Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/7/e1700315/DC1

# Supplementary Materials for

## Community-level regulation of temporal trends in biodiversity

Nicholas J. Gotelli, Hideyasu Shimadzu, Maria Dornelas, Brian McGill, Faye Moyes, Anne E. Magurran

Published 26 July 2017, *Sci. Adv.* **3**, e1700315 (2017)
DOI: 10.1126/sciadv.1700315

**The PDF file includes:**

- table S5. Correlations of species richness and abundance with air or seawater temperature.
- table S6. Variance ratio tests for patterns of compensatory fluctuations in total abundance.
- table S7. Null model tests for the slope of the relationship between the observed number of colonizations at time *t* and the observed number of extinctions at time $t + x$.
- Legend for table S8
- References (*53–137*)

**Other Supplementary Material for this manuscript includes the following:**
(available at advances.sciencemag.org/cgi/content/full/3/7/e1700315/DC1)

table S8 (.csv format). Primary references and metadata for 59 assemblage time series data sets.

## Materials and Methods

Data Selection

We searched the scientific literature and online biodiversity databases for publicly available time series of species abundance estimates for consistently sampled ecological assemblages (sensu (*46*)). Our criteria for inclusion of a dataset were that a) it included 10 or more years (not necessarily consecutive) of sampling, b) the sampling methods were described and relatively consistent through time, c) abundance estimates of all species in the sample were reported (i.e. assemblage data rather than population data). The 59 data sets we used are a subset of the 100 data sets originally compiled in Dornelas et al. (*35*). A full list of the datasets used in this study, their characteristics, and sources are included in the Additional Data table S1 (separate file). Datasets were checked for duplicates, for species with zero abundance and for non-organismal records, which were deleted prior to any analysis.

For access to the data sets, please contact the authors directly. Unfortunately, some of the data sets are proprietary, and we do not have permission to release all of them to third parties. However, we are currently assembling an expanded data base of community time series that will be available in the form of a published data paper.

Most biodiversity metrics are affected by sampling effort, and sampling effort was often not constant throughout the time series. To prevent variation in sampling effort from obscuring temporal biodiversity patterns, we employed sample-based rarefaction within each time series. Specifically, we used year as the temporal grain and for each time series we found the minimum number of samples in a year. We bootstrap resampled the data from each other year to obtain a constant number of samples at each sampling time. Species abundances were then pooled within each year. Some time series included years with an unusually low number of samples. To assess if this was causing an excessive loss of information, we individually assessed each time series and removed any years with less than half of the average number of samples prior to performing the sample based

rarefaction described above. This process did not affect the results of our analysis (fig. S7 in (*35*)) and hence only the first type of rarefaction is used in these analyses.

Marine Temperature Data

The Extended Reconstructed Sea Surface Temperature (ERSST) dataset is a global monthly sea surface temperature dataset derived from the International Comprehensive Ocean–Atmosphere Dataset (ICOADS). It is produced on a 2° × 2° grid with spatial completeness enhanced using statistical methods. This monthly analysis begins in January 1854 continuing to the present and includes anomalies computed with respect to a 1971–2000 monthly climatology. The newest version of ERSST, version 4 (*47*), is based on optimally tuned parameters using the latest datasets and improved analysis methods.

The monthly analysis extends from January 1854 to the present, but because of sparse data in the early years, there is damping of the analyzed signal before 1880. After 1880, the strength of the signal is more consistent over time. ERSST is suitable for long-term global and basin-wide studies, and smoothed local and short-term variations are used in the dataset.

Monthly NetCDF format gridded data are available from 1854 to present (*48*). These were imported and converted to feature layers using the ArcGIS Multidimensional Tools in order to link with central data points for analysis.

Terrestrial (and Freshwater) Temperature Data

Climate simulations from the CCSM4 are generated on a Gaussian grid, where each grid point can be uniquely accessed by one-dimensional latitude and longitude arrays (i.e. the coordinates are orthogonal). In the CCSM4 model output, the longitudes are equally spaced at 1.25º, while the latitudes vary in spacing slightly around 0.94º. Therefore, approximate spatial resolution of global climate projections is 105 km. Because of the irregular grid in the CCSM model, this portal distributes data in a point shapefile format, where each point represents a centroid of a corresponding CCSM grid cell. A shapefile of

irregular rectangular polygons of the original model output is also available. Data are available from 1850 to 2005 (*49*).

Because the data set contains both marine and terrestrial assemblage time series from both hemispheres, we used the average of the July and January temperatures for all data sets. These two months correspond to midsummer in the northern and southern hemisphere respectively, so the temperature measurements are comparable for both hemispheres. Air temperature from the CCSM4 model measured as bulk temperature of the air in degrees Kelvin and was converted to degrees Celsius for analysis.

To calculate the temperature time series associated with each dataset, we used the geographic mid-point of the study and chose the closest temperature point associated with it (see Additional Data table S1). The distance to this closest point varied among studies, but was always less than 2° in latitude and longitude. For both terrestrial and marine data, the temperature time series created for each assemblage is based on the geographic midpoint of the study location, and covers the same years that the assemblage was censused.

**Supplementary Text**

Autoregressive Time Series Models and the ADF Test: A Brief Primer

This section introduces basic AR1 time series models to explain the ADF test that was used to detect community-level regulation of species richness and abundance. In this section, $N_t$ is the species richness or total abundance of the assemblage at time *t*, *c* is a simple equilibrium point, $\varepsilon_t$ is a term for random noise at time *t* that is drawn from a normal (Gaussian) distribution with a specified mean and variance, and $\varphi$ is a parameter that controls the autocorrelation, or "memory" of the system and how much it is affected by earlier values in the time series.

**White Noise**

Consider the following models of a population "bouncing" around an "equilibrium" point *c*:

$$N_t = c + \varepsilon_t \;\; where\; \varepsilon_t \sim N(0, \sigma^2)\; and\; cov(\varepsilon_i, \varepsilon_j) = 0\; for\; i \neq j$$

Or

$$\Delta N = N_{t+1} - N_t = \varepsilon_{t+1} - \varepsilon_t \quad where \quad \varepsilon_{t+1} - \varepsilon_t \sim N(0, 2\sigma^2)$$

This is a white noise (Gaussian) process. In each time step, whether $N_t$ is greater than or less than *c* has no effect on whether ΔN is positive or negative. But the system stays strongly centered around *c*, never being more than the last time step's perturbation away from *c*. This system has no memory or autocorrelation. Each time step is completely independent of the preceding one.

It looks like this (fig. S1):

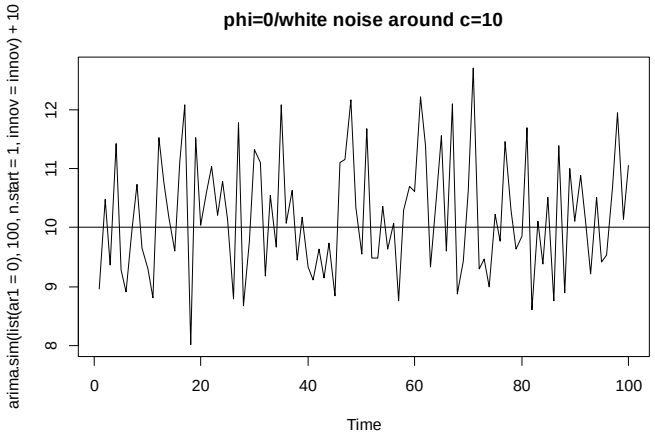



**fig. S1. Time series of uncorrelated white noise.** The odds of being above *c* for *k* times in a row are exactly the same as the odds of flipping heads on a coin *k* times in a row $(1/2)^k$. Runs of more than 3 or 4 times steps are unusual. And large swings from large positive to large negative are equally common.
But it can be centered around a trend as well (fig. S2):

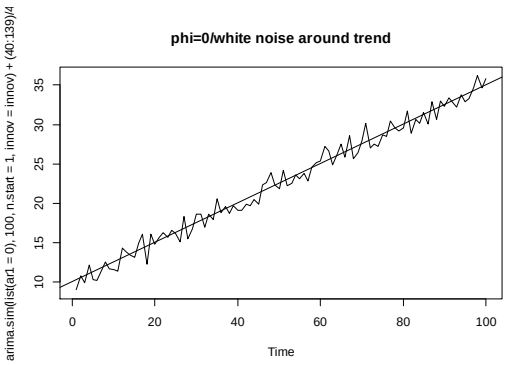



**fig. S2. Time series of uncorrelated white noise with a linear temporal trend.** And when there is a one-time major shock to the system, it disappears in the next time step (fig. S3):

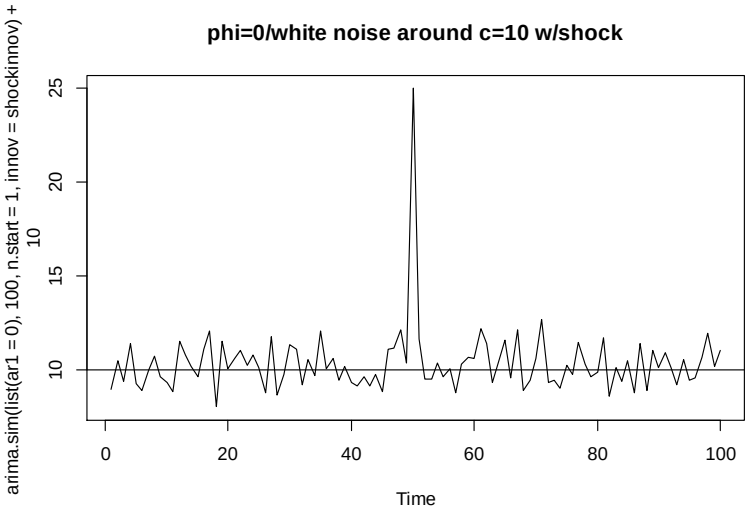



**fig. S3. Time series of uncorrelated white noise with a one-time perturbation.**

**<u>Random</u> <u>walk</u>**

Now consider a random walk starting at *c*:

$$N_0 = c, \qquad N_{t+1} = N_t + \varepsilon_t \ \ where\ \varepsilon_t \sim N(0, \sigma^2)\ and\ cov(\varepsilon_i, \varepsilon_j) = 0\ for\ i \neq j$$

Or

$$\Delta N = N_{t+1} - N_t = \varepsilon_t$$

Again, whether $N_t$ is greater or less than *c* does not predict whether $\Delta N$ is positive or negative. But the system stays only weakly centered around *c*; averaged across many series, it will be centered on *c*, but any one series can walk away from *c* for extremely

long excursions; although it will return to *c* infinitely many times in the future, it may stay away for long periods. This system has extreme memory or autocorrelation. The random walk does "return" to *c* and is "centered" around *c*, but it takes very long excursions on one side of *c* or the other. It looks like (fig. S4):

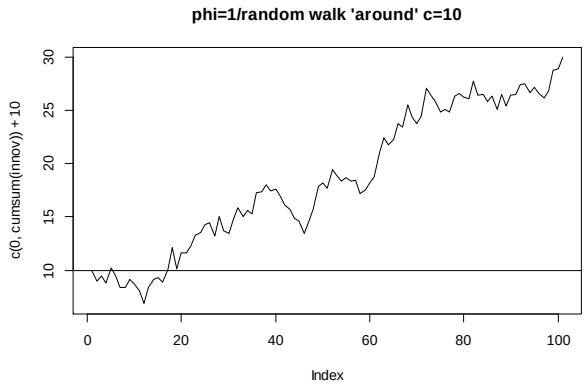



**fig. S4. Time series of random walk.** If this were run long enough it would eventually re-cross *c* = 10. The series can also be centered on a trend line (fig. S5):

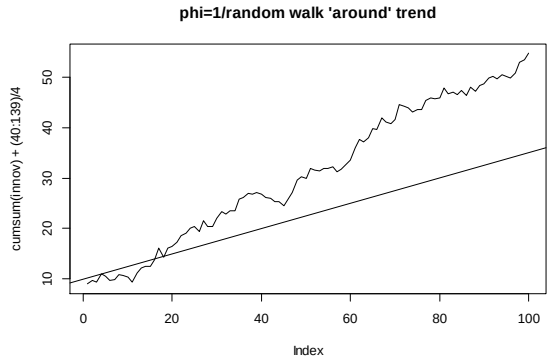



**fig. S5. Time series of a random walk with a linear temporal trend.** But if there is a one-time shock, the perturbation never disappears – a shock of +15 shifts the new *c* to *c* + 15 in the short-term (in the long-term it will eventually return to any set point that is established; fig S6):

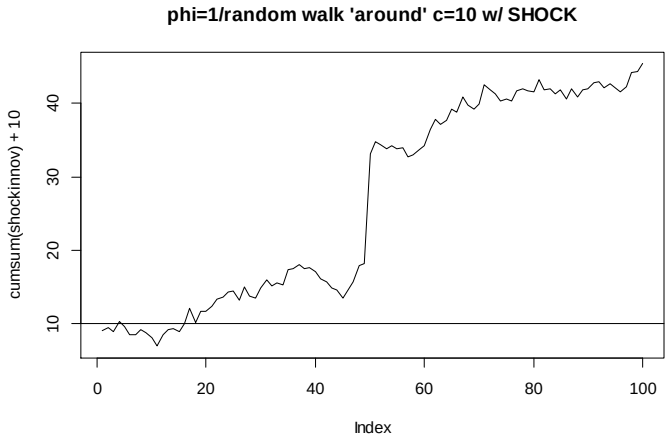



**fig. S6. Time series of a random walk with a one-time perturbation.**

**Auto-regressive Model AR1**

Now consider an auto-regressive model with a time lag of 1 (AR1 process) centered on *c*:

$$N_{t+1} = c + \varphi(N_t - c) + \varepsilon_t \; where \; \varepsilon_t \sim N(0, \sigma^2) \; and \; cov(\varepsilon_i, \varepsilon_j) = 0 \; for \; i \neq j$$

and:

$$\Delta N = N_{t+1} - N_t = -(N_t - c)(1 - \varphi) + \varepsilon_t$$

First note that when φ = 0 this is a white noise model. When φ = 1 it is a random walk model (centered on $c=N_0$). What happens when 0 < φ < 1? The system will have a moderate "memory" and a tendency to return towards *c* when it is perturbed by a shock.

Second, note the common behavior of returning. When $N_t > c$, there is a partial return by an amount $\varphi(N_t - c)$. When φ is close to 1, $N_t$ returns to *c* very quickly, and when φ is close to 0 it returns to *c* very slowly, but in all cases it has a return tendency to *c*. In other words, $\Delta N$ does depend on how far and in which direction $N_t$ differs from *c*.

If we solve recursively, we get:

$$N_t = c\left(\frac{1 - \varphi^{t+1}}{1 - \varphi}\right) + \varphi^t(N_0 - c) + \varepsilon_t + \varphi\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \varphi^3\varepsilon_{t-3} + \cdots$$

Because $|\varphi| < 1$, as $t$ increases, the effect of $N_0$ decreases and the effect of the error terms increases, with the most recent mattering most. And the effect of any past shock decays geometrically fast.

We can also see (with some work for the variance and correlation!):

$E(N_t) = c/(1-\varphi)$ when $t >> 0$

$Var(N_t)=\sigma^2/(1-\varphi^2)$

$Cor(N_t,N_{t-h})= \varphi^h$

Thus, the AR1 process is weak-sense stationary: constant mean, constant variance, autocorrelation depends only on the distance between two points in the time series, not on their position in the time series (early or late).

It looks like (fig. S7):

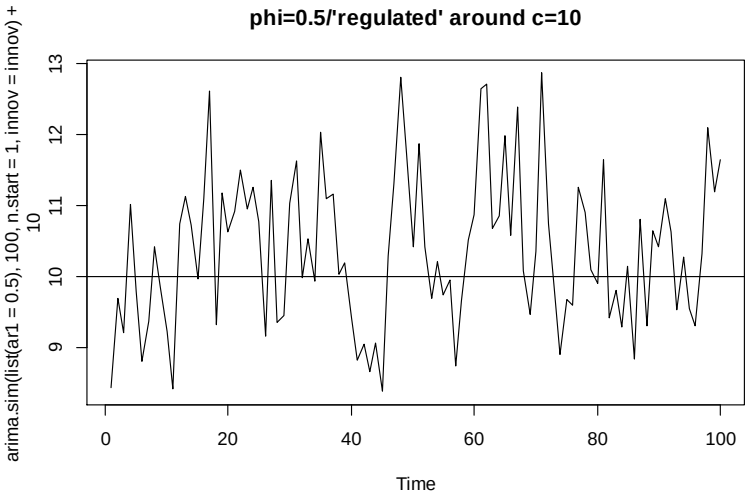



**fig. S7. Time series of a regulated autoregressive process.** Note, when the ε happen to be sampled to be positive 3 or 4 times in a row you get a bit of a deviation above, but this effect decays geometrically fast. So you can get short excursions above or below the line, but they don't last long (e.g. 5-10 time periods in the above example with $\varphi$=0.5).

But AR1 can also be centered on a trend line (fig. S8):

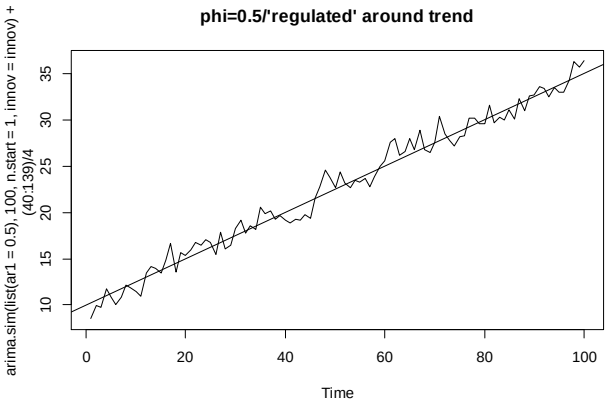



**fig. S8. Time series of a regulated autoregressive process with a linear temporal trend.** When there is a one-time shock, the effect of the perturbation persists for a while (technically it never completely disappears), but it decays back to a zero effect geometrically fast. Very quickly it reaches a point where it once again appears to be bouncing around *c*, driven by ε alone and not by the earlier shock (fig. S9):

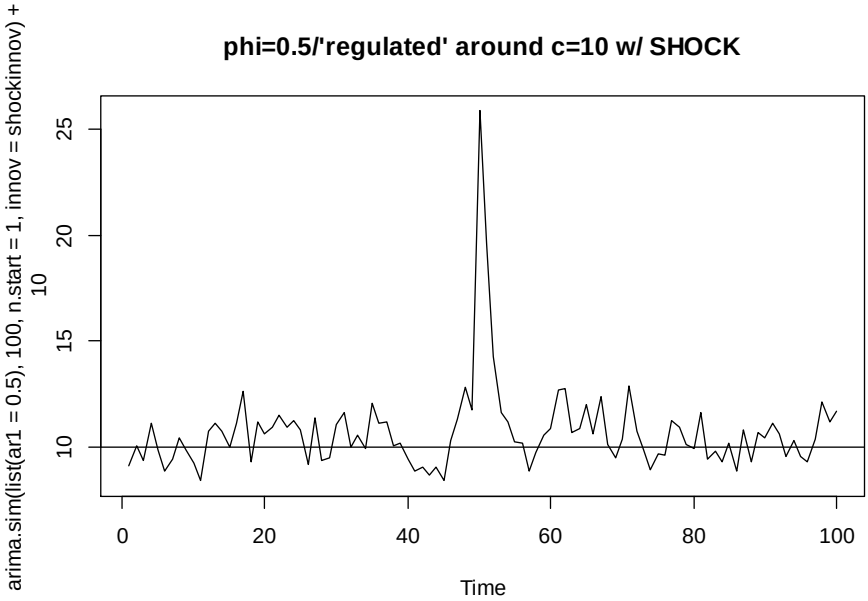



**fig. S9. Time series of a regulated autoregressive process with a one-time perturbation.** ADF Test So what is an Augmented Dickey-Fuller (ADF) test (*11*)? Nothing more or less than a one-sided statistical test for whether φ is significantly less than one ($\varphi < 1$). The null hypothesis is that the time series is best fit by a random walk ($\varphi = 1$). The alternative hypothesis is that the time series is best fit by an AR1 process with a geometrically rapid decay of perturbations ($\varphi < 1$). This process generates a time series strongly centered on *c* with relatively brief excursions above or below *c*, although the excursions last longer than those from a pure white noise process.

Although AR1 is only one of an infinite number of models that shows return tendencies to *c*, it is a simple linear model that it is a generic stand-in for more complicated models. In the same way (and with the same limitations), we typically use a linear regression model as the simplest description of the relationship between two variables when we do not have a more specific hypothesis that is being tested.

ADF Test For Regulated Time Series

Although 59 data sets were originally selected, the ADF test for species richness could not be fit successfully to 3 of them, so the tabled results are only for 56 data sets. Applied to the empirical data series, the ADF test yielded significant deviations from a random walk for 55% of the species richness time series and 54% of the total abundance time series (Table S1). These results did not differ greatly for an AR1 model that incorporated a simple linear trend, and there was still a large proportion of non-random data sets after correcting for multiple testing with the Benjamini and Hochberg (*53*) False Discovery Rate adjustment (Table S1). The results were very similar following logarithmic or square-root transformations of the data (Table S2, S3).

Random Walk Generator For Evaluation Of The ADF Test

As with many statistical analyses that have been used to detect population regulation, the ADF test may be vulnerable to false positives when there is observation error present. Following the lead of Shenk et al. (*54*) and Freckleton et al. (*55*), we explored the potential effects of observation error with simulated data sets. We generated unconstrained random walks with explicit terms for both process error and observation error and subjected them to the ADF test. The random walk model represents the null hypothesis for the ADF test, so these simulations reveal the tendency for Type I errors in the ADF test when observation error is present. The basic ADF test assumes there are no deterministic long-term trends in the data. A variant of the ADF test incorporates simple linear trends. For these data sets, the grand average showed no upward or downward trend, but there were particular data sets in which temporal trends could be detected (*35*). Therefore, we repeated the analyses with this variant of the ADF test to ensure that the results were consistent with or without correction for temporal trends.

The random walk model we simulated for a continuous response variable *N* was:

$$N_{t+1} = N_t + r_t$$

where $r_t \sim N(0, \sigma^2)$.

To incorporate observation error, we first created the random walk series and then contaminated it by adding a random normal variate *z*, also with a mean of zero and a specified variance $z_t \sim N(0, \sigma^2)$

$$N_{t_{\text{observed}}} = N_t + z_t$$

Note that the observation error is simply added to the time series, so the size of the error is independent of the size of the observation. The units (standard deviations of a normal distribution) are the same for process and observation error, which makes it straightforward to compare the relative strength of process versus observation error.

In addition to the effects of process and observation error, the statistical power of the ADF test is also affected by the length of each time series (= sample size). In this case, capturing the variation due to sample size was important, because some of the time series in our analyses had as few as 10 or 15 data points. Therefore, we simulated random walks that had the same number of observations as each of the empirical series, systematically varied the strength of observation error and process noise, tested each series with the ADF test, created box plots for each set of parameter combinations, and compared them visually to the box plot of effect sizes for the empirical data.

Benchmark Performance of the ADF Test

We compared the effects of differing levels of observation error ($S_{noise}$ = σ = 0,10,20,30,40,50,60,70,80,90) crossed with differing levels of process error ($P_{noise}$= σ = 10,20,30,40,50,60). For each of the 10 × 6 = 60 treatment combinations, we generated a single data set with 59 time series. The length of the individual time series matched the length of the time series in the observed data. We next calculated the ADF test for each stochastic series, just as we did for the empirical data. The *P*-value from each test was converted to a standardized effect size with the `qnorm()` function in R. We then created a box plot of standardized effect sizes for each ensemble. We visually compared this set of 60 box plots to the two box plots of standardized effect sizes for the species richness and abundance of the empirical data. We also repeated the random walk analyses, with a square root or logarithmic transformation first applied to the empirical and simulated data series. Figure S2 shows the results for the random walk tests, and fig. S12 and fig. S13 show the results for the logarithmic and square root transformations.

In all of these random walks, statistically significant results are more likely to occur with high levels of measurement error, and less likely to occur with high levels of process error. If the ratio of observation error to process error is less than or equal to 1.0, the stochastic time series exhibit *P*-values (and standardized effect sizes) that are distinctly different from the two empirical box plots for species richness and total abundance. When the process error is very strong (σ =50 or σ =60), the ratio of process to

observation error can be greater than 1.0 without generating results that match the empirical data. For the parameter space we analyzed, 30 of the 60 simulation scenarios generated box plots that were distinctly different from the distribution of the observed data. By "distinctly different", we mean that 75% of the SES values for the simulated time series (= lower boundary of the interquartile box in a boxplot) were greater than -2.0, which is the approximate median of the SES values for the empirical time series of species richness and relative abundance (blue and green box plots in figs. S11-S15).

Process and Observation Error in Assemblage-Level Time Series

These simulation exercises show that the ADF test is not overly sensitive to observation error. With relatively strong process error, the ADF test will not incorrectly reject a data set that is based on a random walk. But how much measurement and process error is there in the real data? We can't answer this question without replicated observations at each time point (*56*). But we can comment on the results of the ADF test applied to these particular data.

Although many of these assemblages have been monitored for decades, 18 of the time series (31%) are only 10 to 15 years long. Small sample size weakens the power of the ADF test, and makes it more likely that a regulated time series will be mistaken for a random walk. The data series analyzed here span a time period from 1874 to 2010, a period of unprecedented increase in the size of the human population (*26*), the velocity of climate change(*27*), and the pace of accidental and deliberate biotic exchange (*28*). These forces should generate a very strong signal of process error for a stochastic random walk, which makes it harder to incorrectly reject the null hypothesis with the ADF test. Finally, we suspect that observation error for the summed abundance of an entire assemblage (which itself is comprised of pooled samples for many of the data series) is likely to be smaller than the observation error measured for any one of its populations, although that is a hypothesis that remains to be tested.

In sum, all statistical tests for regulated time series are sensitive to sample size and the relative strength of observation and process error. There is always a "Goldilocks" configuration of parameter settings for a random walk (large sample size, weak process error, and strong measurement error) that could generate false positives with the ADF tests. But for this diverse set of monitored assemblages, we argue that the results in Fig. 1 more likely reflect the signature of community-wide regulation of species richness and abundance than the effects of observation error.

Portfolio Effects and Other Sample Size Artifacts

In addition to large observation error, a second potential statistical artifact is the portfolio effect (*52*), in which the total variance of a community variable such as abundance is reduced if it is comprised of a large number of component species that are fluctuating independently of one another. However, the strength of statistical significance of the ADF test (as measured by the *z*-score of the probability value) was uncorrelated with the number of species in the assemblage (species richness ADF: $r = 0.03$, $P = 0.80$; total abundance ADF: $r = -0.08$, $P = 0.53$). This result confirms that the ADF test is not simply reflecting the coefficient of variance in $S$ or total $N$ but instead is measuring the constancy of fluctuations through time. Other sampling artifacts also seem unlikely because the strength of the ADF test was uncorrelated with the total abundance in the assemblage (species richness ADF: $r = 0.08$, $P = 0.54$; total abundance ADF: $r = 0.17$, $P = 0.18$), or the number of observations in the time series (species richness ADF: $r = 0.02$, $P = 0.86$; total abundance ADF: $r = -0.05$, $P = 0.72$).

Environmental Tracking

Environmental conditions that vary with time could cause patterns of community regulation, either because they reflect changes in limiting resources (such as energy) or because species in an assemblage show similar responses to changes in abiotic factors, such as temperature or salinity (*25, 57*). If this kind of environmental tracking occurs and the environmental variable itself shows a pattern of regulated fluctuations, community-level regulation would result (*58*).

Although it is not possible to test the role of all environmental drivers, we examined the importance of environmental tracking by testing for correlations with environmental temperature. In both terrestrial and aquatic habitats, temperature has major direct impacts on metabolism, individual growth, life history, and population dynamics (*59*) and is strongly correlated with many other important environmental drivers (e.g. precipitation on land, salinity in the ocean).

To explore the importance of temperature as an environmental driver that is responsible for community-level regulation, we created time series of annual air or sea surface temperatures for each assemblage. Each temperature time series was tailored to the temporal coverage of the assemblage time series and the geographic midpoint of the sampled sites (Materials and Methods). By the ADF test, 97% of these temperature time series (57 of 59 studies) were in in the direction of stationarity (negative *z*-scores for the ADF test), and 64% (38 of 59 studies) were strongly stationary ($P < 0.05$; Table S4).

In spite of the strong stationarity of most temperature time series, there were relatively few assemblages in which fluctuations in $S$ or $N$ were statistically correlated with air or water temperature ($P < 0.05$, two-tailed test; ln($S$) versus temperature: 12 of 59

assemblages = 20%; ln(*N*) versus temperature: 13 of 59 assemblages = 22%; Table S5). Thus, although air and water temperature time series show a stationary pattern, there is little evidence that direct or indirect tracking of temperature is responsible for community-level regulation of species richness or total abundance.

Compensatory Fluctuations

Although there is widespread evidence for community-level regulation in these assemblages (Fig. 1), there was little evidence that the pattern was caused by observation error, portfolio effects, or environmental tracking of air or water temperature. Compensatory fluctuations in abundance or species occurrence that reflect a community-level energy constraint are an important mechanism that could account for these results.

We tested for compensatory dynamics in abundances of populations and for compensatory dynamics in species richness via colonization and extinction rates. To test for compensatory fluctuations in *N*, we used the variance ratio test (*23, 24*). In a regulated assemblage, the variance ratio (variance of the sum of abundances to the sum of the variances of abundances) should be significantly < 1.0 when the average pairwise covariation in abundances is negative. To test for compensatory fluctuations in *S*, we calculated, for each assemblage, the slope of the relationship between the number of colonization events at time ($t$) and the number of extinction events at time ($t + x$), where $x$ is a specified time lag. In a regulated assemblage, the observed slope should be significantly larger than expected by chance. Steep positive slopes would occur when a large number of colonizations in one time period ($t$) is compensated for by a large number of extinctions in a future time period ($t + x$).

For both of these analyses, we used a null model randomization test in which the within-species autocorrelation in occurrence or abundance is preserved through a cyclic-shift permutation (*23*). For the variance ratio test, only 2 of 59 assemblages showed compensatory fluctuations in *N* (observed variance ratio smaller than expected by chance; Table S6). For the slope test, only 2 to 6 (depending on the time lag) of 56 assemblages showed compensatory fluctuations in *S* (observed slope of the relationship between colonization and extinction larger than expected by chance; Table S7).

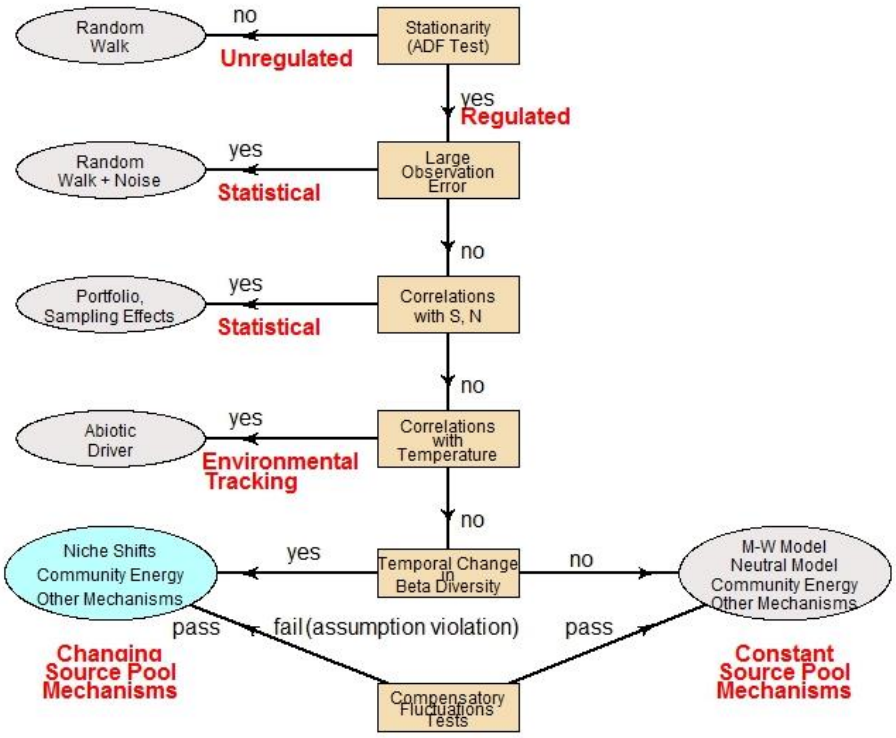
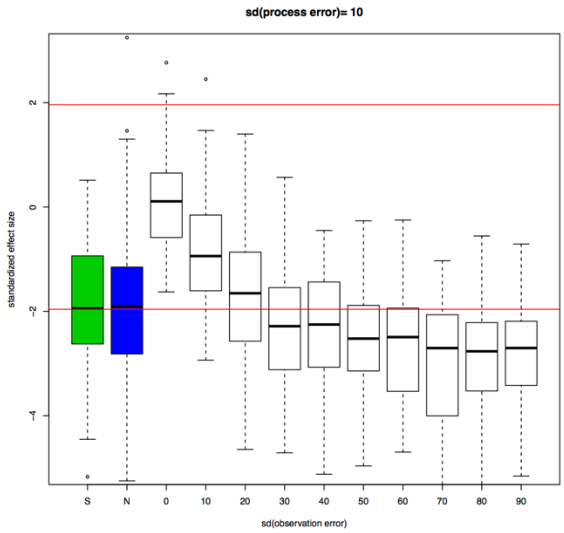
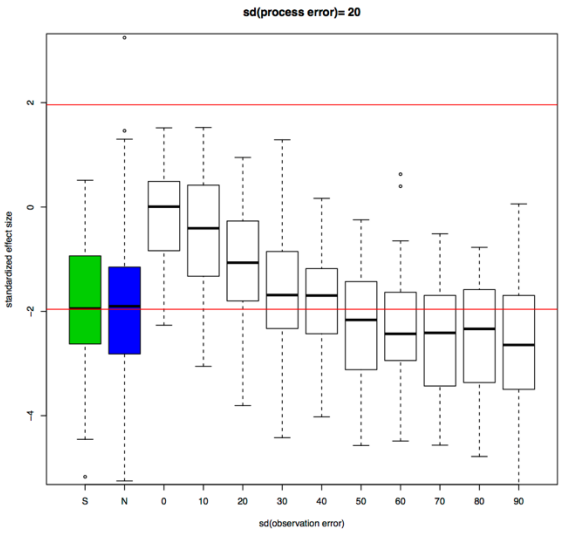
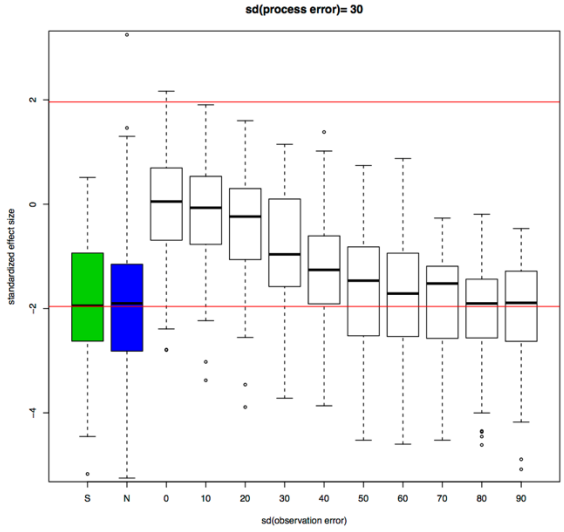
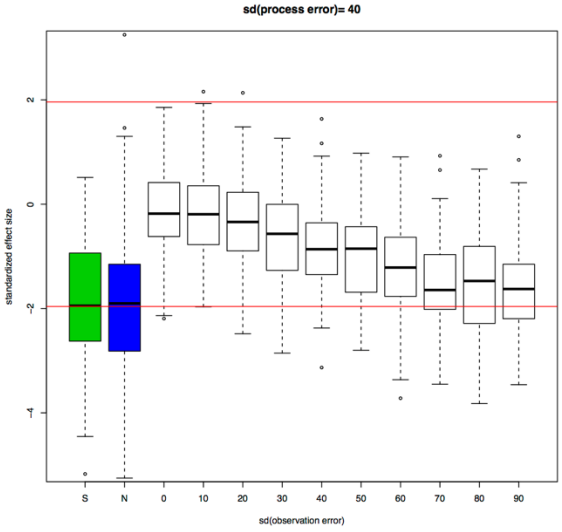


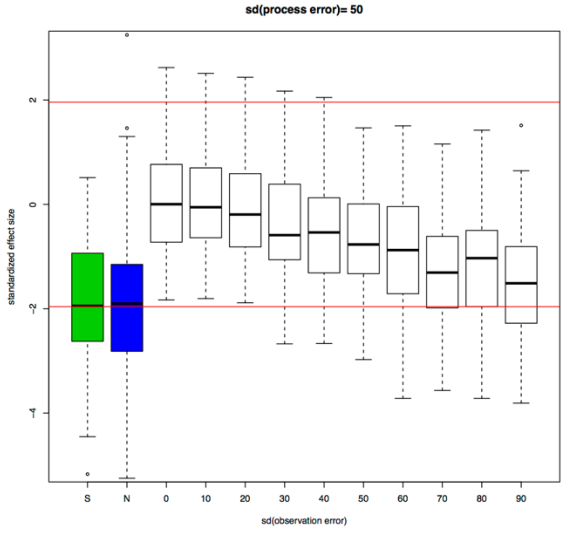

**fig. S10. Logic tree for analysis and interpretation of community time series.** Logic tree for analysis and interpretation of community time series. Each box represents a different statistical test or model analysis, and the arrows represent the different outcomes. Each oval represents the different models or interpretations of the result, and the blue oval represents the outcome for these analyses. Red labels indicate the different broad categories of regulatory mechanisms.


sd(process error)= 10
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90
2
0
−2
−4


sd(process error)= 20
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90
2
0
−2
−4


sd(process error)= 30
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90
2
0
−2
−4


sd(process error)= 40
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90
2
0
−2
−4


sd(process error)= 50
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90
2
0
−2
−4

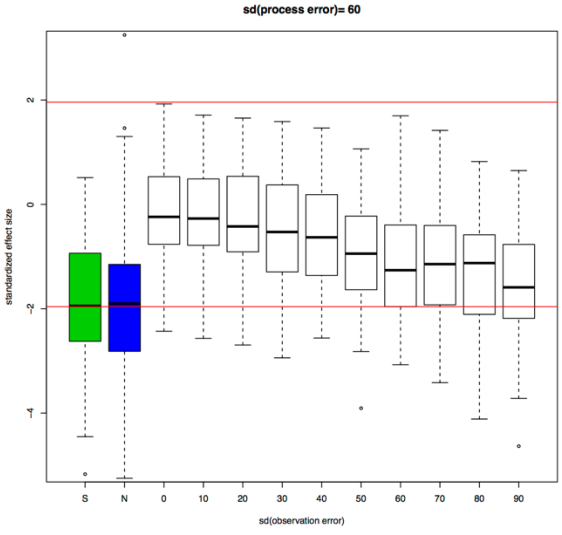
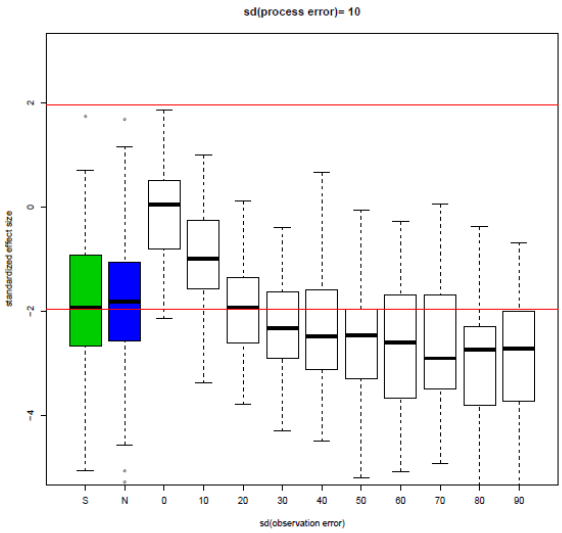
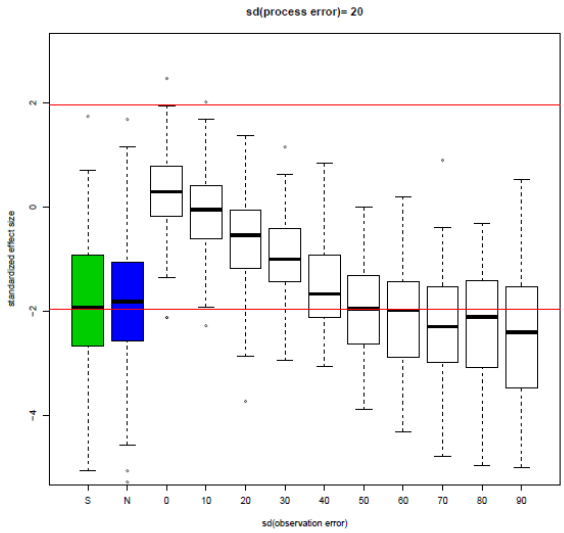
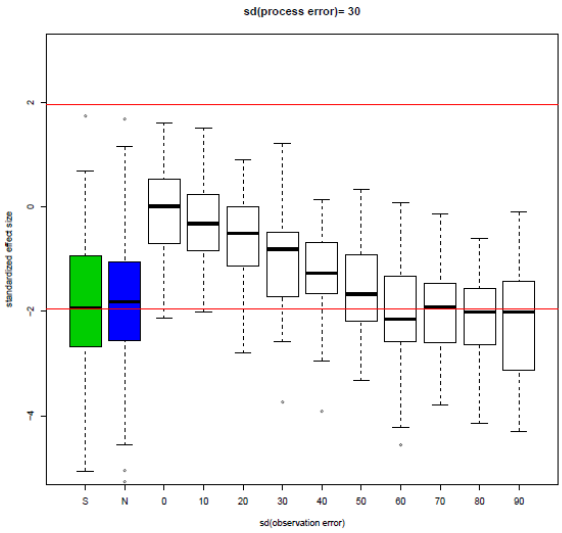
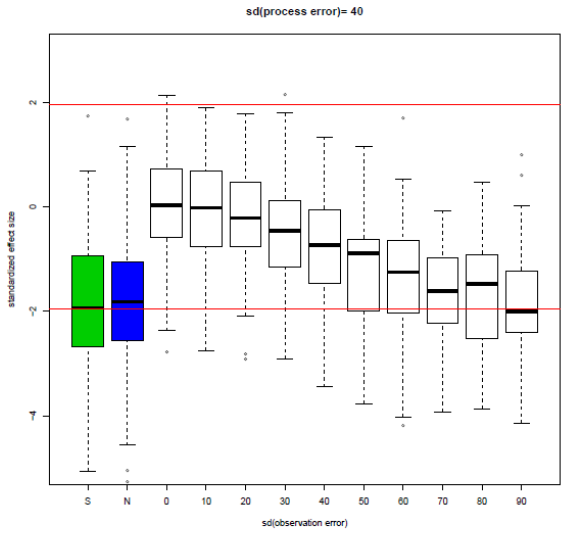

sd(process error)= 60
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90
2
0
−2
−4

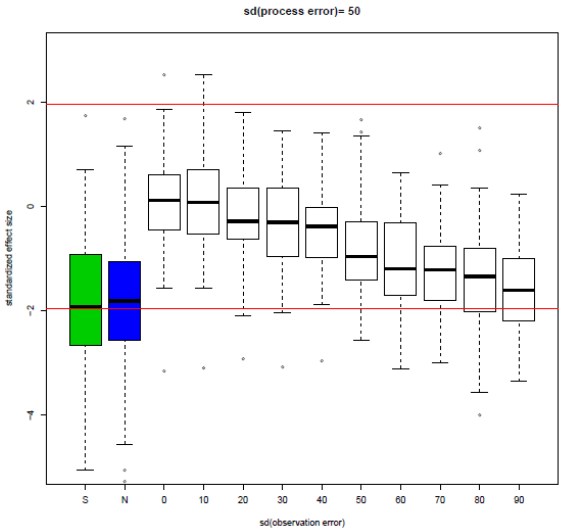
**fig. S11. Benchmark analysis of ADF test.** Benchmark analysis of ADF test. The *Y*-axis is the standardized effect size, which is the *P* value from the ADF test converted to a standardized effect size with the R function qnorm(). The horizontal red lines indicate the approximate significance boundaries of *P* values, with $P < 0.05$ (two-tailed test) corresponding to an SES of ~ -1.96. Each box plot summarizes the results of 59 ADF tests, applied to time series whose length matches those of the empirical data. The shaded box plots show the distribution of SES values for the empirical data for species richness (green) and total abundance (blue; see also Fig. 1). Each unshaded box plot shows the result of an analysis on a comparable set of box plots in which the ADF test was applied to random-walk time series.


sd(process error)= 10
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90


sd(process error)= 20
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90


sd(process error)= 30
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90


sd(process error)= 40
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90


sd(process error)= 50
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90

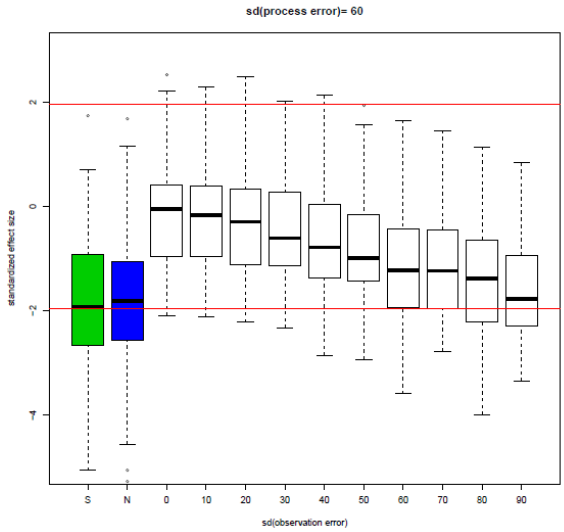
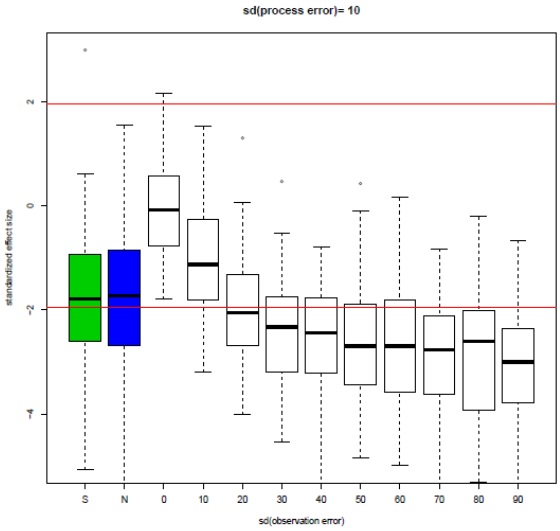
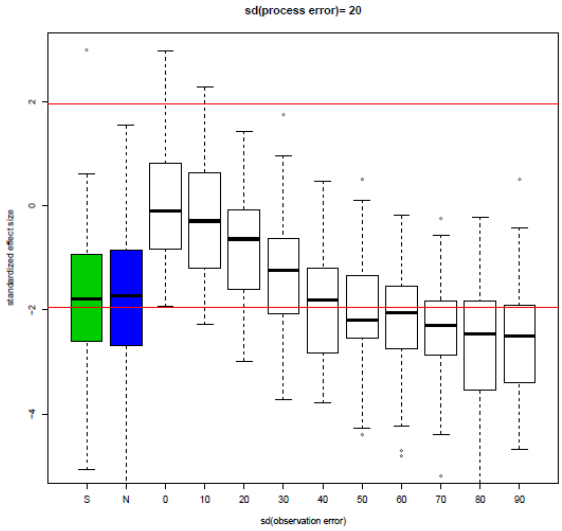
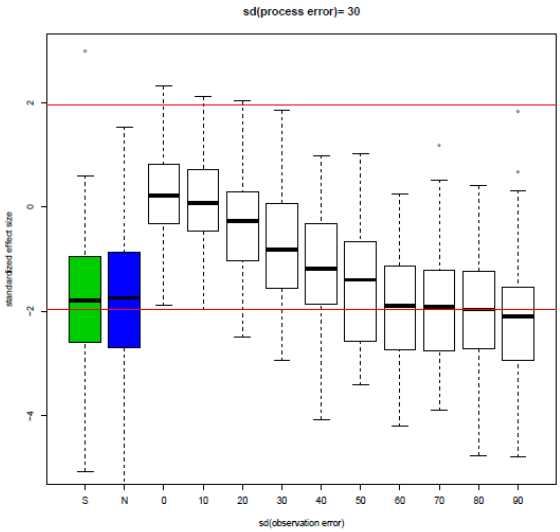
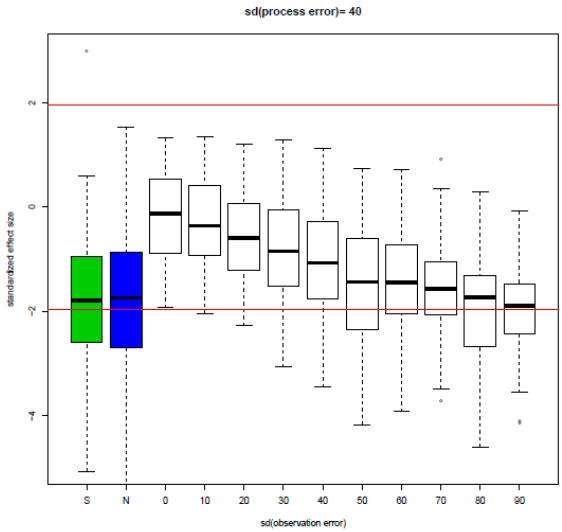

sd(process error)= 60
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90

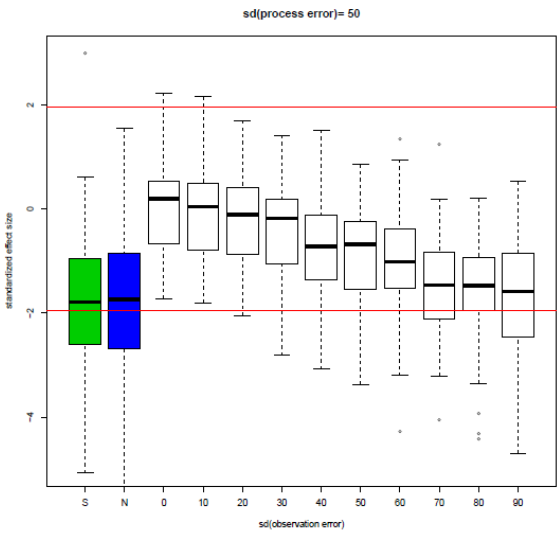
**fig. S12. Benchmark analysis of ADF test.** Benchmark analysis of ADF test. The *Y*-axis is the standardized effect size, which is the *P* value from the ADF test converted to a standardized effect size with the R function qnorm(). The horizontal red lines indicate the approximate significance boundaries of *P* values, with $P < 0.05$ (two-tailed test) corresponding to an SES of ~ -1.96. Each box plot summarizes the results of 59 ADF tests, applied to time series whose length matches those of the empirical data. The shaded box plots show the distribution of SES values for the empirical data for species richness (green) and total abundance (blue; see also Fig. 1). Each unshaded box plot shows the result of an analysis on a comparable set of box plots in which the ADF test was applied to random-walk time series. Empirical and simulated data were $\log_{10}$ transformed before analysis.


sd(process error)= 10
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90


sd(process error)= 20
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90


sd(process error)= 30
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90


sd(process error)= 40
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90


sd(process error)= 50
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90

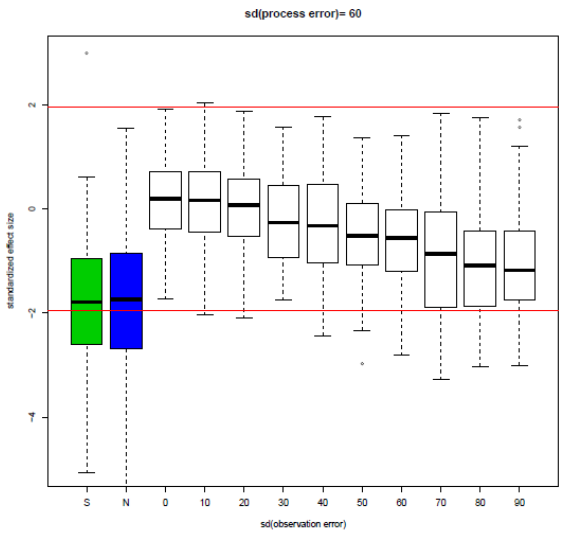

sd(process error)= 60
standardized effect size
sd(observation error)
S
N
0
10
20
30
40
50
60
70
80
90

**fig. S13. Benchmark analysis of ADF test.** Benchmark analysis of ADF test. The *Y*-axis is the standardized effect size, which is the *P* value from the ADF test converted to a standardized effect size with the R function qnorm(). The horizontal red lines indicate the approximate significance boundaries of *P* values, with $P < 0.05$ (two-tailed test) corresponding to an SES of ~ -1.96. Each box plot summarizes the results of 59 ADF tests, applied to time series whose length matches those of the empirical data. The shaded box plots show the distribution of SES values for the empirical data for species richness (green) and total abundance (blue; see also Fig. 1). Each unshaded box plot shows the result of an analysis on a comparable set of box plots in which the ADF test was applied to random-walk time series. Empirical and simulated data were square-root transformed before analysis.

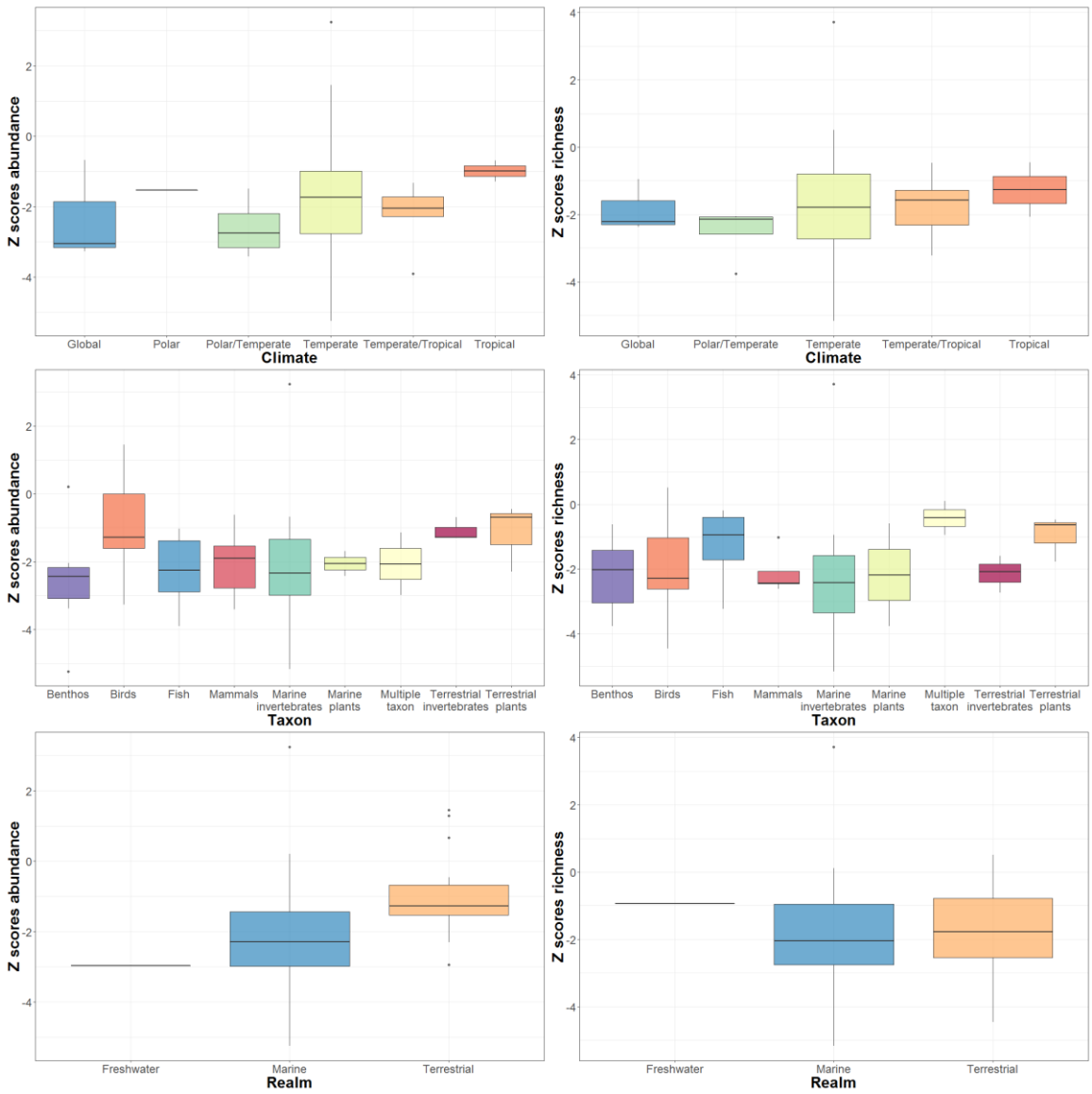



**fig. S14. Statistical tests for effects of latitudinal band (=climate), taxonomic group, and realm on standardized effect sizes (*z* scores) of species richness and total abundance.** Statistical tests for effects of latitudinal band (= climate), taxonomic group, and realm on standardized effect sizes (Z-scores) of species richness and total abundance. Each box plot summarizes the Z-scores for a subset of communities. Groups in each panel were compared by a one-way analysis of variance: Total Abundance x Climate ($P$ = 0.76; model $r^2$ = 0.05), Species Richness x Climate ($P$ = 0.88; model $r^2$ = 0.02), Total Abundance x Taxon ($P$ = 0.40; model $r^2$ = 0.15), Species Richness x Taxon ($P$ = 0.52; model $r^2$ = 0.13), Total Abundance x Realm ($P$ = 0.008; model $r^2$ = 0.16), Species Richness x Realm ($P$ = 0.80; model $r^2$ = 0.01).

**table S1. Number of significant ($P < 0.05$) and nonsignificant test results for assemblage-level regulation of species richness or abundance.**

Number of significant ($P < 0.05$ and non-significant (NS) test results for assemblage-level regulation of species Richness or abundance. The ADF0 test assumes no deterministic trends in the data, whereas the ADF1 test accounts for linear trends. Tests are repeated with the Benjamini and Hochberg (*53*) adjustment for the False Discovery Rate (FDR). Counts of species and total abundance were untransformed.

| | ADF0 (p < 0.05) | ADF0 (NS) | ADF1 (p < 0.05) | ADF1 (NS) |
|---|---|---|---|---|
| Species Richness | 31 | 25 | 29 | 27 |
| Abundance | 32 | 27 | 34 | 25 |
| Species Richness (FDR) | 28 | 28 | 17 | 39 |
| Abundance (FDR) | 29 | 30 | 24 | 35 |

**table S2. Number of significant ($P < 0.05$) and nonsignificant test results for assemblage-level regulation of species richness or abundance.**

Number of significant ($P < 0.05$ and non-significant (NS) test results for assemblage-level regulation of species richness or abundance. The ADF0 test assumes no deterministic trends in the data, whereas the ADF1 test accounts for linear trends. Tests are repeated with the Benjamini and Hochberg (*53*) adjustment for the False Discovery Rate (FDR). Counts of species and total abundance were $\log_{10}$ transformed before analysis.

| | ADF0 (p < 0.05) | ADF0 (NS) | ADF1 (p < 0.05) | ADF1 (NS) |
|---|---|---|---|---|
| Species Richness | 32 | 24 | 32 | 24 |
| Abundance | 35 | 24 | 31 | 28 |
| Species Richness (FDR) | 28 | 28 | 23 | 33 |
| Abundance (FDR) | 24 | 35 | 25 | 34 |

**table S3. Number of significant ($P < 0.05$) and nonsignificant test results for assemblage-level regulation of species richness or abundance.**

Number of significant ($P < 0.05$ and non-significant (NS) test results for assemblage-level regulation of species richness or abundance. The ADF0 test assumes no deterministic trends in the data, whereas the ADF1 test accounts for linear trends. Tests are repeated with the Benjamini and Hochberg (*53*) adjustment for the False Discovery Rate (FDR). Counts of species and total abundance were square-root transformed before analysis.

| | ADF0 (p < 0.05) | ADF0 (NS) | ADF1 (p < 0.05) | ADF1 (NS) |
|---|---|---|---|---|
| Species Richness | 33 | 23 | 31 | 25 |
| Abundance | 31 | 28 | 31 | 28 |
| Species Richness (FDR) | 26 | 30 | 18 | 38 |
| Abundance (FDR) | 28 | 31 | 23 | 36 |

**table S4. Results of ADF tests for temperature time series.**

Results of ADF tests for temperature time series. Each row represents a different assemblage. The first column is the habitat (terrestrial or marine) and the second column is the ID number of the assemblage (see Additional Data Table S1). The third column is the slope of the relationship between temperature (air temperature for terrestrial assemblages, sea water temperature for marine assemblages) and time (units are C°/year). The fourth column is the *P*-value for the relationship (linear regression; 2-tailed test). The fifth column is the *P*-value for the ADF test for stationarity of the temperature time series. The sixth column is the *P*-value for the ADF test for stationarity of the temperature time series accounting for linear trends in temperature. Orange shading indicates $P < 0.05$. For the slope tests, blue shading indicates $P < 0.05$ and a negative slope for temperature versus time.

| Realm | studyID | slopeMeanTe | pvalueMean | meanTempA | meanTempA |
|---|---|---|---|---|---|
| Terrestrial | 18.00 | 5.19 | 0.15 | 0.00 | 0.00 |
| Marine | 33.00 | 8.11 | 0.01 | 0.04 | 0.00 |
| Terrestrial | 39.00 | 5.85 | 0.01 | 0.10 | 0.17 |
| Terrestrial | 42.00 | 4.27 | 0.43 | 0.01 | 0.05 |
| Terrestrial | 44.00 | 1.99 | 0.35 | 0.00 | 0.01 |
| Terrestrial | 46.00 | 12.86 | 0.12 | 0.00 | 0.00 |
| Terrestrial | 47.00 | 0.27 | 0.79 | 0.00 | 0.00 |
| Terrestrial | 51.00 | 0.70 | 0.59 | 0.00 | 0.01 |
| Terrestrial | 52.00 | 0.33 | 0.80 | 0.00 | 0.00 |
| Terrestrial | 54.00 | 18.08 | 0.00 | 0.53 | 0.20 |
| Terrestrial | 56.00 | 1.06 | 0.49 | 0.03 | 0.11 |
| Freshwater | 57.00 | 4.52 | 0.01 | 0.00 | 0.00 |
| Terrestrial | 58.00 | 16.73 | 0.00 | 0.53 | 0.13 |
| Terrestrial | 59.00 | 1.71 | 0.30 | 0.01 | 0.02 |
| Terrestrial | 63.00 | 10.72 | 0.02 | 0.04 | 0.05 |
| Terrestrial | 67.00 | 8.75 | 0.07 | 0.04 | 0.05 |
| Marine | 68.00 | 6.54 | 0.08 | 0.36 | 0.31 |
| Terrestrial | 70.00 | -0.60 | 0.84 | 0.05 | 0.16 |
| Marine | 78.00 | 4.29 | 0.03 | 0.00 | 0.00 |
| Marine | 81.00 | 8.46 | 0.01 | 0.25 | 0.27 |
| Marine | 84.00 | 16.99 | 0.17 | 0.20 | 0.05 |
| Marine | 85.00 | 1.51 | 0.39 | 0.04 | 0.07 |
| Marine | 92.00 | 15.72 | 0.00 | 0.30 | 0.04 |
| Marine | 97.00 | 20.71 | 0.00 | 0.04 | 0.00 |
| Marine | 100.00 | 9.70 | 0.00 | 0.01 | 0.01 |
| Marine | 101.00 | 9.70 | 0.00 | 0.01 | 0.01 |
| Marine | 108.00 | -28.74 | 0.06 | 0.03 | 0.15 |
| Marine | 110.00 | 15.78 | 0.01 | 0.01 | 0.01 |
| Marine | 112.00 | 25.28 | 0.00 | 0.00 | 0.00 |
| Marine | 113.00 | 8.27 | 0.07 | 0.01 | 0.02 |
| Marine | 117.00 | 4.61 | 0.35 | 0.00 | 0.00 |
| Marine | 119.00 | 12.65 | 0.00 | 0.00 | 0.00 |
| Marine | 125.00 | 1.00 | 0.77 | 0.10 | 0.31 |
| Marine | 135.00 | -4.07 | 0.43 | 0.18 | 0.37 |
| Marine | 147.00 | 17.89 | 0.01 | 0.00 | 0.00 |
| Marine | 148.00 | 6.32 | 0.14 | 0.01 | 0.03 |
| Marine | 152.00 | 4.38 | 0.32 | 0.14 | 0.38 |
| Marine | 162.00 | 1.06 | 0.69 | 0.02 | 0.08 |
| Marine | 163.00 | 5.16 | 0.02 | 0.43 | 0.44 |
| Marine | 166.00 | 7.09 | 0.01 | 0.01 | 0.01 |
| Marine | 169.00 | 0.87 | 0.74 | 0.01 | 0.04 |
| Marine | 171.00 | 4.16 | 0.39 | 0.01 | 0.03 |
| Marine | 172.00 | 0.63 | 0.87 | 0.02 | 0.09 |
| Marine | 173.00 | 1.51 | 0.62 | 0.07 | 0.15 |
| Marine | 176.00 | 3.30 | 0.26 | 0.11 | 0.24 |
| Marine | 178.00 | 10.11 | 0.01 | 0.01 | 0.00 |
| Marine | 180.00 | -8.97 | 0.01 | 0.06 | 0.08 |
| Marine | 182.00 | 6.79 | 0.01 | 0.08 | 0.04 |
| Marine | 183.00 | 1.42 | 0.69 | 0.09 | 0.23 |
| Marine | 187.00 | 2.91 | 0.58 | 0.13 | 0.09 |
| Marine | 191.00 | 8.56 | 0.32 | 0.02 | 0.04 |
| Marine | 192.00 | 93.00 | 0.00 | 0.11 | 0.22 |
| Marine | 193.00 | 8.09 | 0.03 | 0.01 | 0.01 |
| Terrestrial | 195.00 | 2.87 | 0.05 | 0.01 | 0.01 |
| Marine | 196.00 | 14.29 | 0.00 | 0.08 | 0.01 |
| Marine | 197.00 | 15.58 | 0.00 | 0.15 | 0.05 |
| Marine | 198.00 | 2.67 | 0.27 | 0.02 | 0.05 |
| Marine | 200.00 | 8.99 | 0.03 | 0.12 | 0.27 |
| Terrestrial | 277.00 | 0.12 | 0.98 | 0.02 | 0.13 |

**table S5. Correlations of species richness and abundance with air or seawater temperature.**

Correlations of species richness and abundance with air or sea water temperature. Each row represents a different assemblage. The first column is the habitat (terrestrial or marine) and the second column is the ID number of the assemblage (see Additional Data Table S1). The third column is the slope of the relationship between ln(abundance) and temperature (air temperature for terrestrial assemblages, sea water temperature for marine assemblages; units are ln(N)/C°). The fourth column is the *P*-value for the relationship (linear regression; 2-tailed test). The fifth column is the slope of the relationship between ln(species richness) and temperature. The sixth column is the *P*-value for the relationship. Orange shading indicates $P < 0.05$ with a positive slope, and blue indicates $P < 0.05$ with a negative slope.

| Realm | studyID | slopelogNxMeanTemp | pvaluelogNxMeanTemp | slopelogSxMeanTemp | pvaluelogSxMeanTemp |
|---|---|---|---|---|---|
| Terrestrial | 18 | -0.08 | 0.15 | -0.02 | 0.41 |
| Marine | 33 | -0.07 | 0.65 | -0.03 | 0.53 |
| Terrestrial | 39 | -0.02 | 0.69 | -0.06 | 0.03 |
| Terrestrial | 42 | 0.03 | 0.59 | 0.01 | 0.67 |
| Terrestrial | 44 | 0.14 | 0.03 | 0.13 | 0.02 |
| Terrestrial | 46 | 0.02 | 0.78 | -0.04 | 0.40 |
| Terrestrial | 47 | -0.03 | 0.47 | -0.01 | 0.23 |
| Terrestrial | 51 | 0.00 | 0.91 | 0.00 | 0.77 |
| Terrestrial | 52 | 0.04 | 0.56 | 0.00 | 0.02 |
| Terrestrial | 54 | -0.28 | 0.10 | 0.08 | 0.18 |
| Terrestrial | 56 | 0.02 | 0.12 | 0.02 | 0.65 |
| Freshwater | 57 | 0.05 | 0.43 | 0.01 | 0.53 |
| Terrestrial | 58 | -0.39 | 0.19 | -0.32 | 0.04 |
| Terrestrial | 59 | 0.02 | 0.66 | -0.03 | 0.20 |
| Terrestrial | 63 | 0.24 | 0.15 | 0.05 | 0.62 |
| Terrestrial | 67 | -0.05 | 0.43 | 0.03 | 0.29 |
| Marine | 68 | 0.05 | 0.17 | 0.00 | 0.96 |
| Terrestrial | 70 | -0.11 | 0.80 | -0.02 | 0.79 |
| Marine | 78 | 0.06 | 0.61 | 0.07 | 0.38 |
| Marine | 81 | -0.42 | 0.02 | 0.15 | 0.25 |
| Marine | 84 | -0.19 | 0.55 | 0.23 | 0.02 |
| Marine | 85 | -0.09 | 0.64 | -0.11 | 0.31 |
| Marine | 92 | 0.41 | 0.00 | 0.23 | 0.09 |
| Marine | 97 | -0.71 | 0.21 | 0.02 | 0.84 |
| Marine | 100 | 0.07 | 0.53 | 0.05 | 0.04 |
| Marine | 101 | 0.25 | 0.01 | 0.11 | 0.00 |
| Marine | 108 | -1.07 | 0.07 | -0.82 | 0.01 |
| Marine | 110 | -0.27 | 0.13 | -0.27 | 0.09 |
| Marine | 112 | 0.16 | 0.19 | -0.04 | 0.42 |
| Marine | 113 | -0.76 | 0.00 | -0.71 | 0.00 |
| Marine | 117 | 0.13 | 0.06 | 0.05 | 0.23 |
| Marine | 119 | 0.13 | 0.01 | 0.06 | 0.03 |
| Marine | 125 | 0.08 | 0.55 | 0.03 | 0.70 |
| Marine | 135 | 0.28 | 0.67 | 0.19 | 0.59 |
| Marine | 147 | 0.04 | 0.80 | 0.04 | 0.80 |
| Marine | 148 | -0.19 | 0.21 | -0.19 | 0.21 |
| Marine | 152 | 0.42 | 0.18 | 0.25 | 0.17 |
| Marine | 162 | -0.02 | 0.79 | -0.04 | 0.60 |
| Marine | 163 | 0.06 | 0.04 | 0.00 | 0.78 |
| Marine | 166 | -0.13 | 0.21 | 0.00 | 0.90 |
| Marine | 169 | 0.09 | 0.21 | 0.00 | 0.86 |
| Marine | 171 | -0.26 | 0.43 | -0.06 | 0.50 |
| Marine | 172 | 0.01 | 0.96 | 0.03 | 0.61 |
| Marine | 173 | -0.07 | 0.67 | -0.01 | 0.96 |
| Marine | 176 | 0.04 | 0.96 | 0.02 | 0.36 |
| Marine | 178 | 0.13 | 0.03 | 0.17 | 0.01 |
| Marine | 180 | -0.02 | 0.75 | 0.03 | 0.17 |
| Marine | 182 | 0.53 | 0.00 | 0.16 | 0.00 |
| Marine | 183 | -0.08 | 0.79 | 0.01 | 0.71 |
| Marine | 187 | -0.45 | 0.28 | -0.50 | 0.28 |
| Marine | 191 | 0.29 | 0.36 | 0.24 | 0.04 |
| Marine | 192 | 0.00 | 1.00 | -0.65 | 0.00 |
| Marine | 193 | 0.63 | 0.00 | -0.03 | 0.32 |
| Terrestrial | 195 | 0.00 | 0.04 | 0.00 | 0.11 |
| Marine | 196 | -0.03 | 0.09 | 0.00 | 0.97 |
| Marine | 197 | 0.35 | 0.00 | 0.06 | 0.01 |
| Marine | 198 | 0.17 | 0.01 | 0.07 | 0.05 |
| Marine | 200 | -0.34 | 0.29 | -0.12 | 0.50 |
| Terrestrial | 277 | 0.00 | 0.99 | 0.00 | 0.99 |

**table S6. Variance ratio tests for patterns of compensatory fluctuations in total abundance.**

Variance ratio tests for patterns of compensatory fluctuations in total abundance. Each row represents a different assemblage. The first column is the ID number of the assemblage (see Additional Data Table S1). The second and third columns are the lower and upper 95% confidence intervals for the variance ratios calculated for 1000 null assemblages generated with the cyclic shift permutation algorithm (*23*). The fifth column is the average variance ratio for the null assemblages. The sixth column is the observed variance ratio for the empirical data. Entries shaded orange are assemblages for which the observed variance ratio is below the lower 95% confidence limit, indicating a pattern of compensatory fluctuations in total abundance.

| ID | lowerCI | upperCI | nullmean | VR |
|---|---|---|---|---|
| 18 | 0.62 | 1.73 | 1.05 | 3.17 |
| 33 | 0.77 | 1.21 | 0.98 | 1.47 |
| 39 | 0.28 | 2.02 | 0.99 | 2.40 |
| 42 | 0.36 | 1.91 | 0.97 | 3.49 |
| 44 | 0.36 | 1.73 | 0.94 | 10.92 |
| 46 | 0.31 | 2.27 | 1.07 | 2.94 |
| 47 | 0.41 | 1.98 | 1.03 | 2.91 |
| 51 | 0.33 | 2.00 | 0.95 | 3.10 |
| 52 | 0.42 | 1.69 | 1.00 | 1.82 |
| 54 | 0.45 | 1.61 | 0.99 | 1.00 |
| 56 | 0.59 | 1.48 | 1.00 | 0.14 |
| 57 | 0.80 | 1.29 | 0.98 | 1.64 |
| 58 | 0.55 | 1.55 | 0.97 | 2.28 |
| 59 | 0.49 | 1.68 | 1.01 | 1.06 |
| 63 | 0.49 | 1.61 | 1.01 | 1.20 |
| 67 | 0.40 | 2.38 | 0.99 | 9.02 |
| 68 | 0.46 | 1.82 | 1.00 | 4.36 |
| 70 | 0.83 | 1.24 | 0.99 | 1.16 |
| 78 | 0.69 | 1.46 | 1.00 | 2.59 |
| 81 | 0.73 | 1.36 | 1.01 | 0.68 |
| 84 | 0.63 | 1.66 | 1.00 | 2.00 |
| 85 | 0.38 | 1.85 | 0.97 | 2.06 |
| 92 | 0.57 | 1.93 | 1.01 | 13.69 |
| 97 | 0.93 | 1.27 | 1.00 | 0.97 |
| 100 | 0.81 | 1.28 | 1.00 | 1.31 |
| 101 | 0.72 | 1.36 | 0.99 | 1.53 |
| 108 | 0.63 | 1.53 | 0.99 | 1.56 |
| 110 | 0.67 | 1.55 | 1.01 | 15.43 |
| 112 | 0.57 | 1.75 | 1.04 | 0.92 |
| 113 | 0.47 | 1.70 | 1.03 | 63.36 |
| 117 | 0.91 | 1.24 | 1.00 | 1.70 |
| 119 | 0.56 | 1.41 | 0.98 | 1.67 |
| 125 | 0.38 | 1.60 | 0.97 | 2.85 |
| 135 | 0.56 | 1.79 | 1.03 | 2.69 |
| 147 | 0.71 | 1.37 | 1.00 | 14.95 |
| 148 | 0.62 | 1.39 | 0.97 | 12.81 |
| 152 | 0.56 | 1.69 | 0.99 | 5.91 |
| 162 | 0.36 | 1.97 | 0.96 | 24.49 |
| 163 | 0.51 | 1.60 | 0.99 | 0.55 |
| 166 | 0.81 | 1.35 | 1.01 | 1.49 |
| 169 | 0.53 | 1.63 | 1.02 | 1.28 |
| 171 | 0.59 | 1.77 | 1.01 | 1.25 |
| 172 | 0.40 | 1.68 | 0.97 | 0.84 |
| 173 | 0.73 | 1.49 | 1.01 | 2.70 |
| 176 | 0.68 | 1.61 | 1.00 | 2.94 |
| 178 | 0.60 | 1.66 | 1.06 | 7.37 |
| 180 | 0.85 | 1.27 | 0.99 | 0.88 |
| 182 | 0.83 | 1.33 | 1.01 | 1.54 |
| 183 | 0.54 | 1.26 | 0.99 | 0.76 |
| 187 | 0.43 | 1.81 | 0.98 | 94.91 |
| 191 | 0.79 | 1.65 | 1.00 | 4.15 |
| 192 | 0.54 | 1.91 | 0.96 | 0.48 |
| 193 | 0.79 | 1.86 | 1.01 | 1.86 |
| 195 | 0.37 | 2.13 | 0.95 | 1.81 |
| 196 | 0.57 | 1.86 | 1.07 | 3.69 |
| 197 | 0.64 | 1.42 | 1.01 | 1.00 |
| 198 | 0.62 | 1.37 | 1.02 | 1.65 |
| 200 | 0.99 | 1.06 | 1.00 | 1.17 |
| 277 | 0.60 | 1.49 | 0.98 | 2.42 |

**table S7. Null model tests for the slope of the relationship between the observed number of colonizations at time *t* and the observed number of extinctions at time *t* + *x*.**

Null model tests for the slope of the relationship between the observed number of colonizations at time *t* and the observed number of extinctions at time $t + x$

Each row represents a different assemblage. The first column is the ID number of the assemblage (see Additional Data Table S1). Columns 2, 3, 4, and 5 give the observed slope of the relationship between colonizations and extinctions for a lag of $x = 0$, 1, 2, or 3 years. Columns 6, 7, 8, and 9 give the corresponding average slopes for these same relationships based on 1000 simulated null assemblages, generated with the cyclic shift permutation algorithm (*23*). Columns 10, 11, 12, and 13 give the corresponding *P* value for a one-tailed test that the observed slope was greater than expected by chance. Rejection of this null hypothesis would correspond to a pattern of extinctions compensating for colonizations with a lag of *x* years.

| ID | obs0 | obs1 | obs2 | obs3 | simAv0 | simAv1 | simAv2 | simAv3 | pval0 | pval1 | pval2 | pval3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | -0.16 | 0.52 | 0.05 | -0.02 | -0.17 | 0.47 | -0.04 | 0.01 | 0.47 | 0.39 | 0.33 | 0.56 |
| 33 | -0.44 | -0.14 | 0.39 | -0.25 | -0.22 | 0.30 | 0.01 | -0.05 | 0.79 | 0.96 | 0.09 | 0.76 |
| 39 | -0.34 | 0.17 | 0.69 | -0.22 | -0.21 | 0.30 | 0.02 | 0.05 | 0.67 | 0.71 | 0.03 | 0.81 |
| 42 | -0.28 | 0.14 | -0.12 | 0.18 | -0.16 | 0.35 | 0.00 | 0.02 | 0.71 | 0.88 | 0.73 | 0.22 |
| 44 | -0.35 | 0.38 | 0.04 | -0.23 | -0.22 | 0.38 | -0.03 | -0.01 | 0.64 | 0.50 | 0.43 | 0.77 |
| 46 | 0.41 | 0.70 | 0.35 | -0.07 | -0.06 | 0.58 | 0.01 | 0.13 | 0.00 | 0.16 | 0.02 | 0.90 |
| 47 | -0.49 | 0.06 | 0.27 | -0.24 | -0.19 | 0.17 | 0.13 | -0.04 | 0.93 | 0.70 | 0.20 | 0.82 |
| 51 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 52 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 54 | -0.34 | 0.70 | -0.39 | 0.58 | -0.26 | 0.31 | 0.00 | 0.02 | 0.63 | 0.07 | 0.91 | 0.03 |
| 56 | -0.42 | 0.40 | -0.14 | 0.38 | -0.24 | 0.36 | 0.04 | 0.10 | 0.77 | 0.42 | 0.75 | 0.14 |
| 57 | 0.67 | 0.51 | -0.11 | 0.65 | -0.14 | 0.44 | 0.02 | -0.03 | 0.00 | 0.34 | 0.75 | 0.00 |
| 58 | -0.23 | 0.75 | 0.22 | 0.22 | -0.17 | 0.40 | 0.05 | -0.04 | 0.57 | 0.06 | 0.26 | 0.18 |
| 59 | 0.03 | 0.51 | 0.08 | -0.25 | -0.11 | 0.42 | 0.08 | 0.02 | 0.24 | 0.31 | 0.48 | 0.89 |
| 63 | -0.06 | 0.08 | -0.07 | -0.07 | -0.05 | 0.12 | -0.04 | -0.05 | 0.65 | 0.56 | 0.54 | 0.54 |
| 67 | -0.38 | 0.48 | 0.27 | -0.16 | -0.19 | 0.38 | 0.10 | -0.07 | 0.81 | 0.31 | 0.19 | 0.66 |
| 68 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 70 | -0.37 | 0.07 | -0.41 | -0.46 | -0.18 | 0.01 | -0.08 | 0.03 | 0.74 | 0.40 | 0.85 | 0.92 |
| 78 | -0.11 | 0.77 | 0.06 | 0.08 | -0.17 | 0.67 | -0.01 | 0.06 | 0.41 | 0.29 | 0.40 | 0.48 |
| 81 | 0.07 | 0.00 | 0.63 | -0.40 | -0.19 | -0.03 | 0.34 | -0.18 | 0.15 | 0.50 | 0.15 | 0.76 |
| 84 | -0.77 | 0.64 | -0.26 | 0.50 | -0.22 | 0.51 | -0.03 | -0.04 | 0.93 | 0.35 | 0.71 | 0.10 |
| 85 | -0.20 | 1.07 | -0.18 | 0.09 | -0.20 | 0.71 | -0.02 | -0.11 | 0.48 | 0.07 | 0.69 | 0.29 |
| 92 | -0.38 | 0.78 | -0.18 | -0.21 | -0.16 | 0.84 | -0.05 | -0.05 | 0.87 | 0.69 | 0.72 | 0.75 |
| 97 | -0.47 | 0.74 | 0.58 | -0.09 | -0.14 | 0.69 | 0.09 | -0.02 | 0.93 | 0.39 | 0.02 | 0.62 |
| 100 | -0.39 | 0.25 | 0.20 | 0.26 | -0.19 | 0.53 | -0.04 | 0.00 | 0.85 | 0.96 | 0.11 | 0.10 |
| 101 | -0.24 | 0.14 | 0.02 | 0.03 | -0.11 | 0.34 | -0.03 | 0.10 | 0.72 | 0.87 | 0.38 | 0.64 |
| 108 | -0.25 | 0.84 | -0.13 | 0.24 | -0.18 | 0.67 | -0.02 | 0.01 | 0.65 | 0.13 | 0.71 | 0.13 |
| 110 | -0.15 | 0.84 | 0.15 | 0.10 | -0.09 | 0.83 | 0.04 | 0.00 | 0.65 | 0.50 | 0.26 | 0.32 |
| 112 | -0.17 | -0.14 | 0.08 | 0.05 | -0.25 | 0.47 | -0.12 | 0.03 | 0.39 | 0.98 | 0.26 | 0.46 |
| 113 | 0.19 | 0.84 | 0.52 | 0.29 | -0.13 | 0.74 | 0.03 | -0.02 | 0.07 | 0.25 | 0.01 | 0.08 |
| 117 | 0.21 | 0.49 | 0.30 | 0.06 | -0.05 | 0.85 | 0.06 | 0.01 | 0.04 | 1.00 | 0.07 | 0.35 |
| 119 | 0.08 | 0.87 | 0.46 | 0.77 | -0.11 | 0.60 | 0.05 | 0.02 | 0.15 | 0.02 | 0.01 | 0.00 |
| 125 | -0.54 | 0.51 | 0.04 | -0.41 | -0.18 | 0.59 | -0.02 | -0.07 | 0.88 | 0.62 | 0.43 | 0.82 |
| 135 | 0.57 | 1.02 | 0.51 | 0.01 | -0.10 | 0.81 | 0.09 | -0.08 | 0.01 | 0.11 | 0.06 | 0.38 |
| 147 | -0.11 | 0.98 | -0.05 | -0.02 | -0.05 | 0.93 | 0.02 | -0.01 | 0.65 | 0.19 | 0.67 | 0.52 |
| 148 | -0.07 | 0.81 | 0.11 | -0.13 | -0.08 | 0.83 | 0.05 | -0.02 | 0.48 | 0.58 | 0.35 | 0.76 |
| 152 | 0.35 | 0.18 | 0.84 | 2.12 | -0.10 | 0.70 | 0.10 | -0.05 | 0.05 | 0.99 | 0.01 | 0.00 |
| 162 | 0.25 | 0.88 | 0.42 | -0.10 | -0.15 | 0.65 | 0.01 | -0.08 | 0.10 | 0.18 | 0.12 | 0.51 |
| 163 | -0.41 | 0.47 | -0.02 | 0.28 | -0.22 | 0.51 | -0.03 | -0.04 | 0.70 | 0.57 | 0.47 | 0.20 |
| 166 | -0.33 | 0.77 | -0.10 | 0.08 | -0.16 | 0.57 | 0.01 | 0.09 | 0.78 | 0.13 | 0.68 | 0.54 |
| 169 | -0.44 | 0.68 | -0.47 | 0.31 | -0.20 | 0.45 | -0.03 | 0.01 | 0.83 | 0.15 | 0.95 | 0.13 |
| 171 | -0.34 | 0.73 | -0.19 | 0.33 | -0.22 | 0.64 | -0.05 | 0.00 | 0.65 | 0.34 | 0.68 | 0.13 |
| 172 | 0.04 | 0.13 | -0.07 | 0.05 | -0.21 | 0.33 | -0.08 | -0.03 | 0.23 | 0.72 | 0.48 | 0.42 |
| 173 | -0.12 | 0.86 | 0.32 | -0.35 | -0.14 | 0.85 | -0.04 | -0.06 | 0.48 | 0.46 | 0.11 | 0.83 |
| 176 | -0.56 | 0.88 | -0.23 | 0.19 | -0.29 | 0.65 | -0.16 | 0.00 | 0.81 | 0.17 | 0.59 | 0.30 |
| 178 | 0.05 | 0.65 | 0.42 | 0.25 | -0.13 | 0.57 | 0.04 | 0.03 | 0.14 | 0.31 | 0.02 | 0.10 |
| 180 | -0.75 | 0.88 | -0.38 | 0.51 | -0.22 | 0.51 | 0.02 | -0.04 | 0.99 | 0.02 | 0.96 | 0.01 |
| 182 | -0.30 | 0.04 | 0.23 | -0.07 | -0.15 | 0.17 | 0.10 | -0.04 | 0.73 | 0.71 | 0.29 | 0.53 |
| 183 | -0.21 | 0.04 | 0.19 | -0.24 | -0.18 | -0.20 | 0.15 | -0.20 | 0.51 | 0.15 | 0.46 | 0.53 |
| 187 | -0.19 | 1.07 | -0.20 | 0.24 | -0.15 | 0.84 | -0.08 | -0.02 | 0.55 | 0.10 | 0.68 | 0.23 |
| 191 | -0.36 | 0.83 | 0.01 | 0.00 | -0.12 | 0.72 | 0.02 | 0.03 | 0.87 | 0.22 | 0.54 | 0.56 |
| 192 | -0.10 | 0.22 | 0.15 | 0.26 | -0.12 | 0.27 | 0.04 | 0.10 | 0.43 | 0.62 | 0.23 | 0.15 |
| 193 | -0.40 | 0.21 | 0.21 | -0.25 | -0.23 | 0.24 | 0.01 | -0.18 | 0.81 | 0.58 | 0.13 | 0.66 |
| 195 | -0.26 | 0.06 | 0.10 | -0.04 | -0.20 | 0.32 | 0.01 | -0.05 | 0.60 | 0.93 | 0.33 | 0.47 |
| 196 | -0.05 | 0.60 | 0.33 | 0.29 | -0.11 | 0.46 | 0.05 | 0.02 | 0.35 | 0.18 | 0.06 | 0.08 |
| 197 | -0.14 | 0.53 | -0.03 | 0.00 | -0.15 | 0.46 | 0.04 | 0.00 | 0.47 | 0.36 | 0.66 | 0.50 |
| 198 | -0.60 | 0.31 | 0.06 | -0.16 | -0.16 | 0.42 | 0.06 | -0.01 | 0.97 | 0.73 | 0.49 | 0.70 |
| 200 | -0.10 | 0.99 | 0.01 | -0.01 | -0.05 | 0.94 | 0.00 | -0.02 | 0.58 | 0.26 | 0.47 | 0.48 |
| 277 | -0.36 | 0.25 | 0.52 | 0.50 | -0.24 | 0.44 | -0.11 | 0.02 | 0.63 | 0.70 | 0.08 | 0.16 |

**table S8 (separate file)**

Primary references and metadata for 59 assemblage time series data sets.