In our experiments, the Avidians are completely asexual, but this is no barrier to using them to address problems involving ecologically diversified populations, as asexual organisms can form phenotypically and functionally distinct phylogenetic clusters just as well as sexual organisms [6].

Each Avidian occupies a cell in the population grid, which is of finite size. Upon division, a daughter organism is placed in the population, killing any previous occupant of the destination cell (various placement options are available). This feature emulates removal of a healthy microbial cell from a chemostat, introducing an element of drift. All Avidians in a population receive a basal number of CPU cycles (analogous to energy), which enable their programs to run. However, Avidians can earn extra energy and accelerate their execution if they evolve the ability to perform certain computational functions. An Avidian that can perform any such functions receives an energetic reward by metabolizing a corresponding virtual resource into additional CPU cycles, enabling faster replication and creating variation among individuals in replication rate. Computationally, more complex functions are built by co-opting evolved mechanisms for simpler ones [62, 105]. Absolute time in *Avida* is measured in *updates*, during which an average of 30 instructions are executed per organism in the population. A generation typically requires 5–10 updates, with the precise number depending on the complexity of the Avidians present.

In our experiments, we used a 60×60 toroidal grid, initially seeded with a single copy of an asexual ancestor 50 instructions in length, capable only of self-reproduction. Mutations occurred during the copy procedure at a rate 0.0075 per instruction copied, and insertion/deletion mutations during division at a rate of 0.05 per divide, permitting change in both genotype and genome size.

2.1.1   Environmental Configuration

The environment configuration used here contains two key features that enhance the model world's ecological realism. First, there are multiple depletable resources, in which use of a resource by one Avidian lowers the availability of that resource for other individuals requiring it. A low concentration of a resource reduces the benefit gained by performing the associated computation,

favoring organisms that target underutilized resources. This resource competition model is analogous to that occurring in bacterial chemostats. Competition between Avidians that need exactly the same resources will favor organisms that use them most efficiently, but organisms with minimal overlap in resource requirements may coexist [23, 25]. Resources have no spatial structure; all organisms access the same resource pools regardless of location on the population lattice. Second, only a limited number of resources are supplied exogenously. The remainder occur as metabolic *by-products* when the Avidians consume available resources by successfully completing associated computations. These by-products then become resources to be metabolized by other Avidians, introducing a degree of ecological interdependence into the population dynamics (Figure 1). This second feature allows for evolution of rudimentary trophic structures through facilitative *niche construction* [79]. Although not a true food web because of the lack of true predation, evolution of such cross-feeding relationships has been observed previously in microbial evolution experiments [14, 57, 91]. The set of particular functions (and associated resources) used by any Avidian defines its *ecotype,* a concept analogous to an ecological guild: All Avidians that belong to the same ecotype perform the same functions and use the same resources, but do not necessarily descend from the same most recent common ancestor. We define ecotypes by the presence/absence of particular functions, but in practice apply the following screening procedure: If a particular function's execution makes up 5% or less of an Avidian's total functional output, it is set to zero, giving greater weight to those functions that are the major components of fitness for the Avidian.

In our experiments, we used an additive reward model. The energy gained for performing a rewarded function was added directly to the Avidian's current execution rate, and could be applied to multiple executions of that function over an organism's gestational cycle. Further information on how computational functions are evaluated and confer additional energy is provided in Supplement A
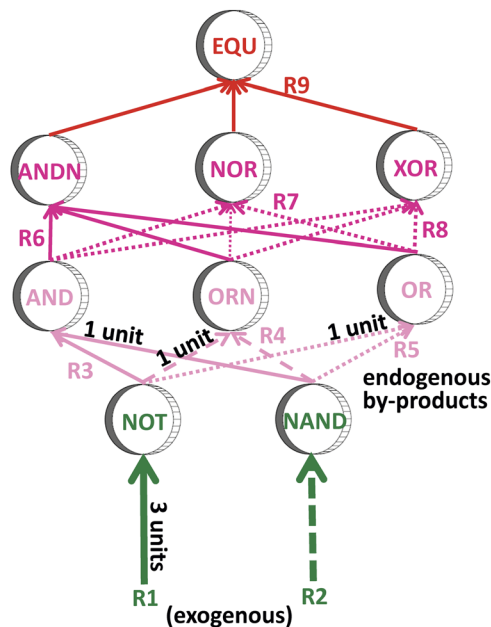

Figure 1. Cross-feeding relationships used in this study. An arrow connecting Avidians signifies that an organism with a lower-level function consumes the incoming resource and makes a by-product that is available for any Avidians with higher-level functions. Arrows are color coded by trophic level (green = 1° production; pink = 1° consumer; magenta = 2° consumer; red = top consumer). Different dashes indicate differently named resources. Only resources R1 and R2 (associated with NOT and NAND) are provided exogenously; all other resources are produced endogenously. Three units of the NOT resource are required to produce one unit each of the resources for AND, ORN, and OR. Similarly, three units of the AND resource are required to produce one unit each of the ANDN, NOR, and XOR resources.

in the online supplementary materials for this article at www.mitpressjournals.org/doi/suppl/10.1162/ARTL_a_00272.

2.2   Mass Extinction in *Avida*

Previous work [104, 106] investigated two principal types of mass extinction: pulse and press extinctions (Figure 2). Geologically, a pulse extinction happens with sufficient speed and power that adaptive change cannot occur during the extinction episode, although adaptation may take place afterwards. Conversely, a press extinction occurs over a longer period that allows for an adaptive response in affected populations [36]. In *Avida*, a pulse extinction is an instantaneous mass culling of individuals from the population, with survivors selected at random from the pool of all viable organisms. By contrast, a press extinction involves a *press episode*, a period of altered environmental conditions—in this case, greatly reduced inflows of basal resources—that persists long enough to allow an adaptive response. The aim is to trigger a massive drop in productivity, leading to bottom-up collapse of the ecological relationships in the community. Such a mechanism has been implicated in several major mass extinctions in the Earth's history [10, 40, 99]. Ecological recovery is then initiated by restoring resource inflows to pre-extinction levels. Unlike other models (e.g., [42, 97]) that used direct, targeted removal of individuals with particular phenotypic characteristics, here the ecosystem is allowed to adjust and evolve entirely on its own in response to the changed abiotic environment.

For the work described herein, the following experimental treatments were performed:

1.   Uninterrupted evolution (*Control*). Each replicate runs unperturbed for 205,000 updates.

2.   *Strong Press* episode. Each replicate runs for 100,000 updates before the press episode; these pre-extinction histories are identical to the Control. Resource inflows are then lowered by two orders of magnitude for 5,000 updates. This treatment is applied uniformly across experiments as absolute time, much as real extinction-driving crises act independently of biological generation times. Resource inflows are then restored for a subsequent 100,000 updates of evolution.

3.   *Weak Press* episode. Pre-extinction histories and press duration are as above, but the reduction in resources is not as severe. The low resource level was about tenfold higher than for Strong Press, so as to most closely match the phylogenetic attrition of Weak Pulse extinction ([104, 107]; see below).

4.   *Strong Pulse* extinction. Pre-extinction histories are as before. At 100,000 updates, an instantaneous mass cull of the population is performed. Survivors are picked randomly
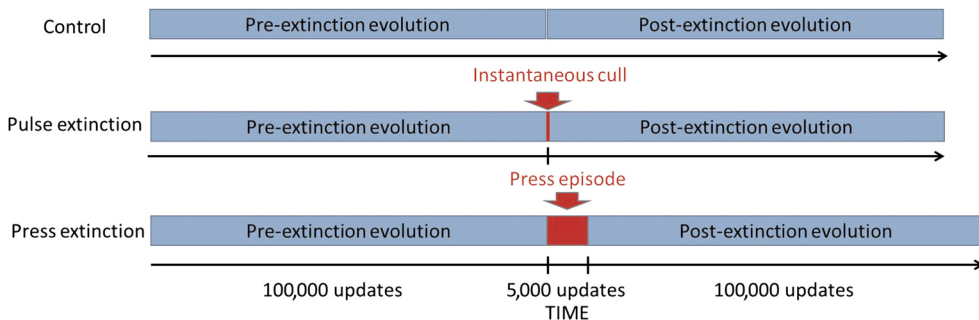

Figure 2. Schematic comparison between setups of Control (top), Pulse (middle), and Press (bottom) treatments. In our experiments, the Control and Pulse treatments are each 200,000 updates in total length; the pulse event is instantaneous. The Press treatment is 205,000 updates in total length, including the press episode of 5,000 updates and pre- and post-treatment intervals of 100,000 updates each.

from among viable organisms, with no environmental alteration. This treatment is followed by 100,000 updates of recovery, again equal to the time for pre-extinction evolution. The survival rate of this strong cull is 4/3600 individuals (0.1%). This was the minimum number needed to ensure population continuity across all replicates; fewer survivors than this often resulted in population collapse due to stochastic effects or retention of highly unfit individuals.

5. *Weak Pulse* extinction. Pre-extinction histories are as before, but the survival rate of the cull is 36/3600 individuals (1%). As Avidians may evolve to perform more than one computational function per organism, this is sufficient to preserve most of the community's ecological integrity.

We performed 100 replicates of each treatment. Each replicate with a particular random seed was identical in its evolutionary history up to 100,000 updates and diverged after that according to the treatment type.

A mass extinction and recovery may lead not only to a change in the total number of ecotypes present in the community, but possibly also to changes in community composition and structure. We address the following questions:

1. To what extent were digital communities stable (by some measure) in the absence and presence of mass extinction?

2. When ecotype identity was considered, did mass extinction and recovery cause greater rearrangement of dominant ecotypes than in corresponding Control replicates?

3. Did mass extinction and recovery often result in the emergence of new stable states, a community with a persistent, differing taxonomic and ecological balance?

4. When eco-phenotypic data suggested stability, was this supported once genealogical considerations were taken into account?

2.2.1   Measuring Stability

We use three different metrics with which to measure stability, each of which captures a different facet of the community. While they are not the ones most commonly used by ecologists, we chose them because of their suitability for the data generated by *Avida* and for ease of interpretability.

(a) *Normalized mutual entropy (NME) as an index of specialization.* Another concern in ecology, biogeography, and sociology is to what extent natural communities are generalized or specialized in some aspect, that is, the degree of division of labor of a set of individuals for a considered set of tasks (e.g., particular behaviors). For systems that may be clearly represented in the aforementioned manner, a promising approach is that of *normalized mutual entropy* (NME [38, 39]), the degree to which knowledge of an individual specifies knowledge of the tasks it performs. This is a somewhat more elaborate application of information-theoretic entropy-based measures to ecological data, a practice that has quite respectable precedent [68]. This metric has previously been applied in investigations of division of labor in social systems [12], but can be adapted for biogeographical studies [37]. It is particularly well suited to analyzing *Avida* data, since Avidian ecotypes in a population may be expressed as two-dimensional matrices, where each individual (rows) performs each rewarded function (columns) a certain number of times over the course of its life cycle. As there are more individuals than functions, the appropriate normalization is division of individuals into tasks (DIT), which is the mutual entropy of individuals and functions, divided by the marginal entropy of individuals [38]. In the formulation used here, the index is bounded between zero (no entropy, corresponding to a pure specialist community) and one (no information, corresponding to a pure generalist community).

We use DIT as an overall index of specialization for the community, showing the extent of shifts in the degree of specialization over time. We can thus distinguish whether these shifts happen spontaneously due to internal community dynamics, or result from our experimental perturbations. As the calculation produces only a single point value of DIT for a given community, we generated nonparametric errors through repeated (5,000×) two-sample bootstrapping of the source community [32]. Details of the calculation of DIT, with a worked example, are shown in Supplement B in the online supplementary materials.

(b) *Mean rank shift of the ten most common ecotypes (MRS$_{10}$)*. Ecologists have often sought to use abundance distributions for making mechanistic inferences about community structuring forces, but such approaches obscure species identity, and cannot detect shifts in dominance. Yet, such internal community dynamics are often the response of interest in manipulative experiments or long-term ecological monitoring, and metrics that incorporate information on compositional identity are highly desirable [70, 71]. An identity-sensitive metric used previously for assessing such community change is *mean rank shift* [24], which quantifies relative changes in species rank abundances over time. We use this metric to assess community change from different temporal perspectives, which we designate the *prospective* (from the past looking into the future) and *retrospective* (from the future looking into the past) comparisons. In the present context, the prospective comparison uses the immediate pre-treatment community as a reference point, while the retrospective comparison uses the end-experiment community state as the reference point. Since *Avida* communities contain many rare ecotypes, we modified this statistic to examine only the ten most common ecotypes in a community, which we designate as MRS$_{10}$. This limitation captures changes in the ecologically most important segment of the community, while filtering out most background noise. Further details of calculation, interpretation, and a worked example are given in online Supplement C. For comparison, we also use the more familiar Bray-Curtis (or Sorenson quantitative) index [68], which works with actual abundances rather than ranks; these results, which are completely congruent with those for MRS$_{10}$, are also presented in online Supplement C (Figure S5). We emphasize the MRS$_{10}$ results because they are more easily interpretable than the Bray-Curtis in terms of the number of ecotypes contributing to a shift of one standardized measurement unit (see the example in online Supplement C).

The DIT (including errors), MRS$_{10}$, and Bray-Curtis metrics were calculated with custom-written MATLAB scripts. Resampling of communities for DIT used functions from MATLAB's Parallel Computing Toolbox.

(c) *Genealogical stability scores (GSS)*. Our DIT and MRS$_{10}$ data are derived purely from *Avida*'s ecotypic output; there is an implicit assumption that for a particular ecotype at a given time, an ancestor-descendent connection exists between the most recent common ancestors (MRCAs) of that ecotype at both that time and earlier times. In order to verify whether or not there were enduring genealogical connections between ecotypes, we derived additional metrics that made use of genealogical information. The first metric is simply the proportion of shared ecotypes between the later and earlier times that had valid genealogical connections. The second is a more elaborate measure that incorporates information from ecotype, abundance, and genealogy. Briefly, when comparing two communities from different points in time that have substantial ecotype overlap and appear stable by one or both of MRS$_{10}$ or DIT, that stability is penalized if there is no genealogical connection between the MRCAs of an ecotype that is present at both times, that is, the ecotype at the later time is an ecological doppelganger of the one at the earlier time, but of different phylogenetic origin. The method, along with worked examples and an illustrative case, is described in detail in online Supplement D. For both these metrics, we compare the ecotypes at the end of the experiment with those in the immediate pre-treatment community (effectively a retrospective comparison).

Transformation of *Avida* population files into phylogenetic tree representations was performed with previously written software [107]. Additional labeling of tree files and filtering of *Avida* lineage and clade data were performed with custom-written MATLAB and R scripts. Tree visualization and MRCA identification were done with *FigTree* v.1.4.2 (http://tree.bio.ed.ac.uk/) and cross-verified with custom-written R scripts using functions from the *ape* [84], *phytools* [86], and *Picante* [55] packages.

2.2.2   Analysis of Time Series Data

Time series of ecological data may be used to infer periods of stability [51], either with reference to a preexisting level of the measured index [69], or by comparing alternative treatments for manipulative experiments [85]. We use both approaches here, particularly for DIT.

We identified stable periods within the MRS_{10} and DIT time series using breakpoint analysis as implemented in the R package *strucchange* [108, 109]. This method has been employed previously in analysis of long-term ecological data to detect ecosystem regime shifts [67]. We use this procedure to investigate the effect of different external perturbations on change in community composition and division of labor. Comparison between different indices can reveal synchronicity and coordination of the changes at different levels, and evidence for stability. Where this method detected significant time series breaks, we always cross-checked the implied community changes against the ecotype lists for time points around the break (before and after), as the event underlying the break may be statistically but not ecologically significant.

We analyzed the DIT data as follows. First, we compared the pre-extinction and end-experiment average levels of DIT. The pre-extinction average level was determined by averaging the values from the immediate pre-treatment time to the first older breakpoint (see above) in the time series. Similarly, the end-experiment level averaged values from the end of the experiment to the first older breakpoint. To handle temporal autocorrelation, we first used iterative ARIMA modeling [26] to determine the optimal values for the autoregressive and moving-average parameters of the time series that would remove most of the autocorrelation. We then used a likelihood ratio test in order to compare the difference between the likelihoods of the ARIMA-adjusted data with those of a null model where the data were subtracted from the global mean of the pre-treatment and end-experiment time series. The aim was to see if the end-experiment state differed systematically from the pre-treatment state in a particular direction with respect to division of labor. We then repeated the above procedure, but this time for the end-experiment averages of the experimental treatments versus those of the corresponding Control replicate. Here, we wished to see if the end-experiment state of a treatment differed systematically from that of a Control. The iterative ARIMA procedure and likelihood ratio test were developed and implemented in R by T. Reitan, using R system library functions.

Second, we further developed a *stability score* based on examination of the entire post-treatment trajectory of each replicate, treating the Control trajectory as a reference level. We took this additional step because the previous analyses may obscure trends in DIT, particularly when a Control replicate itself is unstable. As this additional metric produced results completely congruent with the previously described analyses, we detail its calculation and results in Supplement B (Figure S2, Figure S3, Table S2) in the online supplementary materials.

We analyzed the MRS_{10} data as follows:

(a)   For prospective comparisons, we simply plotted the average time series for each treatment ± two standard errors; trends were sufficiently clear that formal statistical analyses were unnecessary. In order to test whether or not this metric remained static over time for Control, we employed Dunnett's test for multiple comparisons [31], using Control at 1,000 updates after treatment as a reference standard (we could not use the reference community itself, since all starting values for MRS_{10} were zero); this test was performed for prospective comparisons only. Tests were performed in R using standard library functions; the Dunnett test used the *glht( )* function from the *Multcomp* library [45].

(b)   For retrospective comparisons, we noted considerable among-replicate variability in the results. Following treatment, some communities quickly settled into a state indicating low rank shift (and implied high similarity) compared to the end-experiment reference community, whereas others showed clearly more steplike traces when plotted. These results suggested that many populations recovering from extinction did not quickly form

stable, mature post-extinction communities, requiring some way of assessing between-treatment variation in the number of intermediate community states. We therefore used recursive partitioning as implemented in the R system package *rpart* to fit step functions through the retrospective time series data. The plateaus in the predicted data were deemed *locally stable stages (lss)*. If more than 95% of the observations for a particular replicate were less than the control end-experiment lower 95% confidence interval, that replicate was considered to have only a single state. To avoid overfitting of the data (and a resulting very large number of steps), all replicates were analyzed with control parameter settings of *minpart* $= 40$ and *cp* $= 0.01$.

All supporting data and R and MATLAB analysis scripts are available from Dryad (https://doi.org/10.5061/dryad.tq1g3j7). The source code for the versions of *Avida* and *TreeLoader* used to perform the experiments and convert population files into phylogenetic trees can be found at https://www.dropbox.com/home/Luo_etal_ALIFE_SourceCodes.

3 Results

3.1 Overall Summary of Results

In this study, we used digital evolution as instantiated in *Avida* to investigate the stability of cross-feeding communities in response to mass extinction of two different types (Press and Pulse) and two different magnitudes (Strong and Weak). We investigated community stability using three different metrics based on division of labor (DIT), ecotype rank abundance (MRS_{10}), and genealogical connectedness (GSS). We found that:

(i)   The analysis of DIT time series showed that the end-experiment state of Strong Press communities more often differed statistically and ecologically from both the corresponding pre-treatment and end-Control states than did the other treatments, but also had the widest variance of outcomes, with no systematic bias in either direction. Strong Press replicates also tended to establish a more stable post-treatment trajectory of DIT that differed decisively from the corresponding Control trajectory than did the other treatments. However, only Strong Pulse showed a statistically detectable bias towards communities that were more generalized than corresponding Controls. Control and Weak treatments showed few (if any) notable differences in these respects.

(ii)   Using MRS_{10}, we showed that a relatively small portion of the data set could be considered stable, with most (60–65%) Control replicates featuring between one and four replacements of dominant ecotypes over the length of the post-treatment period. Control showed a statistically significant signal of increasing MRS_{10} over time, implying most replicates were not stable by this measure even without treatment. Strong Press communities had the largest and most consistent rank shifts of all treatments, Strong Pulse the second most, and Weak treatments generally did not differ significantly from Control.

(iii)   Both MRS_{10} and DIT showed that recovery was more complicated from Strong treatments than from Control and Weak treatments. Strong treatments required more transitions in ecotype rank abundance for the final community state to take shape, whereas Control and Weak treatments were again quite similar in these characteristics, with fewer rank-abundance transitions and more total replicates that featured only a single locally stable stage (lss) over the length of the post-treatment period.

(iv)   Results using information from genealogy were generally consistent with those from analyses based only on eco-phenotypic data. For ecotypes that were shared between pre-treatment and end-experiment populations, Control and Weak treatments showed the

greatest degree of genealogical continuity (regardless of rank or abundance), Strong Press the least, and Strong Pulse again intermediate between the two groups. When genealogical weightings were applied to a Bray-Curtis-like metric that included both ecotype and abundance information, Control and Weak treatments were all affected about equally, while the two Strong treatments showed greater reduction in stability, with Strong Press again most heavily affected.

3.2   Division of Labor

Using DIT to investigate shifts in overall community specialization clearly illustrates differences in shifts due to internal dynamics versus those due to treatment-driven perturbations. Figure 3 shows two clear cases highlighting those differences. The community shown in panel (a) transitioned spontaneously from a generalized to a pure-specialist state before imposition of treatments, and in the Control replicate remained so for the duration of the simulation, indicating long-term stability of the specialist community. Of the treatments, only Strong Press disrupted this state, causing the emergence of a community dominated by more generalized organisms than either pre-treatment or end-Control states; DIT had still not settled into a new stationary state. All other treatments failed to disrupt the pre-treatment state, and the community remained dominated by specialists. In contrast, the community shown in panel (b) was more strongly generalized before treatments, and remained so in the absence of perturbation. This community was shifted into a more specialized state by the Strong treatments (particularly Strong Pulse), but recovered quickly to its previous state after the Weak treatments. Over the entire data set, using a criterion of a change of ≥0.1 in either direction to indicate ecologically meaningful change, we found among Controls 11 stable communities, 40 average communities, and 49 unstable communities.

3.2.1   Differences between End-Treatment States of DIT from Pre-Treatment and End-Control States

The final state of DIT could be either more generalized or more specialized than the pre-treatment state (Figure 4). Overall, Control had the narrowest range and smallest variance of differences, while Strong Press had the largest (Table 1). However, only for Strong Pulse did the mean difference differ negatively and significantly from zero, indicating a slight tendency for Strong Pulse replicates to recover to a more generalized state than before treatment. The two Strong treatments had a wider
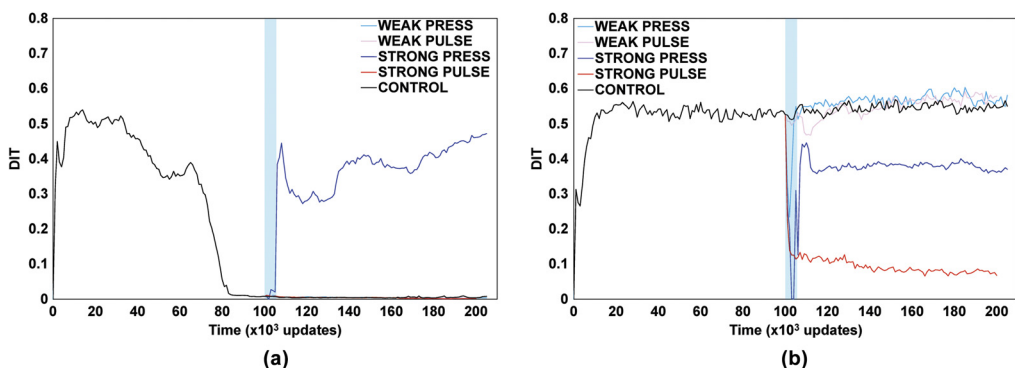

Figure 3. Two exemplar replicates showing change in DIT over time. In both panels, the transparent light blue window represents the Press episode. (a) The pre-treatment community is composed of pure specialists, and the Control remains stable over the second half of the experiment. Strong Press (dark blue trace) results in a much more generalized post-treatment community; all other treatments have no effect on the community makeup as measured through DIT. (b) The pre-treatment community is more generalized than that in panel (a), and remains in this state over the second half of the experiment. Both Strong treatments result in more specialized post-treatment communities (particularly Strong Pulse, red trace). Both Weak treatments have only a slight generalizing effect on the post-treatment community relative to Control.
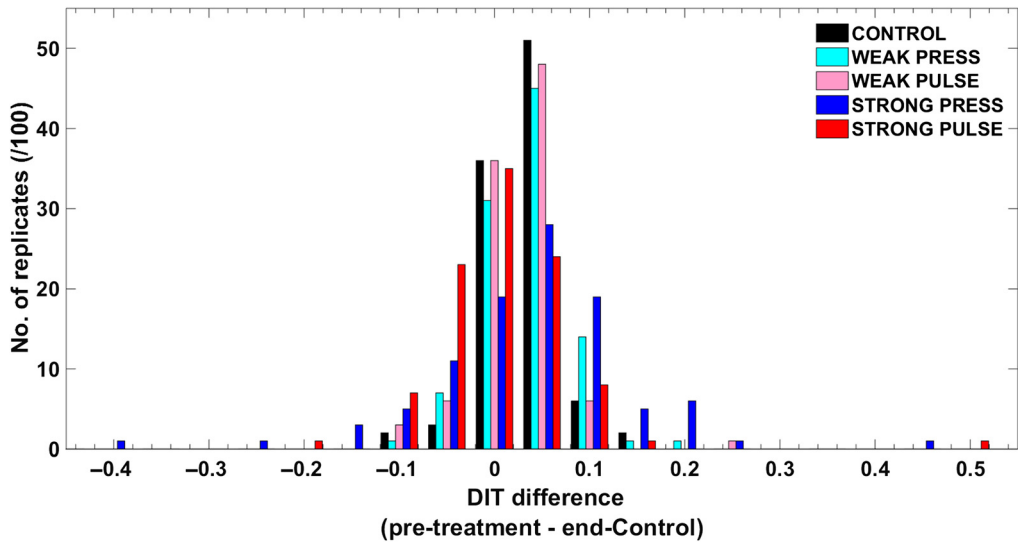
Figure 4. Distributions of differences in DIT scores between pre-treatment and end-treatment states as determined by breakpoint analysis (see Section 2) of DIT time series. Colored bars show number of replicates (vertical axis) with pre-treatment versus end-treatment differences that fall in the specified bins (horizontal axis). Black—Control; pink—Weak Pulse; cyan—Weak Press; red—Strong Pulse; blue—Strong Press. Both Strong treatments show much wider variances than the other treatments (see also Table 1); note also the slight leftward shift for Strong Pulse (modes in bins]−0.1, −0.05] and]−0.05, 0]), indicating tendency towards slightly more generalized post-treatment communities.

variance of differences, differing significantly from Control, Weak treatments, and each other, while the Control and Weak treatments differed only from the Strong treatments. These same tendencies were seen when comparing the end-treatment states with the end-Control state (Supplement B, Table S1; Figure S1 in the online supplementary materials), and when considering the entire post-treatment time series (online Supplement B, Figure S2, Figure S3, Table S2). Except for Strong Pulse, the other treatments showed no systematic bias in end-treatment DIT state versus either pre-treatment or end-Control states.

Table 1. Differences between end-treatment and pre-treatment states of DIT.

Treatment	Range of fitted normal curve	Mean of differences | No. replicates with significant difference (/100)	Variance of differences
(A) Control [D,E]	−0.1–0.1	0.00691 | 43	0.00157
(B) Weak Press [D,E]	−0.13–0.15	0.00854 | 50	0.00205
(C) Weak Pulse [D,E]	−0.13–0.13	0.00244 | 37	0.00204
(D) Strong Press [A,B,C,E]	−0.35–0.35	0.0101 | 58	0.01091
(E) Strong Pulse [A,B,C,D]	−0.25–0.2	−0.0180** | 51	0.00533

Notes. Superscript letters indicate treatments that *are* significantly different from the treatment in the row after correction for multiple comparisons.

Using a change of ≥ 0.1 (in either direction) as the criterion for ecologically meaningful change, we identified among the Controls 11 stable communities, 40 average communities, and 49 unstable communities. There were no significant differences between pre- and post-extinction states for any treatments in stable communities. For average communities, only some replicates yielded significant differences, and differences were always significant in unstable communities.

3.3   Mean Rank Shift

As mentioned previously, we used a criterion of 0.1 as a threshold for stability. We could thus divide the results into communities that were stable (almost always around 0.1 or less), average (between 0.1 and 0.3, up to four replacements of dominant ecotypes), or unstable (0.3 or greater, four or more replacements of dominant ecotypes). Summary results are in Table 2 and Figure 5.

There was significant variation among treatments, with Strong Press in particular having an excess of average and unstable communities (from both perspectives) compared to Control; at the other extreme, Weak Pulse hardly differed at all. The effect of Weak Press was equivocal; while there were fewer stable communities and somewhat more average and unstable communities than for Control, it was not nearly as dramatic as for the Strong treatments. The response of Strong Pulse was intermediate between Strong Press and the Weak treatments, as is more easily seen in the prospective view.

Clearly significant variation among treatments in MRS_{10} was shown for prospective comparisons over nearly all of the post-treatment period (Figure 5). Control communities close in time to the beginning of the post-treatment period did not differ from the (near-)reference community, but on average became increasingly dissimilar over time (Table 3).

Comparing between communities, the largest differences were produced by Strong Press, which differed strongly from all other treatments (Figure 5, dark blue traces). The Control and Weak treatments all showed comparable values and did not differ significantly from each other over

Table 2. Summary results for MRS_{10}.

		No. of stable communities (/100)	No. of average communities (/100)	No. of unstable communities (/100)	χ^2 p (2 d.f.)
Control	Prospective	30	60	10	
	Retrospective	29	64	7	
Weak Press	Prospective*	19	74	7	0.0166
	Retrospective	24	63	13	0.0493
Weak Pulse	Prospective	32	63	5	0.249
	Retrospective	34	63	3	0.206
Strong Press	Prospective***	1	52	47	8.99×10^{-37}
	Retrospective***	0	41	59	1.06×10^{-92}
Strong Pulse	Prospective***	7	82	11	2.5×10^{-06}
	Retrospective***	15	67	18	5.6×10^{-06}

Notes. Chi-square tests use Control values for expectation. No asterisk: not significant; *: $0.01 < p < 0.05$; ***: $p < 0.0001$.
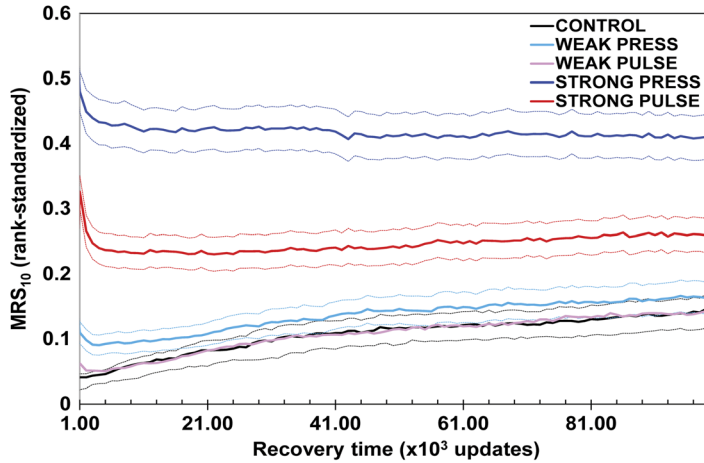
Figure 5. Change in MRS_{10} (see text for definition) over the recovery period, using the immediate pre-extinction populations as reference points for subsequent shifts in relative ranks of dominant ecotypes (prospective comparison). Time has been rescaled so the beginning of the recovery is $t = 0$ for all treatments. Black traces—Control; blue traces—Strong Press; red traces—Strong Pulse; cyan traces—Weak Press; pink traces—Weak Pulse. Bold traces indicate mean MRS_{10} value for the treatment across all 100 replicate communities; thin stippled traces are 95% CIs. Strong Press is the most heavily rank-shifted of all treatments, indicating a tendency to large changes in the ranks of pre-treatment dominant ecotypes. Strong Pulse is intermediate between Strong Press and Control + Weak treatments, indicating less radical but still substantial changes in pre-treatment dominant ecotype ranks. Other treatments tend to change slowly and gradually over recovery time.

the post-treatment interval (except for a brief window very early in the recovery for Weak Press). On average, Weak Press was around the upper 95% confidence intervals of Control and Weak Pulse (Figure 5a, cyan traces), and it had fewer stable replicates than did the former two treatments (Table 2). Strong Pulse showed an intermediate response, but remained statistically distinct from all other treatments (Figure 5a, red traces). These results show that, on average, the Strong treatments resulted in an intermediate to large perturbation of pre-extinction community structure. The Control and Weak treatments resulted in minimal perturbation of community structure (somewhat stronger in the case of Weak Press), with dominant ecotypes generally remaining so. However, even in Control replicates, the large number of average communities (more than half of all replicates: Table 2) and steadily increasing average MRS_{10} over time (Table 3) imply an overall lack of community stability even for this (non-)treatment.

Table 3. Dunnett contrasts for Control MRS_{10} at key post-treatment times, using Control 1,000 updates post-treatment as reference standard.

Post-treatment time (updates)	Estimate (std. err. = 0.0118)
10,000	0.0251 NS
25,000	0.0606***
50,000	0.0908***
75,000	0.0997***
100,000	0.121***

Notes. ***: $p < 0.001$; NS: not significant.

3.3.1   Stepwise Recovery and lss of Retrospective Comparisons

As mentioned previously, for retrospective comparisons, we found a considerable diversity of outcomes for changes in individual replicates, even in Controls (Figure 6). Some cases were quite stable, maintaining low values over the entire post-treatment time period (Figure 6a). In such communities, dominant ecotypes generally maintained their pre-treatment relative ranks, with very little introgression of new types. However, there were also unstable cases, where substantial rank shifts occurred quickly and unpredictably (Figure 6b). Where large rank shifts occurred, at least two pre-treatment dominant ecotypes were replaced by ones with functional combinations that were previously rare or absent; many such communities were mosaics of old and new ecotypes, but complete turnover did not occur. Responses for Strong treatments, however, ranged from intermediate turnover to complete renewal of the dominant types. For prospective comparisons, many communities quickly stabilized at higher MRS_{10} values, implying that the post-extinction community state differed substantially and permanently from the pre-treatment state (Figure 5, dark blue trace; Figure 6c,d, stippled traces). For the corresponding retrospective comparisons, rank shifts could soon stabilize or occur in a more stepwise manner over the post-treatment period (e.g., Figure 6c, solid trace). Our
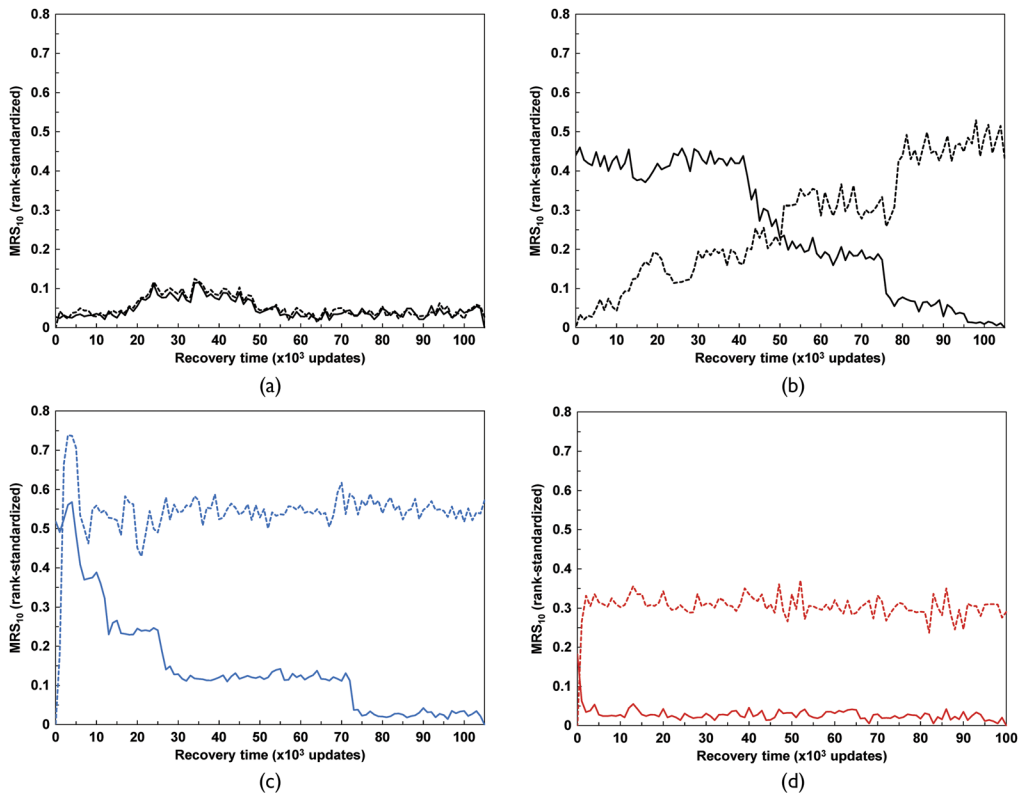

Figure 6. Diversity of outcomes for retrospective MRS_{10}. In all panels, solid trace means retrospective comparison; stippled trace, prospective comparison. (a) A largely stable Control replicate, showing highly synchronous behavior between prospective and retrospective time series. The pre- and end-treatment states are about equivalent in dominant ecotype composition. (b) An unstable Control replicate, showing approximate mirror image symmetry between prospective and retrospective time series. Divergence from the pre-treatment state equates to convergence on the end-treatment state. Three lss are apparent in the retrospective series. (c) A Strong Press replicate showing an asymmetric response between views and clearly stepwise behavior of the retrospective view, indicating transitional post-treatment community states. The prospective view indicates a stable, permanent rank shift, but the retrospective view shows three lss, including that of the end-treatment community. (d) A Strong Pulse replicate where the post-treatment community takes shape rapidly after the mass extinction. There is only a single post-treatment lss. The prospective view again indicates a stable, permanent rank shift of the pre-treatment ecotypes, though not to the extent of the community in panel (c).

Table 4. Between-treatment comparison of lss (see Section 2 for definition) for MRS_{10} retrospective series.

	Lss	No. of replicates with one stage (/100)
(A) Control	2.16 ± 0.129 [D,E]	46
(B) Weak Press	2.05 ± 0.118 [D,E]	48
(C) Weak Pulse	2.10 ± 0.120 [D,E]	43
(D) Strong Press	3.21 ± 0.094 [A,B,C,E]	7
(E) Strong Pulse	2.71 ± 0.110 [A,B,C,D]	19

Notes. Superscript letters indicate treatments that *are* significantly different from the treatment in the row after correction for multiple comparisons.

interpretation of such results is that the formerly dominant pre-treatment ecotypes either quickly went extinct or became very rare. By contrast, the composition of the post-treatment community could take shape either rapidly or in stages over the course of the recovery period, with the largest shifts driven by multiple new dominant types that were ecologically marginal or absent in the pre-treatment community. We analyzed this more closely through determining lss for retrospective comparisons (Table 4).

We found significant variation among treatments for the number of lss of retrospective MRS_{10} ($F_{(4,495)} = 19.04$, $p < 1.41 \times 10^{-14}$). The results were consistent with those for prospective comparisons; there were no significant differences in lss among Control and Weak treatments, Strong Press had on average the most lss, and Strong Pulse was again intermediate between the two groups. When considered on a within-treatment basis, a larger fraction of the data for Control and the Weak treatments showed only a single lss, Strong Press had the fewest such replicates, and Strong Pulse was again intermediate. Seen this way, the Control and Weak treatments tended during recovery to most quickly establish a community state that resembled the end-experiment one (if the immediate pre-treatment community was at all perturbed). Strong Press communities went through the most transitional states before establishing a final state (indeed, many were still in transition at the end of the experiment), while Strong Pulse fell between these extremes.

3.4 Genealogical Stability

The mass extinction treatments affected not only ecological stability as indicated by the preceding metrics, but also the underlying phylogenetic and genealogical structure of the affected populations. The first, simple measure of genealogical stability revealed that markedly fewer ecotypes shared between the end-experiment community and the immediate pre-treatment community maintained genealogical continuity across the treatment and recovery period for the two Strong treatments than for the other treatments (Figure 7a). The fraction of shared ecotypes with connections was quite similar for Control and Weak treatments, with about 78%, 75%, and 72% for Control, Weak Pulse, and Weak Press respectively. This fraction fell to about 53% for Strong Pulse and 42% for Strong Press, indicating an increased occurrence in these treatments of shared ecotypes that were convergently evolved ecological doppelgangers of their pre-treatment counterparts. Importantly, the differences among treatments were in the same direction as MRS_{10} and statistically significant in the same way (one way ANOVA, $F_{(4,495)} = 98.36$, $p < 1.5 \times 10^{-61}$), with significant differences between the Strong treatments after multiple comparison testing.

The second metric, which incorporated information from ecotype identity, abundance, and genealogy, demonstrated that including genealogical information could alter the value of the metric—and thus inference of the degree of community stability—versus the assumption of connections
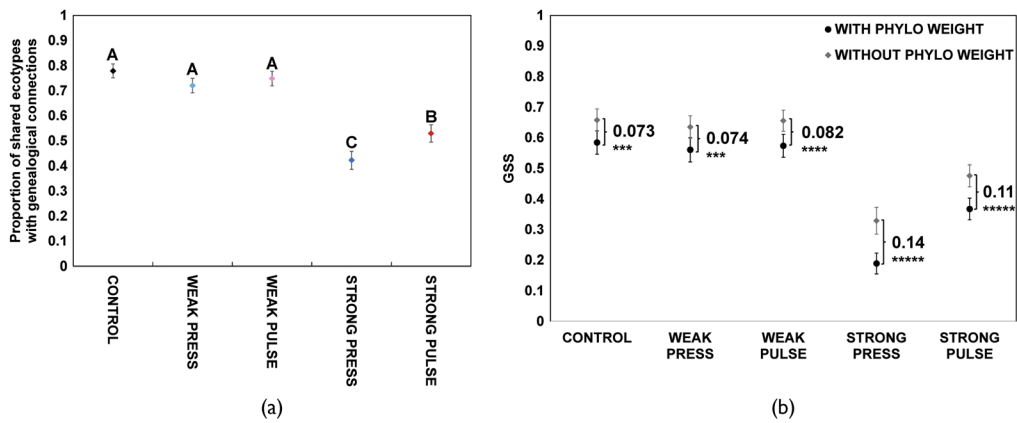
Figure 7. Effect of different treatment types on genealogical stability. Comparison is between end-experiment and pre-treatment populations. In all cases, $n = 100$. All error bars are ±2 standard errors. (a) Proportion of shared ecotypes with genealogical connections. Letters indicate groups of treatments that are *not* significantly different from each other after multiple comparison testing. As with unweighted MRS_{10}, Strong Press shows the strongest effect, followed by Strong Pulse; no significant differences exist between Control and Weak treatments. (b) Abundance-weighted GSS with and without genealogical weightings. Numbers next to brackets indicate absolute difference between weighted and unweighted scores. Weightings have a significant reduction on all treatments; the two Strong treatments show the greatest reduction in stability due to loss of genealogical continuity, while effects for Control and Weak treatments are similar. All comparisons were made with one-tailed *t*-tests at $\alpha = 0.01$. Significance levels: NS, not significant; *, $0.01 < p < 0.05$; **, $0.005 < p < 0.01$; ***, $0.001 < p < 0.005$; ****, $0.0005 < p < 0.001$, *****, $p < 0.0005$.

between all ecotypes (Figure 7b). When unweighted, essentially the same conclusions as using MRS_{10} (Figure 5) or the Bray-Curtis index (online Supplement C, Figure S5) obtain, but these are changes due only to ecotype turnover. Including genealogical weightings reduced the value of the metric further still for all treatments, by an average of 7.3% to 8.2% for Control and Weak treatments, about 11% for Strong Pulse, and about 14% for Strong Press (Figure 7b).

4   Discussion

4.1   Residents versus Tourists in the Ecological Core

Our results provide information about a general pattern in how these cross-feeding digital communities are structured. The generally small values of MRS_{10} (Table 2) and stable values of DIT (Table 1) observed in the Control treatment suggest that most replicate populations contain an *ecological core*: a set of organisms that provides most ecosystem services and ecological structure for the population. Around these core organisms exist many less common, often transient types generated by mutation. This is analogous to natural ecosystems that feature *resident* and *tourist* species, the former characterized by long-standing persistence and high abundance, the latter by infrequent occurrence and low abundance [70], a key difference being that in real communities on ecological time scales, nearly all tourists enter the community through migration, rather than in situ through speciation. Large shifts in either MRS_{10} or DIT indicate an alteration of that core, usually through replacement of several residents by former tourist types. Such shifts were observed mainly in the Strong treatment communities. However, even undisturbed Control communities could also often experience substantial ecotype turnover (e.g., Figure 6b), as demonstrated by the significant signal of increasing departure (Table 3) from near-pre-treatment rank abundance (though not to the extent of the Strong treatments). This may result from delayed initial evolution of ecotypes with top-tier functions (particularly the computationally more difficult XOR and EQUALS functions), or replacement of once-dominant ecotypes by new ones with altered functional groupings, even after all functions have evolved. One or more tourist ecotypes may displace previous core members, though it may take several independent appearances of these tourists for that to occur, as the initially rare types must

first dodge drift [60], and be at least equal (or superior) in fitness to one of the existing dominant types. It is unclear if this slower turnover of ecotypes is due more to Red Queen phenomena [63] or community drift processes [46, 64]. We hypothesize that such replacements might alter relative fitnesses and ecological relationships of other community members, rearranging prevailing ecological structure and influencing subsequent evolution of the community [3, 8], and we plan to test this directly in future work.

4.2 Strong Press Produces the Most Radical Result

On average, the Weak extinction treatments were insufficiently disruptive to produce large changes beyond those of Control populations, though their effect was somewhat more substantial in the case of Weak Press (Figure 5). In stark contrast, the Strong treatments upset the previous ecological core, allowing for the evolution and establishment of a new set of core types. This was most clearly evident in the case of Strong Press: It had the highest average prospective MRS_{10} (Figure 5, Table 2), its communities required on average the most transitions between locally stable post-treatment states and generally passed through the most intermediate steps before settling into a terminal-like state (Table 4), it produced the largest range of differences in DIT (Figure 4, Table 1), and it had the most reduced genealogical stability (Figure 7). This treatment, as we have implemented it, acts as more than a simple filter of organisms, as do the Pulse treatments: It often resulted in near-total ecosystem collapse, effectively erasing the previous core at functional-genomic, ecological, and phylogenetic levels [105–107]. Ecosystem reestablishment started nearly from scratch, from a phenotypically altered, functionally reduced pool of Avidians better adapted to the harsh conditions of the Strong Press episode. This difference in post-treatment evolutionary landscape is also reflected in the relative contributions of genealogically unconnected ecotypes in Press versus Pulse (Supplement D, Table S4 in the online supplementary materials). The greater contribution from unconnected types in Pulse treatments demonstrates that, even with near-total annihilation of the pre-treatment community, the unaltered survivors in many cases retain the genetic and functional potential to regenerate lost ecotypes with the same ecological profiles, though there is also of course room for vacated ecospace to be occupied by newly evolved organisms with different ecological profiles. The kinds of changes resulting from Press treatments, on the other hand, favor replacement of marginalized or lost ecotypes by functionally different ones (which may be ecologically similar, but not identical), leading to a more thoroughly reshaped and restructured community.

4.3 DIT Reveals a Key Difference between (Strong) Random and Selective Extinction

In contrast to the highly visible differences created by Strong Press, Strong Pulse was generally intermediate in most aspects between Strong Press and Control on one hand and the Weak treatments on the other (Figure 5, Tables 1, 2, 4). In our implementation of pulse extinction, survival is random, and a population may be reestablished by a subset of Avidians ranging anywhere from all common, core organisms at one extreme to all rare, tourist organisms at the other. Previous work [104] showed a probability of 0.49 that at least two of the Strong Pulse survivors would come from among the two most common ecotypes. Further, the overall generalized nature of most replicate populations (average pre-treatment DIT $= 0.465 \pm 0.011$ (s.e.), about 2–3 functions per Avidian) means that even a sample as small as those used for Strong Pulse may actually still contain many of the functions needed to rebuild the ecology (albeit in heavily reduced amounts), free of the eroding effects of a press episode. Survivors belonging to previously dominant groups may then quickly re-assume key ecological roles. Thus, communities recovering even from a Strong Pulse extinction might retain some "memory" of their previous state.

This is supported by the lower prospective MRS_{10} than for Strong Press (Figure 5), and, more interestingly, the fact that among the distributions of DIT difference scores (Figure 4; Table 1), only the mean of Strong Pulse was significantly different (albeit small) from zero, indicating a slight tendency towards more generalized post-treatment communities—a result shown by no other

treatment. For Control and Weak treatments, the numerical differences are simply not large enough (even if statistically significant: Figures 4, 5; Tables 1, 2, 4) to imply substantial regular ecological differences between the communities. Under Strong Press, destruction of the pre-treatment community is often so thorough that there is effectively no relation between the pre- and post-treatment communities with respect to DIT. The post-treatment community is just as likely to be more specialized as more generalized, as shown by the wider variance in DIT comparisons (Table 1; Figure 4), as well as the signed stable trajectory score (Supplement B, Table S2 in the online supplementary materials). These results highlight an important difference between the random Pulse treatment and the selective Press treatment. Generalist survivors may well have an early advantage in post-treatment expansion, as they can use multiple resources and may in many cases be able to survive by self-feeding. However, the difference here seems to be more than just a simple short-term expansion of lucky generalized survivors, in that Strong Pulse actually leaves a small, but enduring, signature in the community's functional makeup.

4.4   Shifts to New Stable States and Early Recovery Instability

We were also concerned with whether our digital communities shifted into new stable states following a large disruption, that is to say, a community characterized by different ecological and taxonomic makeup, even with the same underlying abiotic environment, that does not easily revert to the previous one [29, 49, 85, 93]. For this digital system, the possibility of post-extinction alternative states was raised in previous work [106], based on simple phenotypic diversity-through-time results. We have explored it more fully here, as raw phenotypic diversity can indicate recovery to similar diversity levels but hide significant changes in community makeup. Our results show that sufficiently strong disruption could indeed often drive these digital communities into new states with a different taxonomic and ecological balance, shown by high prospective MRS_{10} indicating a heavily altered set of core ecotypes, substantially altered division of labor as revealed through DIT, or both.

However, there are two key caveats to these results. The first is that while strong mass extinction often succeeded in driving communities into new states, we cannot say that they consistently represented *alternative stable* states—the assumption being that the immediate pre-treatment community itself represented a stable state. Indeed, our data indicate that stable pre-treatment communities were relatively uncommon. The overall signal of increasing prospective MRS_{10} for Control (Table 3), as well as the substantial number of Control communities that were deemed average and unstable (Table 2) or showed a significant difference in DIT from the pre-treatment state (Table 1), demonstrate that even without perturbation many communities were still subject to change in the second half of the experiment. The point where the treatment was imposed divided the experiment into convenient before and after phases, but the community itself need not have been stable at that point (e.g., Figure 6b). Hence, it is more accurate to say that the treatments (especially the Strong ones) drove many communities into a state that differed substantially from that achieved by the Control, regardless of Control stability.

The second consideration is that even when new stable states emerged, they often did not result quickly following a Strong treatment. The higher average number of lss in retrospective MRS_{10} and the low number of replicates that quickly converged to a community state with long-term stability (Table 4) indicate that after a Strong treatment, many early post-treatment communities are unstable and not yet mature. A persistent, stable community state may only emerge after a prolonged period of time, after passing through several transitional states (e.g., Figure 6c). Such early post-treatment instability was implied by changes in phylogenetic properties of these communities [107], and we have now demonstrated that more directly. Not all post-treatment instability resulted in transition to an alternative state: There were certainly a number of communities where early instability was followed by an eventual return to pre-treatment Control values. However, caution must be exercised with this interpretation, particularly depending on what metric is used to infer stability. Different types of communities may produce similar DIT values but have different ecotype

composition. For any given replicate experiment, it is better to infer stability from consistency across different metrics—for example, if a return to pre-treatment values of DIT is accompanied by low rank shift, also indicating return to pre-treatment ecotype composition and ranking. Explicit consideration of phylogeny/genealogy then becomes necessary to determine whether ecotypes shared between pre- and post-treatment communities really are genealogically connected.

4.5 Genealogical Changes Compound the Effects of Ecotype Turnover

Since we designated ecotypes based on the presence or absence of rewarded computational functions, the possibility existed that new Avidians with the same ecotype as those from an already existing clade, but of different phylogenetic origin, could arise and replace their eco-equivalents during the course of an experiment, an occurrence that would go undetected. Additionally, the computational functions of Avidians can be re-evolved in lineages where they once existed and were lost due to the erosive effects of the press episode [105]. Hence, we had to establish whether there were indeed genealogical connections between the same ecotypes at widely separated points in time.

Both our measures of genealogical stability revealed that in all treatments, changes in genealogical connections further impacted stability beyond ecotype turnover alone (Figure 7). For all treatments, the genealogical changes alone do not contradict the ecological results (which would make the latter meaningless); neither do they cancel the ecological results, since they occur in the same direction and compound rather than counteract ecological changes. Although Control and Weak treatments differed little from each other, inclusion of genealogical information revealed that while Weak Press had about the same fraction of shared ecotypes as did Weak Pulse, its difference from the unweighted metric was slightly smaller, indicating less of an impact from genealogical changes than for Weak Pulse. Close inspection of the data revealed that in Weak Press, a smaller fraction of total replicates had at least one genealogically unconnected shared ecotype as compared to in Weak Pulse. Instead, there was a somewhat greater fraction of replicates where new dominant ecotypes arose from extreme rarity (Supplement D, Table S4 in the online supplementary materials). In Weak Pulse, on the other hand, a slightly larger fraction of replicates had at least one genealogically unconnected shared ecotype, and fewer with new ecotypes arising from extreme rarity (online Supplement D, Table S4).

Taken together, these results suggest that the Press and Pulse treatments exert their effects in different ways. The (slightly) more ecologically extensive changes seen in Weak Press are due more to press-driven ecotype turnover (and associated ecological restructuring) that has a lesser phylogenetic impact on shared ecotypes (but see [107]). By contrast, in Weak Pulse the phylogenetic impact on shared ecotypes is somewhat greater, with more convergent re-evolution of lost ecotypes and resulting lower ecological impact (which usually hardly differs from that in Control). The aforementioned tendencies also occurred in the Strong treatments, as both MRS_{10} and unweighted GSS showed that the major effect of Strong Press is due to extensive ecotype turnover. Including genealogical information shows that the radical changes seen in most Strong Press communities are in fact due to the combined effect of both this ecotype turnover and tremendous phylogenetic attrition [107]. While the ecological restructuring wrought by Strong Pulse is generally less than that by Strong Press, the additional contribution from genealogy is nearly as large, also consistent with previous results that examined only phylogeny [107]. Hence, for a system such as this one where ecotype identity is defined through performance of particular ecological functions, additional consideration of genealogical/phylogenetic information can further illuminate the degree to which apparent ecological stability conceals an underlying phylogenetic dynamism.

4.6 Whither Community Stability in Digital Evolution?

One of our original goals was to determine if these digital communities showed sufficient evidence for stability that would result in coordinated stasis. While our digital data differ in a number of ways

from available paleontological data, we can still assess coordinated stasis as a *pattern* through relative stability of rank abundance, community specialization, and genealogical connectedness among eco-types over appreciable spans of simulation time. When only our eco-phenotypic metrics are considered together, only a fairly small fraction of Control replicates (25%) could be considered to show coordinated stasis, inferred through a combination of very low MRS_{10} and near-constancy of DIT across the whole second half of the experiment, but this becomes less convincing after allowing for genealogical connectedness. Thus, we must conclude that the experimental conditions used here lack the necessary ecological complexity to consistently produce sufficiently stable communities where even the *pattern* of coordinated stasis is prevalent. Rather, our data better support an alternative paradigm of *structural continuity* [76, 88, 89] for these communities: Community *structure* is more or less persistent, but the particular actors (species, ecotypes, etc.) that fill particular ecological roles come and go dynamically. Our results, and the conclusion of the prevalence of structural continuity, are also consistent with previous results using the same digital system, with a different experimental and analytical setup [81]. However, given the existence of capabilities for movement and spatial heterogeneity of resources in *Avida* [35], examining the current leading hypothesis for coordinated stasis—habitat tracking [15, 18]—is an interesting and plausible direction for future work.

4.7 Parallels and Contrasts with Paleontology

Previous work has covered the broad qualitative similarities of the results produced by this digital system to major extinction and recovery events from Earth history [106]. While many of the particulars of our current results are likely to be system-specific, we can draw further broad parallels based on the present results and advances that have been made in paleontology since that time.

Initially immature post-extinction communities and stepwise assembly. Our results showed that rapid formation of a new community that was stable over the whole post-treatment interval was relatively uncommon for the Strong treatments; a more stepwise dynamic, indicating several different intermediate stages before emergence of the end-experiment community, was more common. Real post-extinction paleocommunities can also display such a variety of dynamics, as the component species and their ecological relationships change and stabilize in a succession-like process [47]. Even if overall trophic structure recovers quickly [94, 100], communities from the earliest stages of recovery may be atypical in comparison with later, more persistent community states, again best observed in post-Permian Triassic communities in both terrestrial [9, 90, 92] and marine [94] contexts, but also evident in the early Paleogene from a number of independent lines of evidence (reviewed in [30]). However, clearly stepwise recovery may depend on geographic context and may not be general at a global scale (compare [11, 22] versus [94]). In our digital system, these post-extinction turnovers and transitions may be due to delayed re-evolution of new functions [106] or a superior new implementation of already existing functions. The Cretaceous terrestrial revolution and Early Eocene Climate Optimum (reviewed in [47]) provide real paleontological analogues.

Radical community turnover after devastating extinction, and new stable states. Our results also showed that Strong Press extinctions in particular resulted in the greatest tendency to produce new ecosystems that differed most substantially in ecological and genealogical makeup from both the pre-treatment and end-Control states, with substantial contribution from ecotypes that were previously marginal or absent. Such extensive restructuring of communities is a hallmark of the most ecologically disruptive mass extinctions and recoveries of the Phanerozoic, particularly the end-Permian [28, 101] and end-Cretaceous [1, 30] extinctions (although the latter is usually interpreted as a canonical pulse extinction [27]), which led to phyletically different sets of taxa becoming dominant members of the biota and the emergence of new ecological structures. The less-extensive restructuring produced by Strong Pulse extinctions more often resulted in post-treatment communities

made up of a variety of pre-treatment survivors, genealogically unconnected re-evolved ecotypes, and formerly rare or new ecotypes with increased ecological prominence. The arrival and establishment of these new ecological players would act to prevent a return to previous states [83, 98]. The early Triassic [96] and early Paleogene [30] recoveries provide perhaps the best analogies, since while many of the available niches were filled through diversification from surviving taxa, others continued to be occupied by carryover taxa (e.g., *Lystrosaurus* across the Permo-Triassic boundary), or were filled by radiation from lineages that were related to, but had previously been less prominent than, the previous occupants—for example, the replacement of ecologically diversified archaic bird lineages by the modern neornithine clade in the early Paleogene [65]. The Weak treatments (particularly Weak Pulse), which generally did not produce significant departure from Control, are perhaps more reminiscent of the end-Ordovician, which, while devastating to organisms on lower taxonomic levels, did not eliminate many higher taxa (order or more inclusive) and did not have the well-documented long-term impacts of the end-Permian and end-Cretaceous events [41].

Shift to increased generalization in Strong Pulse communities. Our DIT results indicated a tendency for a slight (but still detectable) shift to more generalized communities after Strong Pulse. This result has some interesting analogies with the paleontological record, where certain mass extinctions (end-Permian [92], end-Triassic [19], and end-Cretaceous [59]) have resulted in post-extinction communities with increased ecological generalization and/or biogeographic cosmopolitanism. Such a shift is projected for the modern biota, given ongoing anthropogenic extinction [13, 54, 78]. Unlike these paleontological examples, in which the increased generalization/cosmopolitanism eventually declines, our data show a (marginally) persistent effect of the extinction and recovery on division of labor.

5   Concluding Remarks

Our results illustrate the potential for digital evolution to make many aspects of studying mass extinction, recovery, and community stability more transparent than is possible with the fossil record, and treat them in a truly experimental manner. We can study how evolution would proceed under different sets of circumstances and connect observed differences to the types of manipulations. We have shown that different metrics measure different facets of community change, and demonstrated that applying several approaches yields a fuller understanding of community change across multiple levels. This gives us confidence that further use and development of evolving digital systems, combined with measurement approaches borrowed from community ecology, can yield additional insight into problems in paleobiology and long-term evolution that are not amenable to other experimental approaches.

While one must exercise great caution in extrapolating results from a digital system like ours to the real world, we believe that our results have bearing on potential future trends for the ecology and evolution of the world's biota. It has become increasingly accepted that we are in the midst of an anthropogenic sixth mass extinction [20, 73, 75]. Human activities driving the extinction can be considered a continued environmental pressure that favors the survival of certain taxa while threatening and ultimately removing others, which our Press extinctions are analogous to. A key implication of our study is that full recovery from even a relatively mild biotic crisis often entails a transformed community at both ecological and phylogenetic levels, the latter being more pronounced with increasing ecological severity. If the biotic world comes to be dominated by taxa that both survive and thrive in conditions created by human activity, then, depending on the severity of the effects (whether weak or strong), recovery of the biosphere may ultimately result in radically altered ecosystems highly divergent from those we know today, with new taxonomic and ecological balances—as has already happened after the most severe mass extinctions in Earth history—that are

difficult (if not impossible) to reverse and that, depending on their component species, may not be helpful to human well-being.

Acknowledgments

References

1. Aberhan, M., & Kiessling, W. (2015). Persistent ecological shifts in marine molluscan assemblages across the end-Cretaceous mass extinction. *Proceedings of the National Academy of Sciences of the U.S.A.*, *112*(23), 7207–7212. http://doi.org/10.1073/pnas.1422248112.

2. Adami, C. (2006). Digital genetics: Unravelling the genetic basis of evolution. *Nature Reviews Genetics*, *7*(2), 109–118.

3. Adams, J., & Oeller, P. W. (1986). Structure of evolving populations of *Saccharomyces cerevisiae*: Adaptive changes are frequently associated with sequence alterations involving mobile elements belonging to the Ty family. *Proceedings of the National Academy of Sciences of the U.S.A.*, *83*(18), 7124–7127. http://doi.org/10.1073/pnas.83.18.7124.

4. Alroy, J. (2008). Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences of the U.S.A.*, *105*(suppl. 1), 11536–11542.

5. Alroy, J. (2010). The shifting balance of marine diversity among major animal groups. *Science*, *329*(5996), 1191–1194.

6. Barraclough, T. G., Birky, C. W. Jr., & Burt, A. (2003). Diversification in sexual and asexual organisms. *Evolution*, *57*, 2166–2172.

7. Batten, R. L. (1973). The vicissitudes of the gastropods during the interval of Guadalupian–Ladinian time. In A. Logan & L. V. Hills (Eds.), *The Permian and Triassic systems and their mutual boundary. Canadian Society of Petroleum Geologists Memoir 2* (pp. 596–607). Calgary: Canadian Society of Petroleum Geologists.

8. Bell, G. (1997). *Selection: The mechanism of evolution*. New York: Chapman & Hall.

9. Benton, M. J., Tverdokhlebov, V. P., & Surkov, M. V. (2004). Ecosystem remodelling among vertebrates at the Permian-Triassic boundary in Russia. *Nature*, *432*, 97–100. http://doi.org/10.1038/nature02950.

10. Benton, M. J., & Twitchett, R. J. (2003). How to kill (almost) all life: The end-Permian extinction event. *Trends in Ecology & Evolution*, *18*(7), 58–365.

11. Benton, M. J., Zhang, Q., Hu, S., Chen, Z. Q., Wen, W., Liu, J., Huang, J., Zhou, C., Xie, T., Tong, J., & Choo, B. (2014). Exceptional vertebrate biotas from the Triassic of China, and the expansion of marine ecosystems after the Permo-Triassic mass extinction. *Earth-Science Reviews*, *137*, 85–128. http://doi.org/10.1016/j.earscirev.2014.08.004.

12. Bertram, S. M., & Gorelick, R. (2009). Quantifying and comparing mating systems using normalized mutual entropy. *Animal Behaviour*, *77*(1), 201–206. http://doi.org/10.1016/j.anbehav.2008.09.028.

13. Blois, J. L., Zarnetske, P. L., Fitzpatrick, M. C., & Finnegan, S. (2013). Climate change and the past, present, and future of biotic interactions. *Science*, *341*(6145), 499–504. http://doi.org/10.1126/science.1237184.

14. Blount, Z. D., Barrick, J. E., Davidson, C. J., & Lenski, R. E. (2012). Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature*, *489*(7417), 513–518. http://doi.org/10.1038/nature11514.

15. Brett, C. E. (2012). Coordinated stasis reconsidered: A perspective at fifteen years. In J. A. Talent (Ed.), *Earth and Life: Global Biodiversity, Extinction Intervals, and Biogeographic Perturbations through Time* (pp. 23–36). Dordrecht, Heidelberg: Springer.

16. Brett, C. E., & Baird, G. (1995). Coordinated stasis and evolutionary ecology of Silurian to Middle Devonian faunas in the Appalachian Basin. In R. Anstey & D. H. Erwin (Eds.), *Speciation in the fossil record* (pp. 285–315). New York: Columbia University Press.

17. Brett, C. E., Bartholomew, A. J., & Baird, G. C. (2007). Biofacies recurrence in the Middle Devonian of New York state: An example with implications for evolutionary paleoecology. *Palaios*, *22*(3), 306–324. http://doi.org/10.2110/palo.2005.p05-027r.

18. Brett, C. E., Hendy, A., Bartholomew, A. J., Bonelli, J. R., & McLaughlin, P. I. (2007). Response of shallow marine biotas to sea level fluctuations: Faunal replacement and the process of habitat tracking. *Palaios*, *22*(3), 228–244.

19. Button, D. J., Lloyd, G. T., Ezcurra, M. D., & Butler, R. J. (2017). Mass extinctions drove increased global faunal cosmopolitanism on the supercontinent Pangaea. *Nature Communications*, *8*(1). http://doi.org/10.1038/s41467-017-00827-7.

20. Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the U.S.A.*, E6089–E6096. http://doi.org/10.1073/pnas.1704949114.

21. Cleland, H. F. (1903). *A study of the fauna of the Hamilton Formation of the Cayuga Lake section in central New York*. U.S. Government Printing Office.

22. Chen, Z. Q., & Benton, M. J. (2012). The timing and pattern of biotic recovery following the end-Permian mass extinction. *Nature Geoscience*, *5*(3), 375–383. http://doi.org/10.1038/ngeo1475.

23. Chow, S. S., Wilke, C. O., Ofria, C., Lenski, R. E., & Adami, C. (2004). Adaptive radiation from resource competition in digital organisms. *Science*, *305*(5680), 84–86. http://doi.org/10.1126/science.1096307.

24. Collins, S. L., Suding, K. N., Cleland, E. E., Batty, M., Pennings, S. C., Gross, K. L., Grace, J. B., Gough, L., Fargione, J. E., & Clark, C. M. (2008). Rank clocks and plant community dynamics. *Ecology*, *89*(12), 3534–3541. http://doi.org/10.1890/07-1646.1.

25. Cooper, T., & Ofria, C. (2002). Evolution of stable ecosystems in populations of digital organisms. In R. K. Standish & M. A. Bedau (Eds.), *Eighth International Conference on Artificial Life, December 9–13, Sydney, New South Wales* (pp. 227–232). Cambridge, MA: MIT Press.

26. Crawley, M. J. (2013). *The R Book* (2nd ed.). Chichester, U.K.: Wiley. https://doi.org/10.1007/s007690000247.

27. D'Hondt, S., Donaghay, P., Zachos, J. C., Luttenberg, D., & Lindinger, M. (1998). Organic carbon fluxes and ecological recovery from the Cretaceous-Tertiary mass extinction. *Science*, *282*(5387), 276–279.

28. Dineen, A. A., Fraiser, M. L., & Sheehan, P. M. (2014). Quantifying functional diversity in pre- and post-extinction paleocommunities: A test of ecological restructuring after the end-Permian mass extinction. *Earth-Science Reviews*, *136*, 339–349. http://doi.org/10.1016/j.earscirev.2014.06.002.

29. Downing, A. S., van Nes, E. H., Mooij, W. M., & Scheffer, M. (2012). The resilience and resistance of an ecosystem to a collapse of diversity. *PLoS ONE*, *7*, e46135.

30. Dunne, J. A., Labandeira, C. C., & Williams, R. J. (2014). Highly resolved early Eocene food webs show development of modern trophic structure after the end-Cretaceous extinction. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1782), 20133280. http://doi.org/10.1098/rspb.2013.3280.

31. Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, *50*(272), 1096–1121. http://doi.org/10.1080/01621459.1955.10501294.

32. Efron, B., & Tibshirani, R. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman and Hall/CRC Press.

33. Eldredge, N., & Gould, S. J. (1972). Punctuated equilibrium: An alternative to phyletic gradualism. In T. J. M. Schopf (Ed.), *Models in paleobiology* (pp. 82–115). San Francisco: Freeman, Cooper.

34. Eldredge, N., Thompson, J. N., Brakefield, P. M., Gavrilets, S., Jablonski, D., Jackson, J. B. C., Lenski, R. E., Lieberman, B. S., McPeek, M. A., & Miller III, W. (2003). The dynamics of evolutionary stasis. *Paleobiology*, *31*(2), 133–145.

35. Elsberry, W. R., Grabowski, L. M., Ofria, C., & Pennock, R. T. (2009). Cockroaches, drunkards, and climbers: Modeling the evolution of simple movement strategies using digital organisms. In C. L. Nehaniv, H. A. Abbass, & M. Sugisaka (Eds.), *2009 IEEE Symposium on Artificial Life, ALIFE 2009—Proceedings* (pp. 92–99). https://doi.org/10.1109/ALIFE.2009.4937699.

36. Erwin, D. H. (1998). The end and the beginning: Recoveries from mass extinctions. *Trends in Ecology and Evolution*, *13*(9), 344–349. http://doi.org/10.1016/S0169-5347(98)01436-0.

37. Gorelick, R. (2006). Combining richness and abundance into a single diversity index using matrix analogues of Shannon's and Simpson's indices. *Ecography*, *29*(4), 525–530.

38. Gorelick, R., Bertram, S. M., Killeen, P. R., & Fewell, J. H. (2004). Normalized mutual entropy in biology: Quantifying division of labor. *American Naturalist*, *164*(5), 677–682.

39. Gorelick, R., & Bertram, S. M. (2007). Quantifying division of labor: Borrowing tools from sociology, sociobiology, information theory, landscape ecology, and biogeography. *Insectes Sociaux*, *54*(2), 105–112.

40. Hallam, A., & Wignall, P. B. (1997). *Mass extinctions and their aftermath*. Oxford, UK: Oxford University Press.

41. Harper, D. A. T., Hammarlund, E. U., & Rasmussen, C. M. Ø. (2014). End Ordovician extinctions: A coincidence of causes. *Gondwana Research*, *25*(4), 1294–1307. http://doi.org/10.1016/j.gr.2012.12.021.

42. Heard, S. B., & Mooers, A. Ø. (2002). Signatures of random and selective mass extinction in phylogenetic tree balance. *Systematic Biology*, *51*(6), 889–897. http://doi.org/10.1080/10635150290155917.

43. Holland, S. M. (1995). The stratigraphic distribution of fossils. *Paleobiology*, *21*(1), 92–109.

44. Holland, S. M., & Zaffos, A. (2011). Niche conservatism along an onshore-offshore gradient. *Paleobiology*, *37*(2), 270–286. http://dx.doi.org/10.1666/10032.1.

45. Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

46. Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton, N.J.: Princeton University Press.

47. Hull, P. (2015). Life in the aftermath of mass extinctions. *Current Biology*, *25*(19), R941–R952. http://doi.org/10.1016/j.cub.2015.08.053.

48. Ingalls, B. R., & Park, L. E. (2010). Biotic and taphonomic response to lake-level fluctuations in the Greater Green River Basin (Eocene), Wyoming. *Palaios*, *25*(5), 287–298. http://doi.org/10.2110/palo.2009.p09-048r.

49. Isbell, F., Tilman, D., Polasky, S., Binder, S., & Hawthorne, P. (2013). Low biodiversity state persists two decades after cessation of nutrient enrichment. *Ecology Letters*, *16*, 454–460.

50. Ivany, L. C., Brett, C. E., Wall, H. L. B., Wall, P. D., & Handley, J. C. (2009). Relative taxonomic and ecologic stability in Devonian marine faunas of New York State: A test of coordinated stasis. *Paleobiology*, *35*(4), 499–524. http://doi.org/10.1666/0094-8373-35.4.499.

51. Ives, A. R., Dennis, B., Cottingham, K. L., & Carpenter, S. R. (2003). Estimating community stability and ecological interactions from time-series data. *Ecological Monographs*, *73*(2), 301–330.

52. Jablonski, D. (1986). Evolutionary consequences of mass extinctions. In D. M. Raup & D. Jablonski (Eds.), *Life science research reports, vol. 36. Patterns and processes in the history of life* (pp. 313–330). Berlin: Springer.

53. Jablonski, D. (2005). Mass extinctions and macroevolution. *Paleobiology*, *31*(2), 192–210.

54. Jackson, J. B. C. (2008). Colloquium paper: Ecological extinction and evolution in the brave new ocean. *Proceedings of the National Academy of Sciences of the U.S.A.*, *105*(Supplement 1), 11458–11465.

55. Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464. https://doi.org/10.1093/bioinformatics/btq166.

56. Kidwell, S. M., & Holland, S. M. (2002). The quality of the fossil record: Implications for evolutionary analyses. *Annual Review of Ecology and Systematics*, *33*(1), 561–588. http://doi.org/10.1146/annurev.ecolsys.33.030602.152151.

57. Kinnersley, M., Wenger, J. W., Sherlock, G., & Rosenzweig, F. R. (2011). Rapid evolution of simple microbial communities in the laboratory. In P. Pontarotti (Ed.), *Evolutionary biology: Concepts, biodiversity, macroevolution and genome evolution* (pp. 107–120). Berlin: Springer.

58. Kirchner, J. W., & Well, A. (2000). Delayed biological recovery from extinctions throughout the fossil record. *Nature*, *404*(6774), 177–180. http://doi.org/10.1038/35004564.

59. Labandeira, C. C., Johnson, K. R., & Wilf, P. (2002). Impact of the terminal Cretaceous event on plant-insect associations. *Proceedings of the National Academy of Sciences of the U.S.A.*, *99*(4), 2061–2066. http://doi.org/10.1073/pnas.042492999.

60. LaBar, T., & Adami, C. (2017). Evolution of drift robustness in small populations. *Nature Communications*, *8*(1). http://doi.org/10.1038/s41467-017-01003-7.

61. Lenski, R. E., Ofria, C., Collier, T. C., & Adami, C. (1999). Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, *400*(6745), 661–664. http://doi.org/10.1038/23245.

62. Lenski, R. E., Ofria, C., Pennock, R. T., & Adami, C. (2003). The evolutionary origin of complex features. *Nature*, *423*(6936), 139–144. http://doi.org/10.1038/nature01568.

63. Liow, L. H., Van Valen, L., & Stenseth, N. C. (2011). Red Queen: From populations to taxa and communities. *Trends in Ecology and Evolution*, *26*(7), 349–358. http://doi.org/10.1016/j.tree.2011.03.016.

64. Loreau, M., & de Mazancourt, C. (2008). Species synchrony and its drivers: Neutral and nonneutral community dynamics in fluctuating environments. *American Naturalist*, *172*(2), E48–E66. http://doi.org/10.1086/589746.

65. Longrich, N. R., Tokaryk, T., & Field, D. J. (2011). Mass extinction of birds at the Cretaceous-Paleogene (K-Pg) boundary. *Proceedings of the National Academy of Sciences of the U.S.A.*, *108*(37), 15253–15257. https://doi.org/10.1073/pnas.1110395108.

66. Lu, P. J., Yogo, M., & Marshall, C. R. (2006). Phanerozoic marine biodiversity dynamics in light of the incompleteness of the fossil record. *Proceedings of the National Academy of Sciences of the U.S.A.*, *103*(8), 2736–2739. http://doi.org/10.1073/pnas.0511083103.

67. Lynam, C. P., Lilley, M. K. S., Bastian, T., Doyle, T. K., Beggs, S. E., & Hays, G. C. (2011). Have jellyfish in the Irish Sea benefited from climate change and overfishing? *Global Change Biology*, *17*(2), 767–782. http://doi.org/10.1111/j.1365-2486.2010.02352.x.

68. Magurran, A. E. (2003). *Measuring biological diversity*. Malden, MA: Wiley-Blackwell.

69. Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M., Elston, D. E., Scott, E. M., Smith, R. I., Somerfield, P. J., & Watt, A. D. (2010). Long-term datasets in biodiversity research and monitoring: Assessing change in ecological communities through time. *Trends in Ecology and Evolution*, *25*(10), 574–582. https://doi.org/10.1016/j.tree.2010.06.016.

70. Magurran, A. E., & Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, *422*(6933), 714–716. http://doi.org/10.1038/nature01547.

71. Magurran, A. E., & Henderson, P. A. (2010). Temporal turnover and the maintenance of diversity in ecological assemblages. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1558), 3611–3620. http://doi.org/10.1098/rstb.2010.0285.

72. Martin, L. D., & Meehan, T. J. (2005). Extinction may not be forever. *Naturwissenschaften*, *92*(1), 1–19.

73. McCallum, M. L. (2015). Vertebrate biodiversity losses point to a sixth mass extinction. *Biodiversity and Conservation*, *24*(10), 2497–2519. http://doi.org/10.1007/s10531-015-0940-6.

74. McGhee, G. R., Clapham, M. E., Sheehan, P. M., Bottjer, D. J., & Droser, M. L. (2013). A new ecological-severity ranking of major Phanerozoic biodiversity crises. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *370*, 260–270. http://doi.org/10.1016/j.palaeo.2012.12.019.

75. McKinney, M. L. (1997). Extinction vulnerability and selectivity: Combining ecological and paleontological views. *Annual Review of Ecology and Systematics*, *28*(1), 495–516. http://www.annualreviews.org/doi/10.1146/annurev.ecolsys.28.1.495.

76. Miller III, W. (1996). Ecology of coordinated stasis. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *127*(1–4), 177–190. http://doi.org/10.1016/S0031-0182(96)00094-6.

77. Morris, P. J., Ivany, L. C., Schopf, K. M., & Brett, C. E. (1995). The challenge of paleoecological stasis: Reassessing sources of evolutionary stability. *Proceedings of the National Academy of Sciences of the U.S.A.*, *92*(24), 11269–11273. http://doi.org/10.1073/pnas.92.24.11269.

78. Myers, N., & Knoll, A. H. (2001). The biotic crisis and the future of evolution. *Proceedings of the National Academy of Sciences of the U.S.A.*, *98*(10), 5389–5392.

79. Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton, N.J.: Princeton University Press.

80. Ofria, C., Bryson, D., & Wilke, C. (2009). Avida: A software platform for research in computational evolutionary biology. In A. Adamatzky & M. Komosinski (Eds.), *Artificial Life Models in Software* (2nd ed., pp. 3–35). London, U.K.: Springer.

81. Olson, R. S., Mirmomeni, M., Brom, T., Bruger, E., Hintze, A., Knoester, D. B., & Adami, C. (2013). Evolved digital ecosystems: Dynamic steady state, not optimal fixed point. In P. Liò, O. Miglino, G. Nicosia, S. Nolfi, & M. Pavone (Eds.), *Advances in artificial life (ECAL 2013*, pp. 126–133). Cambridge, MA: MIT Press.

82. Olszewski, T. D., & Erwin, D. H. (2009). Change and stability in Permian brachiopod communities from western Texas. *Palaios*, *24*(1), 27–40. https://doi.org/10.2110/palo.2008.p08-061r.

83. Palumbi, S. R., McLeod, K., & Grunbaum, D. (2008). Ecosystems in action: Lessons from marine ecology about recovery, resistance, and reversibility. *Bioscience*, *58*(1), 33–42.

84. Paradis, E. (2012). *Analysis of phylogenetics and evolution with R* (2nd ed.). New York: Springer.

85. Petraitis, P. S., & Dudgeon, S. R. (2004). Detection of alternative stable states in marine communities. *Journal of Experimental Marine Biology and Ecology*, *300*(1–2), 343–371. https://doi.org/10.1016/j.jembe.2003.12.026.

86. Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*, 217–223.

87. Reymond, C. E., Bode, M., Renema, W., & Pandolfi, J. M. (2011). Ecological incumbency impedes stochastic community assembly in Holocene foraminifera from the Huon Peninsula, Papua New Guinea. *Paleobiology*, *37*(4), 670–685. http://dx.doi.org/10.1666/09087.1.

88. Rodríguez, J. (2004). Stability in Pleistocene Mediterranean mammalian communities. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *207*(1–2), 1–22. http://doi.org/10.1016/j.palaeo.2003.12.016.

89. Rodríguez, J. (2006). Structural continuity and multiple alternative stable states in Middle Pleistocene European mammalian communities. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *239*(3–4), 355–373. http://doi.org/10.1016/j.palaeo.2006.02.001.

90. Roopnarine, P. D., Angielczyk, K. D., Wang, S. C., & Hertog, R. (2007). Trophic network models explain instability of Early Triassic terrestrial communities. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1622), 2077–2086. http://doi.org/10.1098/rspb.2007.0515.

91. Rozen, D. E., Schneider, D., & Lenski, R. E. (2005). Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *Journal of Molecular Evolution*, *61*(2), 171–180. http://doi.org/10.1007/s00239-004-0322-2.

92. Sahney, S., & Benton, M. J. (2008). Recovery from the most profound mass extinction of all time. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1636), 759–765. http://doi.org/10.1098/rspb.2007.1370.

93. Scheffer, M., & Carpenter, S. (2003). Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends in Ecology and Evolution*, *18*, 648–656.

94. Scheyer, T. M., Romano, C., Jenks, J., & Bucher, H. (2014). Early Triassic marine biotic recovery: The predators' perspective. *PLoS ONE*, *9*(3), e88987. http://doi.org/10.1371/journal.pone.0088987.

95. Sepkoski, J. J. (1984). A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. *Paleobiology*, *10*(2), 246–267. http://doi.org/10.1017/S0094837300008186.

96. Sidor, C. A., Vilhena, D. A., Angielczyk, K. D., Huttenlocker, A. K., Nesbitt, S. J., Peecook, B. R., Steyer, S. J., Smith, R. M. H., & Tsuji, L. A. (2013). Provincialization of terrestrial faunas following the end-Permian mass extinction. *Proceedings of the National Academy of Sciences of the U.S.A.*, *110*(20), 8129–8133. http://doi.org/10.1073/pnas.1302323110.

97. Solé, R. V., Montoya, J. M., & Erwin, D. H. (2002). Recovery after mass extinction: Evolutionary assembly in large-scale biosphere dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*(1421), 697–707. http://doi.org/10.1098/rstb.2001.0987.

98. Suding, K. N., Gross, K. L., & Houseman, G. R. (2004). Alternative states and positive feedbacks in restoration ecology. *Trends in Ecology and Evolution*, *19*(1), 46–53.

99. Twitchett, R. J. (2001). Incompleteness of the Permian-Triassic fossil record: A consequence of productivity decline? *Geological Journal*, *36*(3–4), 341–353. http://doi.org/10.1002/gj.883.

100. Twitchett, R. J., Krystyn, L., Baud, A., Wheeley, J. R., & Richoz, S. (2004). Rapid marine recovery after the end-Permian mass-extinction event in the absence of marine anoxia. *Geology*, *32*(9), 805–808. http://doi.org/10.1130/G20585.1.

101. Van Roy, P., Orr, P. J., Botting, J. P., Muir, L. A., Vinther, J., Lefebvre, B., el-Hariri, K., & Briggs, D. E. G. (2010). Ordovician faunas of Burgess Shale type. *Nature*, *465*(7295), 215–218. http://doi.org/10.1038/nature09038.

102. Vrba, E. S. (1985). Environment and evolution: Alternative causes of the temporal distribution of evolutionary events. *South African Journal of Science*, *81*, 229–236.

103. Wagner, P. J., Kosnik, M. A., & Lidgard, S. (2006). Abundance distributions imply elevated complexity of post-Paleozoic marine ecosystems. *Science*, *314*(5803), 1289–1292. http://doi.org/10.1126/science.1133795.

104. Yedid, G., & Heier, L. (2012). Effects of random and selective mass extinction on community composition in communities of digital organisms. In P. Pontarotti (Ed.), *Evolutionary biology: Mechanisms and trends* (pp. 43–64). Berlin, Heidelberg: Springer.

105. Yedid, G., Ofria, C. A., & Lenski, R. E. (2008). Historical and contingent factors affect re-evolution of a complex feature lost during mass extinction in communities of digital organisms. *Journal of Evolutionary Biology*, *21*(5), 1335–1357. http://doi.org/10.1111/j.1420-9101.2008.01564.x.

106. Yedid, G., Ofria, C. A., & Lenski, R. E. (2009). Selective press extinctions, but not random pulse extinctions, cause delayed ecological recovery in communities of digital organisms. *American Naturalist*, *173*(4), E139–E154. http://doi.org/10.1086/597228.

107. Yedid, G., Stredwick, J., Ofria, C. A., & Agapow, P. M. (2012). A comparison of the effects of random and selective mass extinctions on erosion of evolutionary history in communities of digital organisms. *PLoS ONE*, *7*, e37233. http://doi.org/10.1371/journal.pone.0037233.

108. Zeileis, A., Kleiber, C., Walter, K., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics and Data Analysis*, *44*(1–2), 109–123. http://doi.org/10.1016/S0167-9473(03)00030-6.

109. Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, *7*(2), 1–38. http://doi.org/10.18637/jss.v007.i02.