Nonequilibrium evolution of volatility in origination and extinction explains fat-tailed fluctuations in Phanerozoic biodiversity

Andrew J. Rominger1*, Miguel A. Fuentes1,2,3, Pablo A. Marquet1,4,5,6,7,8

Fluctuations in biodiversity, large and small, pervade the fossil record, yet we do not understand the processes generating them. Here, we extend theory from nonequilibrium statistical physics to describe the fat-tailed form of fluctuations in Phanerozoic marine invertebrate richness. Using this theory, known as superstatistics, we show that heterogeneous rates of origination and extinction between clades and conserved rates within clades account for this fat-tailed form. We identify orders and families as the taxonomic levels at which clades experience interclade heterogeneity and within-clade homogeneity of rates, indicating that families are subsystems in local statistical equilibrium, while the entire system is not. The separation of timescales between within-clade background rates and the origin of major innovations producing new orders and families allows within-clade dynamics to reach equilibrium, while between-clade dynamics do not. The distribution of different dynamics across clades is consistent with niche conservatism and pulsed exploration of adaptive landscapes.

INTRODUCTION

Biodiversity has not remained constant nor followed a simple trajectory through geologic time (1–5). Instead, it has been marked by fluctuations in the richness of taxa, both positive in the case of net origination and negative in the case of net extinction. Major events, such as adaptive radiations and mass extinctions, have received special attention (6, 7), but fluctuations of all sizes are ubiquitous (2, 5) and follow a fat-tailed distribution, where large events are more probable compared to, e.g., a Gaussian distribution. Understanding the fat-tailed nature of these fluctuations continues to elude paleobiologists and biodiversity theoreticians.

The fat-tailed distribution of fluctuations in taxon richness inspired earlier researchers to invoke ideas from complex systems with similar distributions. These ideas include the hypotheses that biological systems self-organize to the brink of critical phase transitions (8, 9) and that environmental perturbations are highly nonlinear (10). More recent data and analyses have not, however, supported these hypotheses at the scale of the entire Phanerozoic marine invertebrate fauna (5, 11). Other studies have modeled the mean trend in taxon richness as tracking a potentially evolving equilibrium (2, 12, 13) and yet ignore the role of stochasticity and nonequilibrium dynamics in producing observed patterns (4, 14, 15). Individual-, population-, and local ecosystem-scale processes that could produce complex dynamics, such as escalatory coevolutionary interactions (16), have not been documented to scale up to global patterns (17) and should not be expected to do so (18). Thus, we still lack a theory to describe the striking fat-tailed nature of fluctuations throughout the Phanerozoic.

Despite the heterogeneity of explanations of Phanerozoic biodiversity, consensus has emerged on one property of macroevolution: Clades experience different rates of morphological evolution, origination, and extinction (2, 3, 19, 20). Here, we show that the simple fact of conserved rates within clades and variable rates across clades is sufficient to describe pervasive, fat-tailed fluctuations in taxonomic richness throughout the marine Phanerozoic. This biological mechanism has a precise correspondence to the nonequilibrial theory from statistical physics known as “superstatistics” (21), which has been applied across the physical and social sciences (22, 23). We leverage this correspondence to explain the distribution of fluctuations in the standing richness of marine invertebrates preserved in the Phanerozoic fossil record. We further show that the specific mathematical form of this superstatistical distribution is consistent with niche conservatism (24, 25) and pulsed exploration on an adaptive landscape by higher taxa (19, 25–27). We operationally define “adaptive landscape” to mean a clade’s set of characteristics, and the fitness they impart to the clade, that influences its macroevolution. Those characteristics could be ecological [e.g., substrate preference (25, 28, 29)], morphological [e.g., body plan (14)], or macroecological [e.g., range size (30, 31)].

Superstatistics of fossil biodiversity

Superstatistics (21) proposes that nonequilibrial systems can be decomposed into many local subsystems, each of which attains a unique dynamic equilibrium. The evolution of these dynamic equilibria across subsystems occurs more slowly. This separation in time scales allows local systems to reach equilibrium, while the system as a whole is not (21). In the context of macroevolution, we propose that a clade with conserved macroevolutionary rates corresponds to a subsystem in dynamic equilibrium.

In statistical mechanics, local subsystems can be defined by a simple statistical parameter $\beta$ often corresponding to inverse temperature. In macroevolutionary “mechanics,” we define $\beta_k$ of clade $k$ as the inverse variance of fluctuations $x_k$ in the number of genera within that clade, i.e., fluctuations in the genus richness. $\beta_k$ values thus represent the inverse variances, what we term volatilities, of the origination-extinction
processes of genera with clades. The details of this origination-extinction process, e.g., whether it is linear or subject to a carrying capacity, are not of central importance to our analysis; so long as fluctuations can reach a stationary distribution and are observed over time-averaged intervals in a temporally coarse-grained fossil record, they will be approximately Gaussian [see section S1; (32)].

We make the hypothesis of dynamic equilibrium within a clade following MacArthur and Wilson (33) in recognition that while the identity and exact number of taxa will fluctuate stochastically from random origination and extinction (taking the place of local immigration and extinction in island biogeography), the overall process determining the number of taxa, and, by extension, fluctuations in that number, is in equilibrium. The different regions of adaptive space occupied by different clades can be conceptualized as islands with unique dynamic equilibria, albeit with macroevolutionary processes determining the “colonization” of adaptive peaks, as opposed to short time scale biogeographic processes.

The volatility of richness fluctuations will vary across these islands in adaptive space as an emergent trait of a clade resulting from the macroevolutionary fitness of the clade and the shape of the surrounding adaptive landscape. Ultimately, volatility emerges from the life histories, ecologies, and evolutionary histories that drive each clade’s macroevolutionary fitness and characterize its occupancy of different regions of an adaptive landscape. We do not attempt to diagnose which characteristics of different regions account for volatility differences, but others have found rates of origination and extinction to depend on larval type (34), body plan (14), body size (30), range size (30, 31), and substrate preference (25). Not all of these traits would be considered dimensions of an ecological niche or characteristics of a guild (28, 29, 35), but they all point to different strategies that influence a clade’s macroevolutionary success. These characteristics result from interactions between heritable traits and environments, which themselves may be thought of as semi-heritable (36). Thus, different regions of adaptive space, and the clades occupying them, will experience different magnitudes of stochastic fluctuations in taxonomic richness. As clades occasionally split to fill new regions of adaptive space, their pulsed diversification determines the nonequilibrium nature of the entire biota.

Real paleontological data to test superstatistics
To uncover the superstatistical nature of the marine invertebrate Phanerozoic fauna, we analyzed the distribution of fluctuations in genus richness (the lowest reliably recorded taxonomic resolution) using the Paleobiology Database (PBDB; paleobiodb.org). We corrected these raw data for incomplete sampling and bias using a new method described in Materials and Methods. Occurrences from the PBDB were matched to 49 standard time bins, all of approximately 11–million year (Ma) duration following previous publications (5, 12). Fluctuations in genus richness were calculated as the simple difference between bias-corrected richnesses in adjacent time bins.

To focus attention on the variance of fluctuations, we zero-centered each clade’s fluctuation distribution. In this way, we focus on fluctuations about any possible trend toward net diversification or extinction. Because “equilibrium” in the statistical mechanical sense means that a system undergoes coherent, concerted responses to perturbation, the mean trend line (positive or negative) is of less interest than deviations from it. We also note that the distributions of fluctuations for most clades are already very close to a mean of 0 (mean at the family level: 0.038 ± 0.176 SD), and so centering has little influence on clade-specific fluctuation distributions, consistent with the observation that origination is often roughly equal to extinction (37). Following (23), we also ignore all instances of no change (i.e., zero fluctuation).

We define potentially equilibrial subsystems based on taxonomic hierarchies, as a full phylogenetic hypothesis for all marine invertebrates is lacking. Taxa ideally represent groups of organisms that descend from a common ancestor and share similar ecologically and evolutionary relevant traits (38, 39). Thus, our model assumes that at a given higher taxonomic level, within-taxon fluctuations in richness are driven by equilibrial processes characterized by Gaussian distributions. We further assume that new higher taxa arise due to the emergence of sufficiently novel traits (be they ecological, morphological, life history, or macroecological) so that those new taxa occupy a new region of an adaptive landscape. We lastly assume that different regions of adaptive space are characterized by different volatilities in origination and extinction.

To evaluate the optimal taxonomic level for subsystem designation, we test our superstatistical theory using taxonomic levels from family to phylum. In addition, we compare our results to randomized taxonomies and confirm that the observed fit of superstatistics is not an artifact of arbitrary classification but instead represents real, biologically relevant diversification processes within and between clades. We find that families and orders conform to the assumptions of our superstatistical model, while classes and phyla do not.

RESULTS
We first evaluate the local equilibria of clades from family level to phylum. We find that family-level fluctuation distributions are well approximated by Gaussians (Fig. 1 and fig. S3). Three exemplar family-level dynamics are highlighted in Fig. 1 to illustrate how different volatility equilibria express themselves as actual richness time series. This Gaussian approximation also largely holds for orders, but classes and phyla increasingly show deviations from Gaussian with greater kurtosis, corresponding to more frequent outliers at these taxonomic levels (fig. S3).

To predict the superstatistical behavior of the entire marine invertebrate Phanerozoic fauna, we must integrate over all possible local equilibria that each clade could experience. The stationary distribution of \( \beta_k \) values describes these possible equilibria, specifying the probability that a given clade, chosen at random, will occupy a region of adaptive space characterized by \( \beta_0 \).

We estimate the distribution of \( \beta_k \) values as the maximum likelihood distribution describing the set of volatilities for all families, orders, classes, or phyla. Phanerozoic marine invertebrate families follow a gamma distribution in their \( \beta_k \) values (Fig. 1). The gamma distribution also holds for orders but shows increasing deviations again for classes and especially for phyla (fig. S4).

Using the observation of within-family statistical equilibria and gamma-distributed \( \beta_k \) parameters, we can calculate, without further adjusting free parameters, the distributions of family-level fluctuations for the entire marine Phanerozoic, \( P(x) \), as

\[
P(x) = \int_0^\infty p_k(x \mid \beta)f(\beta)d\beta
\]

where \( p_k(x \mid \beta) = \sqrt{\frac{\pi}{2\beta}} \exp\left(-\frac{x^2}{2\beta}\right) \) is the distribution of fluctuations within a family and \( f(\beta) = \frac{1}{\Gamma(b_k/2)} \left(\frac{b_k}{2\beta_k}\right)^{b_k/2} \beta_k^{b_k/2-1} \exp\left(-\frac{b_k\beta}{2\beta_k}\right) \) is the stationary
distribution of volatilities in richness fluctuations. The integral in Eq. 1 leads to

\[ P(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \sqrt{\frac{b_0}{\pi b_1}} \left( 1 + \frac{b_0 x^2}{b_1} \right)^{-\frac{n+1}{2}} \]  

(2)

This corresponds to a non-Gaussian, fat-tailed prediction for \( P(x) \), which closely matches aggregated family-level fluctuations in the bias-corrected PBDB (Fig. 2).

To quantitatively evaluate how well the superstatistical prediction matches the family- and order-level data, we constructed a 95% confidence envelope from bootstrapped maximum likelihood estimates of \( P(x) \). Observed fluctuations for both taxonomic levels fall within these 95% confidence envelopes (Fig. 2), indicating that the data do not reject the superstatistical prediction. For further comparison, we fit a Gaussian distribution to the observed fluctuations, which corresponds to the equilibrium hypothesis that all families conform to the same dynamic. Using Akaike Information Criterion (AIC), we find that observed fluctuations are considerably better explained by the superstatistical prediction than by the Gaussian hypothesis (\( \Delta AIC = 1895.622 \)). Thus, as expected under the superstatistical hypothesis, the fat-tailed distribution of fluctuations arises from the superposition of independent Gaussian statistics of fluctuations within families. Computing the distribution of aggregated fluctuations using orders also closely matches the observed data (Fig. 2), but as we further coarsen the taxonomy to classes and phyla, we see increasingly poorer correspondence between data and theory (Fig. 2).

We quantify this change in the goodness of fit with the Kolmogorov-Smirnov statistic (Fig. 3). We can see that both families and orders have low Kolmogorov-Smirnov statistics, and order-level designation of equilibrial subsystems performs slightly better than the family level. Classes are substantially worse, and phyla are worse yet, with the Kolmogorov-Smirnov statistic of phyla being no different than the null randomized taxonomies described below.

However, if superstatistical theory explains the data, then this worsening fit with increasing taxonomic scale is expected, as the different classes and phyla should not represent dynamically equilibrial subsystems in their fluctuation dynamics. Instead, classes and phyla aggregate increasingly disparate groups of organisms and thus effectively mix their associated Gaussian fluctuations, meaning that one statistic should no longer be sufficient to describe class- and phylum-level dynamics. We see this confirmed by the increasing frequency of outlier fluctuations in within-class and phylum-level fluctuation distributions (fig. S3). We can also see that families and orders represent, on average, one to two ecospace hypercubes [defined by taxon environment, motility, life habit, vision, diet, reproduction, and ontogeny (28, 29, 35)], respectively. In contrast, classes and phyla represent, on average, 8 to 30 hypercubes, respectively (fig. S5).

Our analysis indicates that orders and families are evolutionarily coherent units with all subsumed taxa sharing key ecological and evolutionary attributes, allowing them to reach steady-state diversification independently from other clades at the global scale. The fact that both orders and families conform to theoretical predictions is consistent with superstatistics. If superstatistics operates at the order level, then
the families subsumed by these orders should represent random realizations of their order’s stationary $\beta_k^{(\text{order})}$ volatility. The sum of gamma random variables is still gamma, but with new parameters, and so the family-level distribution of $\beta_k^{(\text{family})}$ is still gamma.

To further test the evolutionary coherence of families, we conducted a permutation experiment in which genera were randomly reassigned to families while maintaining the number of genera in each family. For each permutation, we calculated the superstatistical prediction and its Kolmogorov-Smirnov statistic. The permutation simulates a null model in which common evolutionary history is stripped away (genera are placed in random families), but the total number of observed genera per family is held constant. Because we ignore all instances of no change (i.e., zero fluctuation), we remove any possible large and artificial gaps in the genus occurrences of these permuted clades. Controlling for the total number of genera per family is key because this could be purely an artifact of an arbitrary taxonomic prescription of fluctuations, this landscape cannot be flat (i.e., equal fitness everywhere) but instead must be rugged. Thus, niche conservatism accounts for all downstream superstatistical results. The permutation test is specifically designed to determine if the $\beta_k$ values imposed by this correlation with richness are sufficient to explain the observed superstatistical fit.

Repeating the null permutation of genera in families 500 times yields a null distribution of Kolmogorov-Smirnov statistics that is far separated from the observed values at the family and order levels (Fig. 3), suggesting that the good fit at these levels is not merely a statistical artifact of classification or the richness of clades but carries important biological information. Classes approach the null and phyla are no different. It should also be noted that the width of 95% confidence interval of this null distribution is not far from the distance between the Kolmogorov-Smirnov statistics of orders versus families, suggesting that differences of fit between these taxonomic levels are at least partially accounted for by the randomness of the sampling distribution of Kolmogorov-Smirnov statistics.

**DISCUSSION**

Our analysis makes no assumption that orders and families should correspond to superstatistical subsystems, but identifies them as the appropriate level for marine invertebrates. Our study is the first to demonstrate that complex patterns in the fluctuation of taxon richness in the fossil record are the result of a simple underlying process analogous to the statistical mechanisms by which complexity emerges in large, nonequilibrium physical (22) and social systems (23). We do so by identifying the biological scale at which clades conform to locally independent dynamic equilibria in fluctuations. Equilibrium could result from many processes, including neutrality (33, 42), diversity dependence (43, 44), and processes that dampen, rather than exacerbate, fluctuations in complex ecological networks (45). These candidate processes are directly opposed to the presumption of instability underlying the self-organized criticality hypothesis for paleobiodiversity (8, 9).

We show that the distribution describing the evolution to different equilibria between orders and families is gamma (Fig. 1). A gamma distribution, while consistent with multiple processes, could result from evolution of diversification rates across an adaptive landscape that promotes niche conservatism and pulsed exploration of niche space (46). Specifically, if $\beta_k$ values are associated with a clade’s macroevolutionarily relevant traits, and those traits evolve via Ornstein-Uhlenbeck–like exploration of an adaptive landscape, then the resulting stationary distribution of $\beta_k$ will be gamma (46). For macroevolutionary rates to vary in a way consistent with the observed superstatistical description of fluctuations, this landscape cannot be flat (i.e., equal fitness everywhere) but instead must be rugged. Thus, niche conservatism around local fitness optima in adaptive space interrupted by adaptive exploration is likely (27, 47). The specifics of how this adaptive landscape is shaped and is traversed by evolving clades determine the exact form of the distribution of $\beta_k$ volatilities, in the case of the marine Phanerozoic resulting in a gamma distribution. Our work thus motivates further study of the trait spaces and evolutionary shifts consistent with gamma-distributed equilibria in richness fluctuation volatilities.

We show that the pulsed shift to different equilibria between orders and the families they subsume is sufficient to explain the characteristically fat-tailed distribution of richness fluctuations when the marine Phanerozoic invertebrate fauna is viewed as a whole macrosystem. Armed with an understanding of the statistical origin of this diversification pattern, we can explore which models of niche conservatism...
and pulsed adaptive radiation are consistent with the statistical behavior of the Phanerozoic. Our statistical theory provides new motivation for identifying the eco-evolutionary causes of innovations between lineages and how those innovations are eventually conserved within lineages. Using the superstatistical prediction as a theoretical baseline, we can also go on to identify and robustly examine the mechanisms underlying deviations from statistical theory. For example, some clades wax and wane systematically, and possibly nonsymmetrically, through time (4, 13, 31), a pattern that we cannot explain with superstatistics alone.

Superstatistics could also be applied to other areas of evolution and macroecology. For example, recently proposed phylogenetic models already consider heterogeneous rates of diversification (e.g., (20)), as expected between different subsystems. The superstatistics of clades in adaptive landscapes could motivate models that jointly predict changes in traits and diversification, a research area currently struggling with model inadequacy (48). This framework could also provide a previously unexplored paradigm in modeling the distributions of richness, abundance, and resource use in non-neutral communities, which can be viewed as emerging from the combination of locally equilibrium subsystems. Non-neutral models in ecology are criticized for their overparameterization (49), yet a persistent counterargument to neutral theory (42) is the unrealistic assumption of ecological equivalency and poor prediction of real dynamics (49). If ecosystems are viewed as the superposition of many individualistically evolving clades, each exploiting the environment differently and thus obeying a different set of statistics, then diversity dynamics could be parsimoniously predicted with superstatistics while incorporating real biological information on ecological differences between taxa.

Superstatistics is a powerful tool to derive macroscale predictions from locally fluctuating subsystems whose evolution is driven by interesting, but complex and difficult to model, biological mechanisms. Hence, applications of superstatistics to a wide variety of patterns in ecological and evolutionary systems are ripe for exploration.

MATERIALS AND METHODS
All data processing and analyses were performed in R (50), and all codes needed to reproduce our study are provided, with added explanation, in Supplementary Appendix A.

Paleobiology database data download and filtering
Data on individual fossil occurrences and the ecospace characteristics of Phanerozoic marine invertebrates were downloaded from PBDB (https://paleobiodb.org) on 16 November 2018 via the database’s application programming interface (API) (data retrieval and processing script available in the Supplementary Materials). Collections were filtered using the same approach as Alroy et al. (5) to ensure that only well-preserved marine invertebrate occurrences were used in subsequent analyses. This filtering resulted in 815,222 unique genus-level occurrences. These were further filtered to exclude those occurrences without family-level taxonomy and those collections with age estimate resolutions outside the 11-Ma time bins proposed by Alroy et al. (5), resulting in 454,033 occurrences. Time bins were compiled from http://fossilworks.org with a custom script reproduced in the Supplementary Materials. The first and last of these time bins, corresponding to the earliest Cambrian and the latest Cenozoic, were excluded from analysis because their sampling completeness (see below) could not be assessed.

Correcting for imperfect and potentially biased sampling
We developed and used a new and flexible method to correct for known sampling incompleteness and biases in publication-based specimen databases (5, 12). Incompleteness is inherent in all biodiversity samples, the fossil record being no exception (51–54). In addition to incompleteness, bias may result from preferential publication of novel taxa (12), which exacerbates the difference between poorly sampled and well-sampled time periods. We therefore developed a simple two-step method: We first corrected genus richness for incomplete sampling using the “three-timer” correction (5) and then further corrected this three-timer richness estimate by accounting for any correlation between the number of genera and the number of publications in a time period.

The three-timer correction estimates the probability of failure to observe a genus in a given time period \( p_t \) as the number of times any genus is recorded before and after that period but not during, divided by the number of genera whose occurrence histories span the period \( t \). To calculate the sampling-corrected richness \( \hat{D}_k \) of a clade \( k \) in the time period in question, the observed genera within that clade and time period were divided by \( 1 - p_t \), and their occurrences were summed

\[
\hat{D}_k = \sum_{j \in k} \frac{I_{jt}}{1 - p_t}
\]

where \( j \in k \) designates genera in clade \( k \) and \( I_{jt} \) is an indicator equal to 1 if genus \( j \) occurs in time period \( t \).

\( \hat{D}_k \) is the maximum likelihood estimator of richness in a simple occupancy through time type model assuming binomial sampling (55) and, in that way, mimics other proposed methods for the fossil record (52, 53). We avoid parametrically modeling the sampling process through time by instead taking a sliding window of time bins from the Cambrian to the Cenozoic. It should be noted that the three-timer correction compares favorably to other similar methods to account for imperfect detection (56).

To eliminate further bias due to preferential publication of novel taxa (12), we divided the three-timer–corrected number of genera per family per time period by the expected number of genera, given publications in that time period. The expected number was calculated by regressing the log-transformed three-timer–corrected number of genera on log-transformed number of publications. There is only a weak trend toward higher richness with more publications (fig. S1), meaning that the most important correction comes from the three-timer correction.

Our new method rescales each genus occurrence from 0 or 1 (absent or present) to a weighted number continuously ranging between 0 and 1. Because these weighted numbers represent sampling and bias-corrected occurrences, we can add them arbitrarily, corresponding to the membership of any given genus in any given higher taxonomic group. We must, however, choose a taxonomic level at which to evaluate the relationship between richness and publications; we chose the level of family because this is the most finely resolved option.

We opted not to use subsampling methods (12, 51, 57) because these approaches would not be advisable for clades with few genera. However, our new method achieves similar results to subsampling procedures at the global scale across all clades. We directly compared our predicted time series of global fluctuations in genus richness, with results derived from rarefaction and shareholder quorum subsampling.
Supplementary materials

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/6/eaat0122/DC1

Section S1. Limit distribution of a time-averaged homogeneous origination-extinction process

Section S2. Evaluation of sampling bias correction methods

Section S3. Understanding deviations from superstatistics at higher taxonomic levels

Section S4. Ecospace occupation of higher taxa

Section S5. Relationship between $\beta_0$ and clade richness

Fig. S1. Relationship between number of publications and three-timer corrected genus richness at the family level as recorded by the PBDB.

Fig. S2. Comparison of rarefaction (black line) and SQS (blue) with our three-timer and publication bias correction method (red).

Fig. S3. Change in within-clade richness fluctuation distributions with increasing taxonomic level.

Fig. S4. Change in the distributions of $\beta_0$ across clades of increasing taxonomic level.

Fig. S5. Relationship between number of ecospace hypercubes occupied and taxonomic level.

Fig. S6. Distribution of genus richnesses within families.

Fig. S7. Distribution of genus richnesses within families.

Appendix A R code to reproduce the study

References and notes


47. S. Gavrilets, Fitness Landscapes and the Origin of Species (Princeton Univ. Press, 2004), vol. 41.


Acknowledgments: We thank J. Harte, R. Gillespie, L. Schneider, J. Ying Lim, and D. Jablonski for helpful discussion. We thank A. Clauset and four anonymous reviewers for greatly improving the quality of this manuscript. We thank the many contributors to the PBDB for making data available. Funding: A.J.R. acknowledged funding from the Fulbright Program (U.S. Department of State, Bureau of Educational and Cultural Affairs), the National Science Foundation Graduate Research Fellowship Program, and the Omidyar Program at the Santa Fe Institute. M.A.F. acknowledged FONDECYT 1140278, and P.A.M. acknowledged CONICYT AFB170008, ICM-P05-002, and FONDECYT 1161023. Author contributions: A.J.R., M.A.F., and P.A.M. designed the study. A.J.R. and M.A.F. performed the analyses. A.J.R., M.A.F., and P.A.M. interpreted the results and wrote the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: Data are available through the PBDB (paleobiodb.org), and all codes needed to interface with the paleobiodb.org API, process, clean, and ultimately analyze the data are at github.com/ajrominger/paleo_supStat. This github repository also hosts the exact download from paleobiodb.org used in this analysis. All required scripts are also available and explained in Supplementary Appendix A.

Submitted 15 January 2018
Accepted 17 May 2019
Published 26 June 2019
10.1126/sciadv.aat0122

Nonequilibrium evolution of volatility in origination and extinction explains fat-tailed fluctuations in Phanerozoic biodiversity

Andrew J. Rominger, Miguel A. Fuentes and Pablo A. Marquet

Sci Adv 5 (6), eaat0122
DOI: 10.1126/sciadv.aat0122