Available online at www.sciencedirect.com

ScienceDirect

Palaeoworld 33 (2024) 1170–1178

Palaeoworld

www.elsevier.com/locate/palwor

Per-capita change rates of the Phanerozoic Earth-life system exhibited Zipf distributions

Haitao Shang

Institute of Ecology and Evolution, University of Oregon, OR 97403, USA

Received 25 April 2023; received in revised form 16 November 2023; accepted 21 December 2023
Available online 28 December 2023

Abstract

The Zipf distribution is widespread in physical and biological systems on the modern Earth; whether such distributions existed in the ancient Earth-life system, however, remains understudied. Here I demonstrate that per-capita change rates (PCCRs) of the atmospheric CO_2 level, global average temperature, genus richness, and body size during the Phanerozoic exhibited Zipf distributions. Statistical analyses, including the goodness-of-fit test, likelihood-ratio test, and complementary cumulative distribution function (CCDF), support these Zipf distributions. Moreover, correlation analyses show that these environmental variables and biological metrics are correlated while their PCCRs are not, suggesting that Zipf distributions in PCCRs are unlikely to directly connect to each other; instead, there probably existed a set of common but unknown factors responsible for these systematic Zipf distributions in PCCRs. The results in this study offer new perspectives on the evolutionary patterns of the ancient Earth-life system.

Keywords: Zipf distribution; Per-capita change rate; Phanerozoic; Environmental variables; Biological metrics

1. Introduction

Geological and biological systems on Earth's surface underwent significant changes during the Phanerozoic. Modern geochemical techniques and theoretical models are remarkably successful at reconstructing the evolutionary trajectories of environmental variables such as the atmospheric CO_2 level and global average temperature. These variables are closely tied to the course of life evolution. Changes in the atmospheric CO_2 level manifest the variations in the geological and biological components of Earth's carbon cycle (Hayes and Waldbauer, 2006; Galvez et al., 2020). Temperature significantly influences oceanic redox states and therefore the metabolisms and evolution of marine life (Sheridan and Bickford, 2011; Deutsch et al., 2015). As a methodological counterpart of environmental variables, fossil records provide complementary information to investigate the variations in the Earth-life system. Significant efforts have been dedicated to revealing the patterns, and underlying mechanisms, of the Phanerozoic evolution of life from fossil records (Sepkoski, 2002; Alroy et al., 2008). Several quantitative metrics, including genus richness and body size, have been widely used to distill meaningful information from fossil data. Many intriguing phenomena appearing in the evolutionary course of Phanerozoic life, such as cyclicity (Raup and Sepkoski, 1984; Rohde and Muller, 2005), contingency (Gould, 1990; Blount et al., 2018), and self-organization criticality (Newman, 1996; Bak, 2013), have been brought to light, enriching our understanding of the fundamental mechanisms governing the evolution of life at geologic timescales.

The Zipf distribution, also referred to as Zipf's law or in a more general form the Zipf-Mandelbrot law, is a mathematical distribution in which the r-th most frequent event

E-mail address: htshang.research@gmail.com

https://doi.org/10.1016/j.palwor.2023.12.006

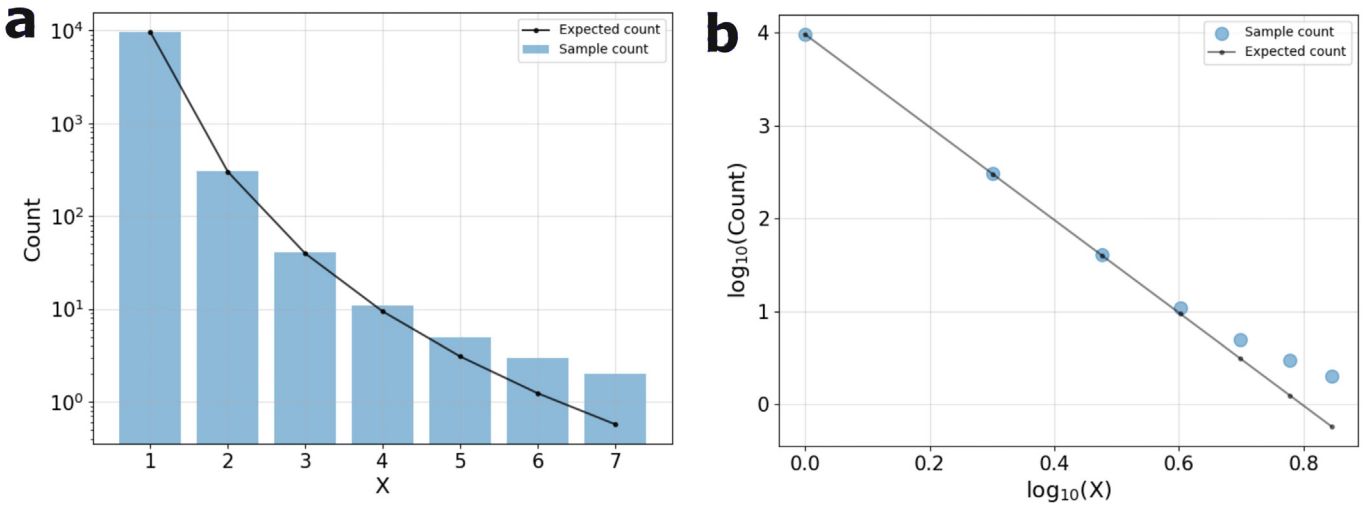


Fig. 1. (a) Histogram of sample counts (blue bars) of 10,000 data points randomly sampled from a Zipf distribution: $p(X) = X^{-5}$. Black dots represent expected counts. (b) Log-log plot of the sample counts versus X (blue circles) of 10,000 data points randomly sampled from the same Zipf distribution: $p(X) = X^{-5}$ and log-log plot of expected counts versus X (black dots). The expected counts in panels (a) and (b) are calculated based on the given Zipf distribution: $p(X) = X^{-5}$.

has a frequency $f(r)$ that scales as $f(r) \propto r^{-\alpha}$, where r and $\alpha > 0$ are the frequency rank and Zipf exponent, respectively (Zipf, 1949). In other words, the Zipf distribution describes a process in which the probability of an observation is inversely proportional to its rank. Fig. 1 shows the distributions of data randomly sampled from Zipf's law. Fig. 1a illustrates the distribution on a histogram and Fig. 1b demonstrates the distribution on a log-log plot. The Zipf distribution is manifested by a straight line on a double logarithmic plot (Fig. 1b), which is also called a Zipf plot in this context, and implies that the underlying regularity of such a relation is independent of a specific scale (Bak et al., 1988; Schroeder, 2009; Bak, 2013). The Zipf distribution appears widely in natural and social phenomena, such as the frequencies of amino acids (Mora et al., 2010) and expressed genes (Furusawa and Kaneko, 2003) in various organisms and tissues, spiking activities of neurons (Mora and Bialek, 2011; Tyrcha et al., 2013), magnitudes of earthquakes (Mega et al., 2003; Gerlach and Altmann, 2019), frequencies of words in written texts (Cancho and Solé, 2003; Zipf, 1949), and fluctuations in financial markets (Ausloos and Bronlet, 2003; Fujiwara, 2004). However, whether Zipf distributions existed in the ancient Earth-life system remains understudied.

Here, to explore the Zipf distribution in deep-time geological and biological systems, I investigate the time series of Phanerozoic atmospheric CO_2 level, global average temperature, genus richness, and body size reconstructed by previous studies (Heim et al., 2015; Foster et al., 2017; Kocsis et al., 2019; Scotese et al., 2021). I show that the Zipf distribution exists in the PCCRs of these four quantities; the goodness-of-fit test, likelihood-ratio test, and CCDF support this result. These Zipf distributions in the four key environmental variables and biological metrics suggest that extreme changes were much rarer than small ones during the Phanerozoic, implying that the Earth-life system likely had developed strong resilience to crises in this eon. The fundamental mechanisms responsible for these systematic Zipf distributions remain unknown; further studies are required to explore the factors resulting in such patterns.

2. Materials and methods

The datasets on the Phanerozoic atmospheric CO_2 level and global average temperature are from Foster et al. (2017) and Scotese et al. (2021), respectively. Original data points in these two time series are evenly spaced with time steps of 0.5 million years (Myr) and 1 Myr, respectively. The dataset on the Phanerozoic genus richness is originally from the Paleobiology Database (https://paleobiodb.org/) and standardized/subsampled in the same manner as the work by Alroy et al. (2008) and Kocsis et al. (2019). The dataset on body sizes of Phanerozoic marine animals is

Fig. 2. Zipf distributions in the PCDRs, PCIRs, and PCVRs of (a–c) atmospheric CO_2 level (CO_2), (d–f) global average temperature (T), (g–i) genus richness (GR), and (j–l) body size (BS) in the Phanerozoic. The time step between two consecutive data points in the interpolated time series of these four quantities is 0.1 Myr (Section 2). Filled green squares, red diamonds, blue triangles, and orange circles represent the PCCRs versus counts (i.e., N's) for the Phanerozoic atmospheric CO_2 level, global average temperature, genus richness, and body size, respectively. The initial small values in panels (b) and (g) that do not follow the Zipf distribution are not included in data fitting (Section 2). Gray dashed lines are the best-fitting Zipf distributions for the PCCRs of each dataset.

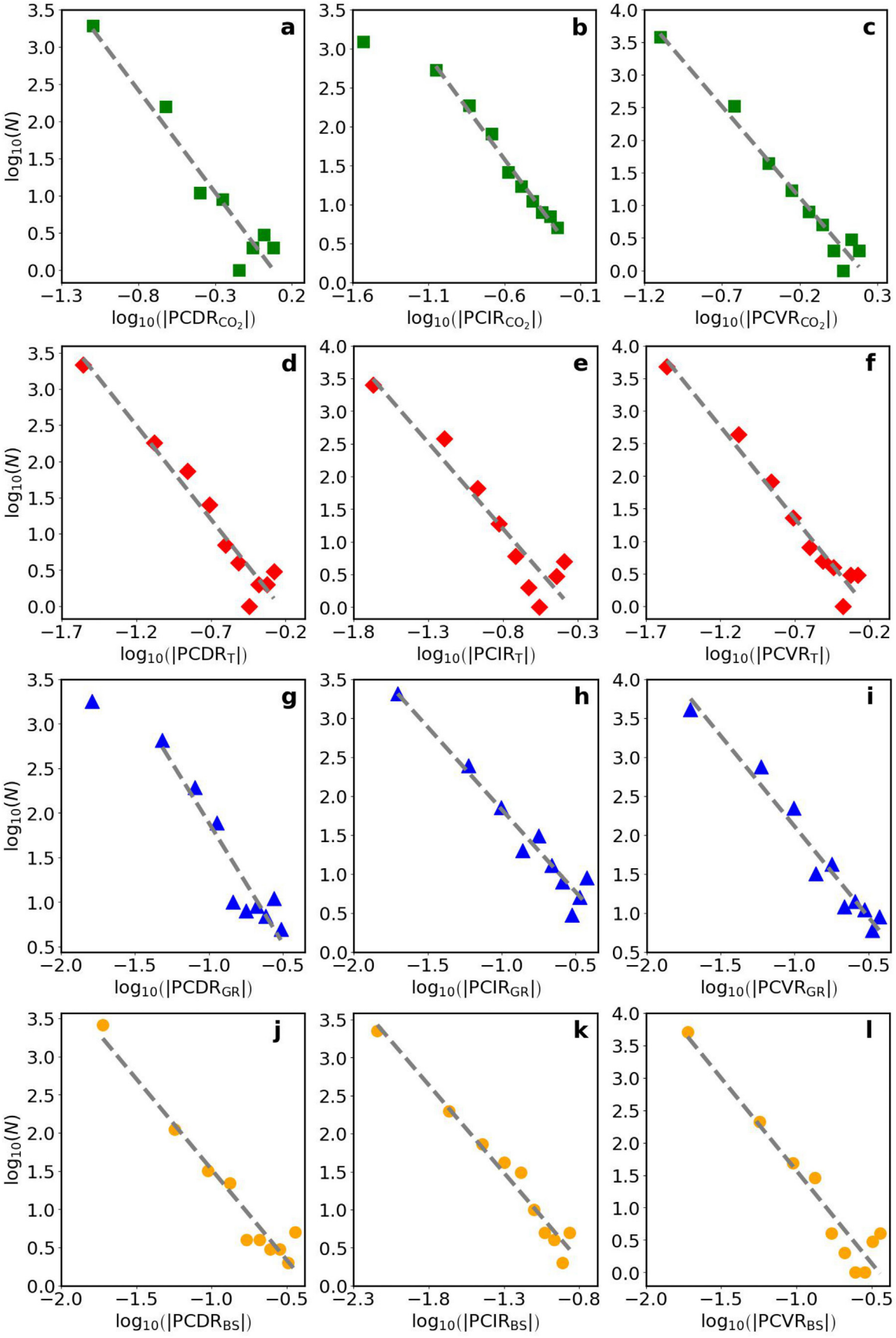

a
b
c
d
e
f
g
h
i
j
k
l
log10(N)
log10(|PCDRCO2|)
log10(|PCIRCO2|)
log10(|PCVRCO2|)
log10(|PCDRT|)
log10(|PCIRT|)
log10(|PCVRT|)
log10(|PCDRGR|)
log10(|PCIRGR|)
log10(|PCVRGR|)
log10(|PCDRBS|)
log10(|PCIRBS|)
log10(|PCVRBS|)

Table 1
Zipf distributions for the PCDRs, PDICs, and PCVRs of the atmospheric CO_2 level (CO_2), global average temperature (T), genus richness (GR), and body size (BS). The time step between two consecutive data points on the interpolated time series of these four quantities is 0.1 Myr (Section 2). The R^2, RMSE, p_{KS}, and p_{CM} represent the coefficient of determination, root mean square error, the p-value of the KS test, and the p-value of the CM test, respectively, of the least-squares fitting on the log-log plots. The LR_{EXP}, $LR_{S\text{-}EXP}$, and LR_{LN} represent the log-likelihood ratios for the Zipf distribution against exponential, stretched exponential, and lognormal distributions, respectively, for each dataset. The p_{EXP}, p_{S-EXP}, and p_{LN} are the p-values for the statistical significance of these tests.

Category	Quantity	Zipf distribution	Goodness-of-fit test				Likelihood-ratio test					
			R^2	RMSE	p_{KS}	p_{CM}	LR_{EXP}	p_{EXP}	$LR_{S\text{-}EXP}$	p_{S-EXP}	LR_{LN}	p_{LN}
PCDR	Atmospheric CO_2 level	$N \sim \lvert PCDR_{CO_2}\rvert^{-2.67}$	0.97	0.11	0.96	0.98	8.39	4.65×10^{-17}	4.31	4.31×10^{-5}	5.25	1.49×10^{-7}
	Global average temperature	$N \sim \lvert PCDR_T\rvert^{-2.33}$	0.98	0.12	0.93	0.88	16.24	2.55×10^{-59}	3.58	0.005	0.19	0.197
	Genus richness	$N \sim \lvert PCDR_{GR}\rvert^{-2.71}$	0.95	0.10	1.00	1.00	14.33	1.49×10^{-46}	2.20	0.014	2.20	0.006
	Body size	$N \sim \lvert PCDR_{BS}\rvert^{-2.24}$	0.93	0.20	0.93	0.76	14.59	9.35×10^{-48}	6.22	4.99×10^{-10}	5.97	2.04×10^{-9}
PCIR	Atmospheric CO_2 level	$N \sim \lvert PCIR_{CO_2}\rvert^{-2.82}$	0.94	0.17	0.94	0.89	8.56	1.10×10^{-17}	6.84	7.46×10^{-12}	8.65	5.28×10^{-18}
	Global average temperature	$N \sim \lvert PCIR_T\rvert^{-2.73}$	0.94	0.13	0.92	0.89	7.79	6.51×10^{-15}	2.02	0.034	1.10	0.013
	Genus richness	$N \sim \lvert PCIR_{GR}\rvert^{-2.32}$	0.89	0.25	0.99	1.00	23.74	1.39×10^{-124}	11.6	3.92×10^{-31}	12.86	7.40×10^{-38}
	Body size	$N \sim \lvert PCIR_{BS}\rvert^{-2.34}$	0.95	0.21	0.99	0.92	5.32	1.01×10^{-7}	2.28	0.003	1.17	0.023
PCVR	Atmospheric CO_2 level	$N \sim \lvert PCVR_{CO_2}\rvert^{-2.63}$	0.96	0.19	0.93	0.93	8.21	2.21×10^{-16}	2.86	0.004	0.81	0.015
	Global average temperature	$N \sim \lvert PCVR_T\rvert^{-2.72}$	0.87	0.16	0.95	0.98	13.11	3.06×10^{-39}	2.04	0.041	3.54	0.004
	Genus richness	$N \sim \lvert PCVR_{GR}\rvert^{-2.33}$	0.91	0.27	0.99	1.00	16.15	1.15×10^{-58}	9.79	1.20×10^{-22}	10.04	9.69×10^{-24}
	Body size	$N \sim \lvert PCVR_{BS}\rvert^{-2.85}$	0.97	0.14	1.00	0.99	7.99	1.27×10^{-15}	1.01	0.147	1.47	0.009

from the study by Heim et al. (2015); the original data points in this time series are at the stage level (i.e., time intervals between consecutive points have varying lengths). I group these data points of body sizes into bins with the width of 1 Myr and calculate the mean in each bin; the time step of the time series for mean body size therefore is 1 Myr. The Zipf distribution usually spans several orders of magnitude; a time series with a finer time step offers more data points and therefore a more prominent Zipf-distribution pattern. Therefore, I interpolate the datasets of these four quantities using smoothing splines, which have been demonstrated to be effective for fitting these quantities (Rohde and Muller, 2005; Foster and Rohling, 2013; Vérard and Veizer, 2019), with the time step of 0.1 Myr. On the other hand, such a high resolution of stratigraphic age is generally difficult to achieve in reality. To mimic the coarser stratigraphic resolution in practice, I interpolate the four time series with the time step of 1 Myr. These two sets of time series with finer and coarser time resolution (i.e., 0.1 Myr and 1 Myr) are both used for the following analyses.

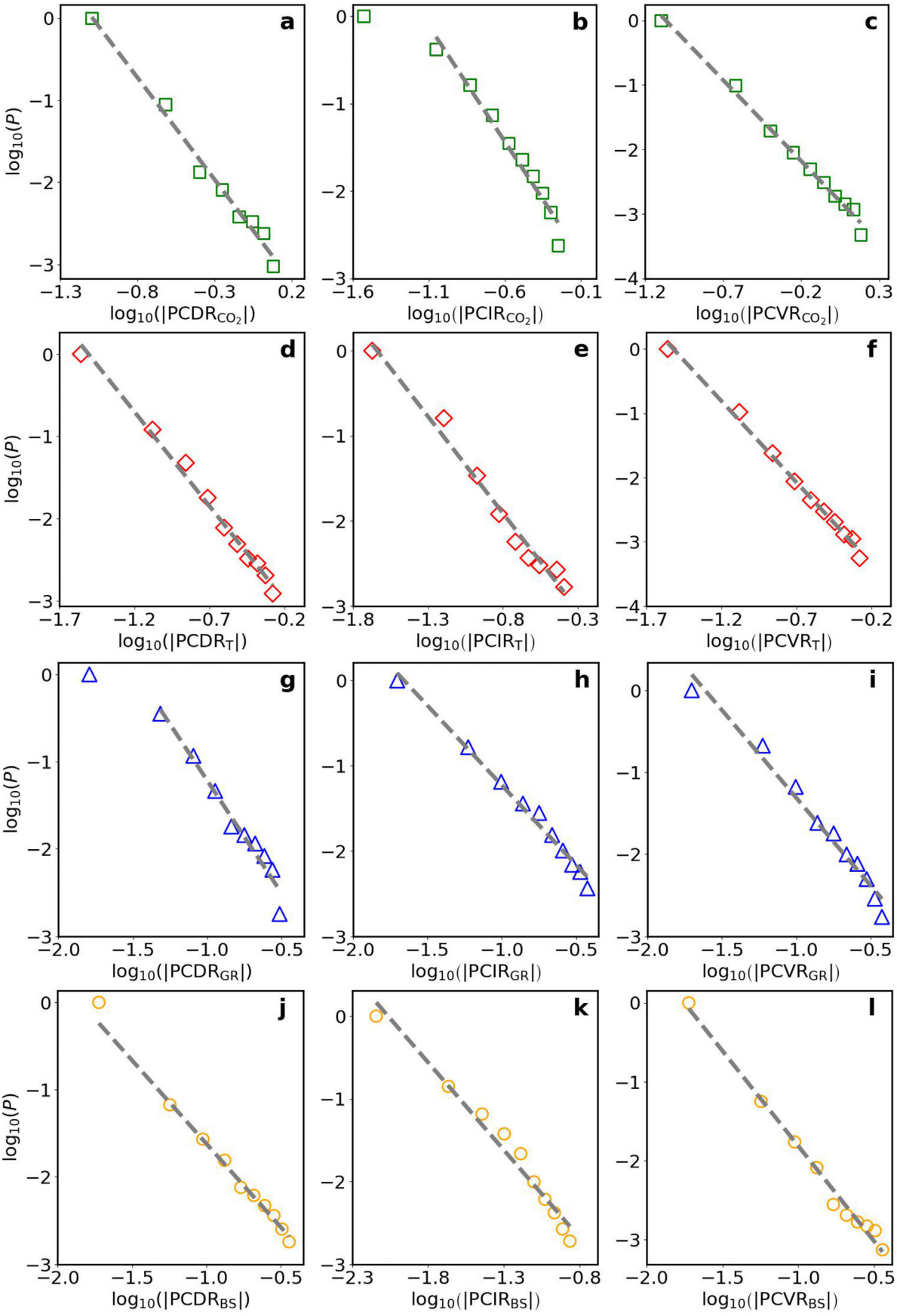
I denote the difference between two consecutive data points by $\Delta\varphi_i = \varphi_{i+1} - \varphi_i$ and the length of a time step by $\Delta t_i = t_{i+1} - t_i$. The PCCR at time t_i, $PCCR_i$, which measures the variation $\Delta\varphi_i$ relative to the original value φ_i during the unit time Δt_i is expressed as

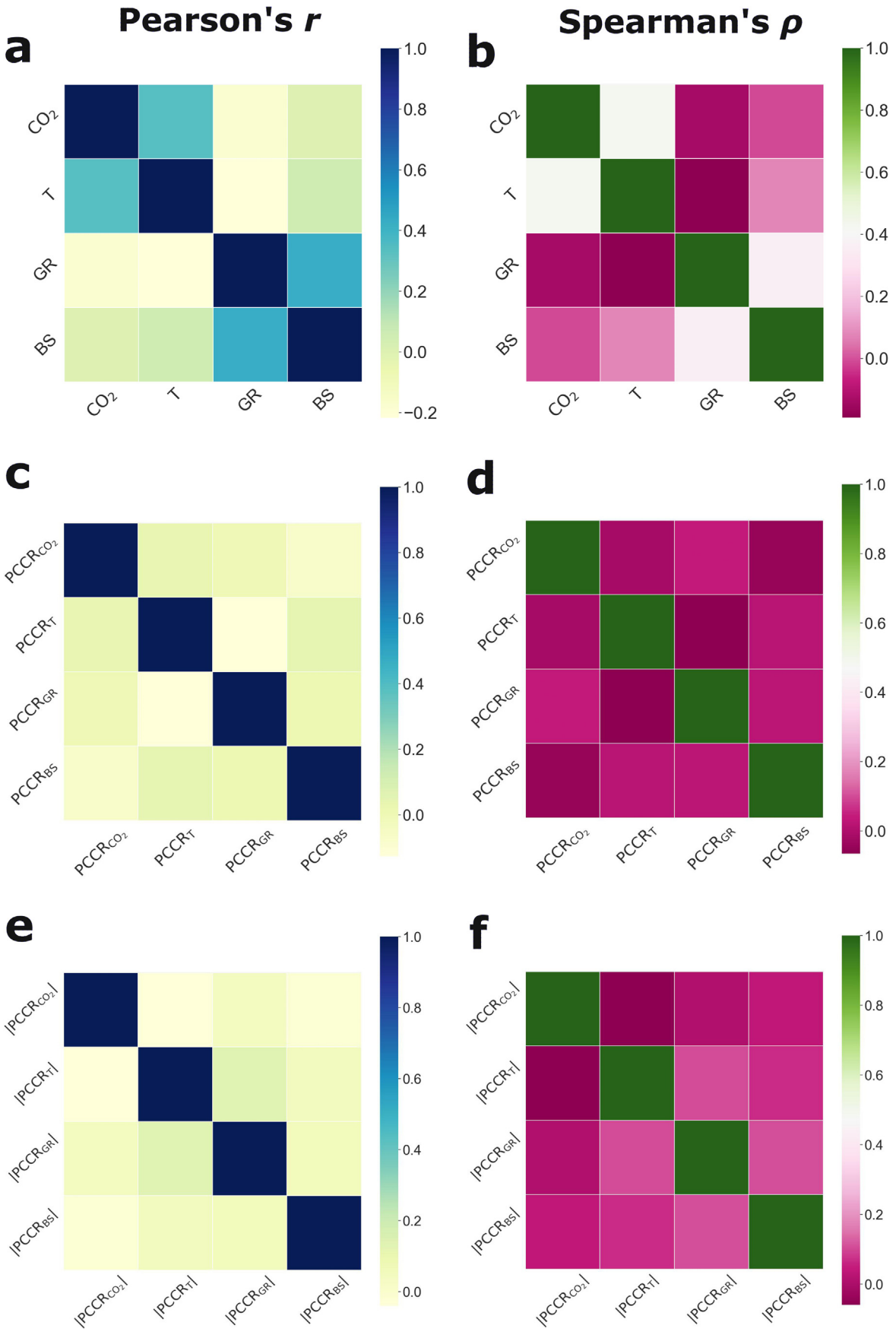
$$PCCR_i = \frac{1}{\Delta t}\frac{\Delta\varphi_i}{\varphi_i}.$$

I divide PCCRs of each variable into three groups: per-capita decrease rate (PCDR), per-capita increase rate (PCIR), and per-capita variation rate (PCVR); the r_i's that fall into these three groups are denoted by $PCDR = \{|PCCR_i \| PCCR_i < 0\}$, $PCIR = \{PCCR_i | PCCR_i > 0\}$, and $PCVR = \{|PCCR_i \| PCCR_i \neq 0\}$, respectively. Logarithms of negative values are undefined; absolute values of data in these three groups, denoted by $|PCDR| = \{|PCCR_i \| PCCR_i < 0\}$, $|PCIR| = \{PCCR_i | PCCR_i > 0\}$, and $|PCVR| = \{|PCCR_i \| PCCR_i \neq 0\}$, are used to analyze the Zipf distributions in the datasets. Zero values (i.e., $PCCR_i = 0$) are not included in the analyses for two reasons: they indicate no change, and an exponential function with a zero base is undefined mathematically when the exponent in a Zipf distribution is negative.

Zipf distributions are heavy-tailed, which means that large values appear at their right ends (Clauset et al., 2009; Alstott et al., 2014). The interesting properties of Zipf distributions are in their heavy tails; conventionally, the initial data points that do not follow a Zipf distribution are discarded prior to fitting the data (Clauset et al., 2009; Alstott et al., 2014). Identifying where to truncate the time series $\{PCCR_i\}$ requires finding the data point

Fig. 3. Complementary cumulative distribution functions (CCDFs) P for the PCDRs, PCIRs, and PCVRs of (a–c) atmospheric CO_2 level (CO_2), (d–f) global average temperature (T), (g–i) genus richness (GR), and (j–l) body size (BS) in the Phanerozoic. The time step between two consecutive data points in the interpolated time series of these four quantities is 0.1 Myr (Section 2). Empty green squares, red diamonds, blue triangles, and orange circles represent the CCDFs for PCCRs of the Phanerozoic atmospheric CO_2 level, global average temperature, genus richness, and body size, respectively. The initial values in panels (b) and (g) that do not follow the Zipf distribution are not included in data fitting (Section 2). Gray dashed lines are the best-fitting Zipf distributions for the CCDFs of each dataset.


a
b
c
d
e
f
g
h
i
j
k
l
log10(P)
log10(|PCDRCO2|)
log10(|PCIRCO2|)
log10(|PCVRCO2|)
log10(|PCDRT|)
log10(|PCIRT|)
log10(|PCVRT|)
log10(|PCDRGR|)
log10(|PCIRGR|)
log10(|PCVRGR|)
log10(|PCDRBS|)
log10(|PCIRBS|)
log10(|PCVRBS|)


Pearson's r
Spearman's ρ
a
b
c
d
e
f
CO2
T
GR
BS
PCCRCO2
PCCRT
PCCRGR
PCCRBS
|PCCRCO2|
|PCCRT|
|PCCRGR|
|PCCRBS|
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

($PCCR_c$) beyond which the Zipf distribution starts to appear. To determine $PCCR_c$, I first establish a Zipf-distribution fit beginning at each individual data point in $\{PCCR_i\}$ and then select the point giving the minimum distance between the data and the fit (Clauset et al., 2009; Alstott et al., 2014). The truncated dataset starting at $PCCR_c$ thus is the optimal Zipf distribution for the original dataset $\{PCCR_i\}$. The Kolmogorov-Smirnov (KS) test (Massey, 1951) and Cramer-von Mises (CM) two-sample test (Anderson, 1962) are performed to justify that the fitted Zipf distributions adequately describe the datasets. Following the convention, I set the critical p-value to 0.05. If the p-value is greater than 0.05, one fails to reject the null hypothesis and concludes that F_{data} is very likely to be the same as F_0; otherwise, one rejects the null hypothesis and concludes that F_{data} is unlikely to be the same as F_0.

To verify whether the Zipf distribution performs better than other heavy-tailed distributions, such as stretched-exponential and lognormal distributions, I perform the likelihood-ratio test (Clauset et al., 2009; Alstott et al., 2014): $LR = \log\left[\mathscr{L}_Z\left(\text{Data}|\widehat{\Theta_Z}\right)/\mathscr{L}_A\left(\text{Data}|\widehat{\Theta_A}\right)\right]$, in which $\mathscr{L}_Z$ is the likelihood of a Zipf distribution (fitted using the above procedure), $\mathscr{L}_A$ is the likelihood of an alternative distribution (fitted with maximum likelihood estimation), and $\widehat{\Theta_Z}$ and $\widehat{\Theta_A}$ are the parameter values that maximize $\mathscr{L}_Z$ and $\mathscr{L}_A$, respectively, over the parameter space. If $LR > 0$, then the Zipf distribution outperforms the alternative distribution for a given dataset; nevertheless, if $LR < 0$, then the alternative distribution fits a given dataset better than the Zipf distribution. Furthermore, I standardize the LR with the standard deviation to correct for chance fluctuations (Vuong, 1989; Clauset et al., 2009) and obtain a p-value for justifying whether the sign of LR is statistically significant. I set the significance level for the p-value as 0.05. If $p < 0.05$, then the likelihood-ratio test is conclusive (because the sign of LR is unlikely to originate from random fluctuations) and therefore one distribution performs better than the other for a given dataset. However, if $p \geq 0.05$, then the likelihood-ratio test is inconclusive (because the sign of LR is likely to be caused by random fluctuations) and thus both $\mathscr{L}_Z$ and $\mathscr{L}_A$ may provide plausible fits for a given dataset.

3. Results and discussion

3.1. Zipf distributions

Fig. 2 illustrates Zipf distributions in the PCCRs of (a–c) atmospheric CO_2 level, (d–f) global average temperature, (g–i) genus richness, and (k–l) body size in the Phanerozoic. The original time series of these four quantities are interpolated with the time step of 0.1 Myr (Section 2). Table 1 presents the fitted mathematical expressions of these Zipf distributions, coefficients of determination (R^2's), root mean square errors (RMSEs), and p-values of the Kolmogorov-Smirnov (p_{KS}) and Cramer–von Mises (p_{CM}) tests (Section 2). The p-values of these tests are all much greater than the critical threshold of 0.05 (Table 1), suggesting that the mathematical expressions of Zipf distributions in Table 1 fit the data well. A straight line on a log-log plot is a necessary but not sufficient condition for a Zipf distribution; datasets generated by other heavy-tailed distributions such as stretched-exponential and lognormal distributions may exhibit patterns close to Zipf distributions due to chance fluctuations (Clauset et al., 2009).

To further confirm that the straight lines in Fig. 2 are Zipf distributions, I perform the likelihood-ratio test against exponential, stretched-exponential, and lognormal distributions (Section 2); the results are presented in Table 1. For most of these tests, the standardized likelihood ratios are greater than zero and their associated p-values are smaller than the critical value of 0.05, implying that the Zipf distribution fits a given dataset better than other heavy-tailed distributions (Section 2). The only two exceptions are (1) the test for the Zipf distribution against the lognormal distribution for $|PCDR_T|$ and (2) the test for the Zipf distribution against the stretched-exponential distribution for $|PCVR_{BS}|$. In these two situations, while the likelihood ratios are greater than 0, the accompanying p-values are larger than 0.05, suggesting that these two tests are not conclusive. In other words, the results of these two tests imply that (1) both the Zipf and lognormal distributions may offer plausible fits for $|PCDR_T|$ and (2) both the Zipf and stretched-exponential distributions may be plausible for fitting $|PCVR_{BS}|$. On the other hand, Fig. 2 and Table 1 show that the Zipf distribution systematically appears in the PCCRs of the four quantities and generally outperforms alternative distributions for these PCCRs. These observations together suggest that the Zipf distribution also serves as a good option for fitting $|PCDR_T|$ and $|PCVR_{BS}|$. Moreover, an important characteristic of the Zipf distribution is that its CCDF equals $P(X) = \Pr(X \geq x) \propto X^{-\beta}$, where X is a random variable and $\beta > 0$ is an exponent (Velarde and Robledo, 2017; Corral et al., 2020). Fig. 3 shows that the CCDFs for the PCDRs, PDICs, and PCVRs of the four quantities also exhibit such patterns, which provides extra support for the Zipf distributions presented in Fig. 2 and Table 1.

Fig. 4. Heatmaps of the Pearson correlation coefficient (Pearson's r) and Spearman's rank correlation coefficient (Spearman's ρ) for (a, b) each pair of quantities (i.e., the Phanerozoic atmospheric CO_2 level (CO_2), global average temperature (T), genus richness (GR), and body size (BS)), (c, d) each pair of PCCRs of these quantities, and (e, f) each pair of the absolute values of these PCCRs. The time step between two consecutive data points in the interpolated time series of the four quantities is 0.1 Myr (Section 2). The numeric values of the Pearson's r and Spearman's ρ for each pair and their p-values are presented in Table S2 in Supplementary data.

While the results presented in Figs. 2, 3, and Table 1 are plausible, the temporal resolution (i.e., time step) of these time series, 0.1 Myr, is usually difficult to reach with the current chronological techniques. To investigate the results of using datasets with a coarser temporal resolution, I perform the same analyses on the four time series with the time step of 1 Myr (Section 2); the Zipf distributions, CCDFs, and statistical analyses are presented in Figs. S1, S2, and Table S1, respectively, in Supplementary data. The patterns and statistical results of the time series with a coarser temporal resolution are similar to those of the time series with a finer time step, implying that time steps used for interpolation do not influence the basic patterns of the time series and that the Zipf distribution is an intrinsic property of the PCCRs of the four quantities studied in this work.

3.2. Correlations

To investigate whether the Zipf distributions in the PCCRs of the four quantities — atmospheric CO_2 level, global average temperature, genus richness, and body size — have certain connections to each other, I calculate the Pearson correlation coefficient (i.e., Pearson's r) and Spearman's rank correlation coefficient (i.e., Spearman's ρ) for each pair of the four quantities, of the PCCRs of these quantities, and of the absolute values of these PCCRs. The results are illustrated in Fig. 4 and summarized in Table S2 in Supplementary data. The values of Pearson's r and Spearman's ρ for (1) CO_2 level vs. temperature and (2) genus richness vs. body size are positive while their p-values are much less than 0.05 (the critical p-value), suggesting that these pairs of quantities are positively correlated. Such relations are consistent with the observations in previous research, which has explained these positive correlations as consequences of (1) the greenhouse effect of CO_2 (Berner, 1990; Sarmiento, 2006) and (2) the passive diffusion of body size during evolution (Stanley, 1973; Trammer, 2005), respectively. On the other hand, the values of Pearson's r and Spearman's ρ for (1) CO_2 level vs. genus richness and (2) temperature vs. genus richness are negative while their p-values are much smaller than 0.05, implying that these pairs of quantities are negatively correlated. These relations may be attributed to (1) the ability of diverse species in natural ecosystems to remove the atmospheric CO_2 and store carbon over geologic timescales (Cardinale et al., 2012; Hooper et al., 2012) and (2) the changes in extinction rates with temperature (Mayhew et al., 2008; Song et al., 2021), respectively. However, statistically significant correlations appear neither among the PCCRs of these quantities nor among the absolute values of these PCCRs (Fig. 4 and Table S2 in Supplementary data), indicating that the Zipf distributions presented in Fig. 2 and Table 1 are unlikely to directly connect to each other. Instead, the systematic Zipf distributions in the PCCRs of geological and biological systems during the Phanerozoic suggest that these patterns may derive from some common but unknown mechanisms.

3.3. Closing remarks

This study shows that the Zipf distribution exists in the PCCRs of the atmospheric CO_2 level, global average temperature, genus richness, and body size during the Phanerozoic. Statistical analyses, including the goodness-of-fit test, likelihood-ratio test, and CCDF, support these Zipf distributions. The systematic appearance of Zipf distributions in the PCCRs of these environmental variables and biological metrics, which characterize the evolutionary trajectories of the Phanerozoic Earth-life system, indicates that the frequency/probability of the occurrence of an extreme change is much smaller than that of a small change, implying that Earth's surface environments might have been stabilized and life's resilience to crises probably had been established across the Phanerozoic (Payne et al., 2020). Moreover, most of these four quantities are temporally correlated while their PCCRs are not (Fig. 4 and Table S2 in Supplementary data), indicating that the Zipf distributions in the PCCRs seem not to be directly connected to each other. What factors drove these uncorrelated PCCRs to systematically exhibit Zipf distributions and why the coefficients and exponents in these Zipf distributions take the specific values (Table 1) require future investigation. Furthermore, studies have attributed the Zipf distribution in a variety of natural systems on the modern Earth to self-organized criticality, a crucial characteristic of non-equilibrium systems possessing nonlinear dynamics (Bak et al., 1988; Schroeder, 2009; Bak, 2013). However, the existence of Zipf distributions is neither a necessary nor a sufficient condition for self-organized criticality (Solow, 2005; Touboul and Destexhe, 2017). Interpreting the fundamental mechanisms responsible for the Zipf distributions presented in this study needs further exploration.

Acknowledgements

I thank Editor Qun Yang for valuable suggestions and securing reviewers for the manuscript and Fernando Momo and an anonymous reviewer for thoughtful and constructive comments.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.palwor.2023.12.006.

References

Alroy, J., Aberhan, M., Bottjer, D.J., Foote, M., Fürsich, F.T., Harries, P. J., Hendy, A.J.W., Holland, S.M., Ivany, L.C., Kiessling, W., Kosnik, M.A., Marshall, C.R., McGowan, A.J., Miller, A.I., Olszewski, T.D., Patzkowsky, M.E., Peters, S.E., Villier, L., Wagner, P.J., Bonuso, N., Borkow, P.S., Brenneis, B., Clapham, M.E., Fall, L.M., Ferguson, C. A., Hanson, V.L., Krug, A.Z., Layou, K.M., Leckey, E.H., Nürnberg, S., Powers, C.M., Sessa, J.A., Simpson, C., Tomašových, A., Visaggi,

C.C., 1989. Phanerozoic trends in the global diversity of marine invertebrates. Science 321, 97–100.
Alstott, J., Bullmore, E., Plenz, D., 2014. Powerlaw: A python package for analysis of heavy-tailed distributions. PLoS ONE 9, e85777.
Anderson, T.W., 1962. On the distribution of the two-sample Cramer-von Mises criterion. The Annals of Mathematical Statistics 33 (3), 1148–1159.
Ausloos, M., Bronlet, P., 2003. Strategy for investments from Zipf law(s). Physica A: Statistical Mechanics and its Applications 324, 30–37.
Bak, P., 2013. How Nature Works: The Science of Self-Organized Criticality. Copernicus, New York, 212 pp.
Bak, P., Tang, C., Wiesenfeld, K., 1988. Self-organized criticality. Physical Review A 38 (1), 364–374.
Berner, R.A., 1990. Atmospheric carbon dioxide levels over Phanerozoic time. Science 249 (4975), 1382–1386.
Blount, Z.D., Lenski, R.E., Losos, J.B., 2018. Contingency and determinism in evolution: Replaying life's tape. Science 362 (6415), eaam5979.
Cancho, R.F.I., Solé, R.V., 2003. Least effort and the origins of scaling in human language. Proceedings of the National Academy of Sciences of the United States of America 100, 788–791.
Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012. Biodiversity loss and its impact on humanity. Nature 486, 59–67.
Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. SIAM Review 51, 661–703.
Corral, Á., Serra, I., Ferrer-i-Cancho, R., 2020. Distinct flavors of Zipf's law and its maximum likelihood fitting: Rank-size and size-distribution representations. Physical Review E 102, 052113.
Deutsch, C., Ferrel, A., Seibel, B., Portner, H.O., Huey, R.B., 2015. Climate change tightens a metabolic constraint on marine habitats. Science 348 (6239), 1132–1135.
Foster, G.L., Rohling, E.J., 2013. Relationship between sea level and climate forcing by CO_2 on geological timescales. Proceedings of the National Academy of Sciences of the United States of America 110, 1209–1214.
Foster, G.L., Royer, D.L., Lunt, D.J., 2017. Future climate forcing potentially without precedent in the last 420 million years. Nature Communications 8, 14845.
Fujiwara, Y., 2004. Zipf law in firms bankruptcy. Physica A: Statistical Mechanics and its Applications 337, 219–230.
Furusawa, C., Kaneko, K., 2003. Zipf's law in gene expression. Physical Review Letters 90, 088102.
Galvez, M.E., Fischer, W.W., Jaccard, S.L., Eglinton, T.I., 2020. Materials and pathways of the organic carbon cycle through time. Nature Geoscience 13 (8), 535–546.
Gerlach, M., Altmann, E.G., 2019. Testing statistical laws in complex systems. Physical Review Letters 122, 168301.
Gould, S.J., 1990. Wonderful Life: The Burgess Shale and the Nature of History. WW Norton & Company, New York, 352 pp.
Hayes, J.M., Waldbauer, J.R., 2006. The carbon cycle and associated redox processes through time. Proceedings of the Royal Society of London, Series B: Biological Sciences 361 (1470), 931–950.
Hooper, D.U., Adair, E.C., Cardinale, B.J., Byrnes, J.E.K., Hungate, B.A., Matulich, K.L., Gonzalez, A., Duffy, J.E., Gamfeldt, L., O'Connor, M.I., 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature 486, 105–108.
Heim, N.A., Knope, M.L., Schaal, E.K., Wang, S.C., Payne, J.L., 2015. Cope's rule in the evolution of marine animals. Science 347 (6224), 867–870.
Kocsis, Á.T., Reddin, C.J., Alroy, J., Kiessling, W., 2019. The R package divDyn for quantifying diversity dynamics using fossil sampling data. Methods in Ecology and Evolution 10, 735–743.
Massey Jr., F.J., 1951. The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association 46 (253), 68–78.
Mayhew, P.J., Jenkins, G.B., Benton, T.G., 2008. A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. Proceedings of the Royal Society of London, Series B: Biological Sciences 275 (1630), 47–53.
Mega, M.S., Allegrini, P., Grigolini, P., Latora, V., Palatella, L., Rapisarda, A., Vinciguerra, S., 2003. Power-law time distribution of large earthquakes. Physical Review Letters 90, 188501.
Mora, T., Bialek, W., 2011. Are biological systems poised at criticality? Journal of Statistical Physics 144, 268–302.
Mora, T., Walczak, A.M., Bialek, W., Callan, C.G., 2010. Maximum entropy models for antibody diversity. Proceedings of the National Academy of Sciences of the United States of America 107, 5405–5410.
Newman, M.E., 1996. Self-organized criticality, evolution and the fossil extinction record. Proceedings of the Royal Society of London, Series B: Biological Sciences 263 (1376), 1605–1610.
Payne, J.L., Bachan, A., Heim, N.A., Hull, P.M., Knope, M.L., 2020. The evolution of complex life and the stabilization of the Earth system. Interface Focus 10, 20190106.
Raup, D.M., Sepkoski, J.J., 1984. Periodicity of extinctions in the geologic past. Proceedings of the National Academy of Sciences of the United States of America 81 (3), 801–805.
Rohde, R.A., Muller, R.A., 2005. Cycles in fossil diversity. Nature 434 (7030), 208–210.
Sarmiento, J.L., 2006. Ocean Biogeochemical Dynamics. Princeton University Press, New Jersey, 528 pp.
Schroeder, M., 2009. Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise. Dover Publications, New York, 448 pp.
Scotese, C.R., Song, H., Mills, B.J., van der Meer, D.G., 2021. Phanerozoic paleotemperatures: The Earth's changing climate during the last 540 million years. Earth-Science Reviews 215, 103503.
Sepkoski, J.J., 2002. A compendium of fossil marine animal genera. Bulletins of American Paleontology 203, 1–560.
Sheridan, J.A., Bickford, D., 2011. Shrinking body size as an ecological response to climate change. Nature Climate Change 1 (8), 401–406.
Solow, A.R., 2005. Power laws without complexity. Ecology Letters 8, 361–363.
Song, H., Kemp, D.B., Tian, L., Chu, D., Song, H., Dai, X., 2021. Thresholds of temperature change for mass extinctions. Nature Communications 12, 4694.
Stanley, S., 1973. An explanation for Cope's Rule. Evolution 27 (1), 1–26.
Touboul, J., Destexhe, A., 2017. Power-law statistics and universal scaling in the absence of criticality. Physical Review E 95, 012413.
Trammer, J., 2005. Maximum body size in a radiating clade as a function of time. Evolution 59 (5), 941–947.
Tyrcha, J., Roudi, Y., Marsili, M., Hertz, J., 2013. The effect of nonstationarity on models inferred from neural data. Journal of Statistical Mechanics: Theory and Experiment 2013, P03005.
Velarde, C., Robledo, A., 2017. Rank distributions: Frequency vs. magnitude. PLoS ONE 12, e0186015.
Vérard, C., Veizer, J., 2019. On plate tectonics and ocean temperatures. Geology 47, 881–885.
Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307.
Zipf, G.K., 1949. Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Addison-Wesley, Cambridge, 573 pp.

Per-Capita Change Rates of the Phanerozoic Earth-Life System Exhibited Zipf Distributions
— Supplementary Information —

Haitao Shang[1] *

[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA 97403

*To whom correspondence should be addressed; E-mail: htshang.research@gmail.com.

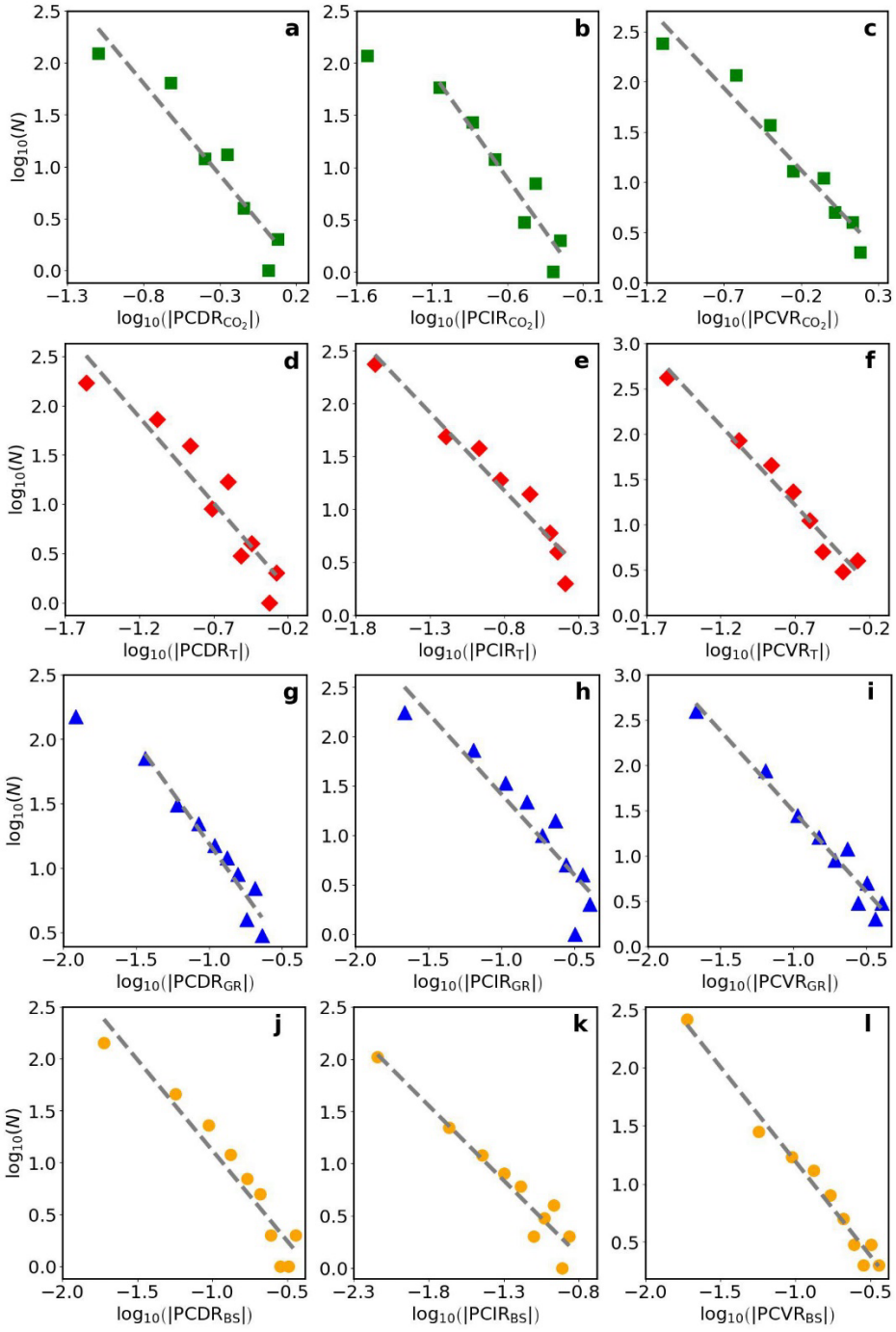


Fig. S1. Zipf distributions in the PCDRs, PCIRs, and PCVRs of (a–c) atmospheric CO_2 level (CO_2), (d–f) global average temperature (T), (g–i) genus richness (GR), and (j–l) body size (BS) in the Phanerozoic. The time step between two consecutive data points in the interpolated time series of these four quantities is 1 Myr (Section 2). Filled green squares, red diamonds, blue triangles, and orange circles represent the PCCRs versus counts (i.e., *N*'s) for the Phanerozoic atmospheric CO_2 level, global average temperature, genus richness, and body size, respectively. The initial small values in panels (b) and (g) that do not follow the Zipf distribution are not included in data fitting (Section 2). Gray dashed lines are the best-fitting Zipf distributions for the PCCRs of each dataset.

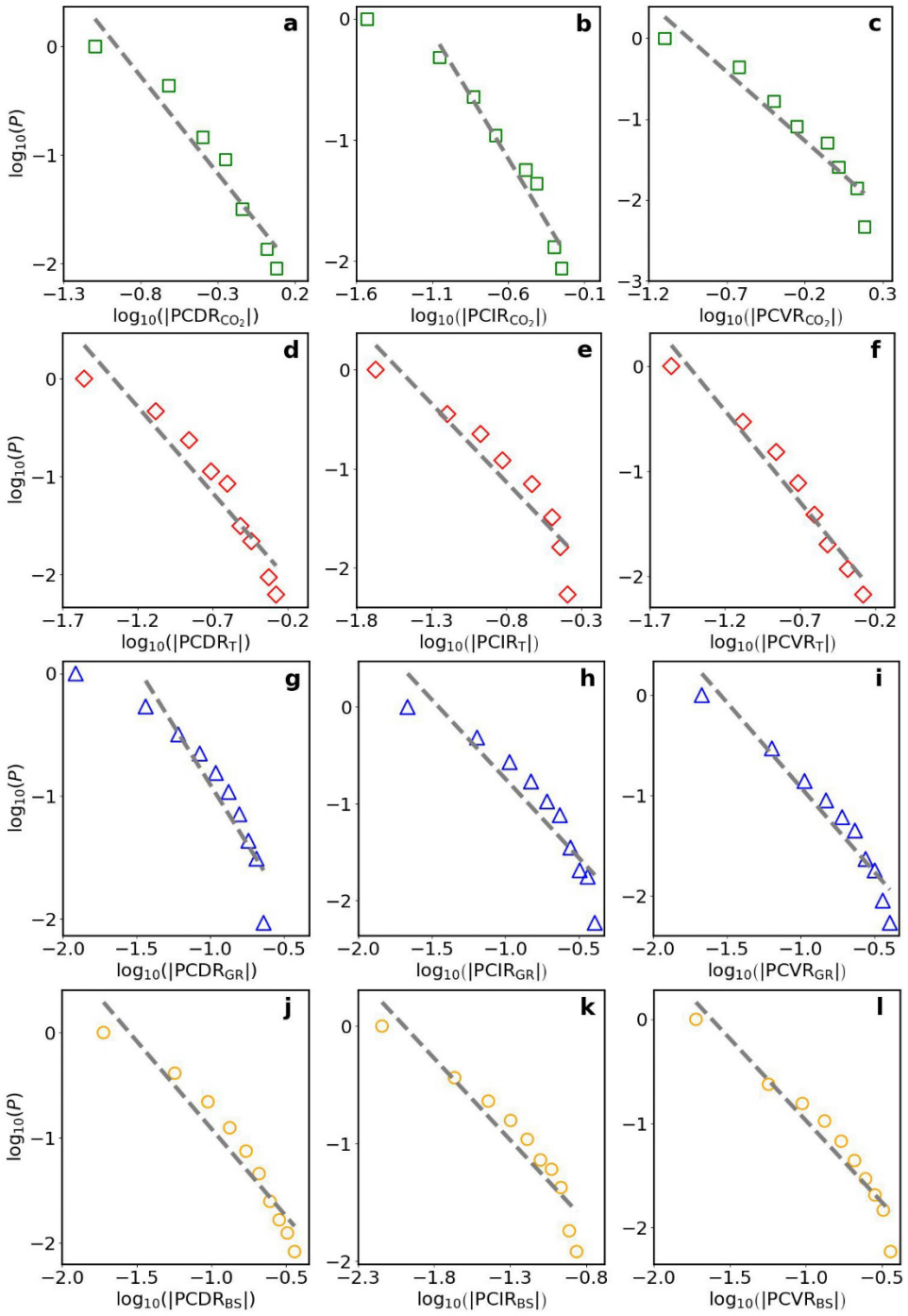


Fig. S2. Complementary cumulative distribution functions (CCDFs) *P* for the PCDRs, PCIRs, and PCVRs of (a–c) atmospheric CO_2 level (CO_2), (d–f) global average temperature (T), (g–i) genus richness (GR), and (j–l) body size (BS) in the Phanerozoic. The time step between two consecutive data points in the interpolated time series of these four quantities is 1 Myr (Section 2). Empty green squares, red diamonds, blue triangles, and orange circles represent the CCDFs for PCCRs of the Phanerozoic atmospheric CO_2 level, global average temperature, genus richness, and body size, respectively. The initial values in panels (b) and (g) that do not follow the Zipf distribution are not included in data fitting (Section 2). Gray dashed lines are the best-fitting Zipf distributions for the CCDFs of each dataset.

Table S1. Zipf distributions for the PCDRs, PDICs, and PCVRs of the Phanerozoic atmospheric CO_2 level (CO_2), global average temperature (T), genus richness (GR), and body size (BS). The time step between two consecutive data points in the interpolated time series of these four quantities is 1 Myr (Section 2). The R^2, RMSE, p_{KS}, and p_{CM} represent the coefficient of determination, root mean square error, the p-value of the KS test, and the p-value of the CM test, respectively, of the least-squares fitting on the log-log plots. The LR_{EXP}, $LR_{\mathrm{S\text{-}EXP}}$, and LR_{LN} represent the log-likelihood ratios for the Zipf distribution against exponential, stretched exponential, and lognormal distributions, respectively, for each dataset. The p_{EXP}, $p_{\mathrm{S-EXP}}$, and p_{LN} are the p-values for the statistical significance of these tests.

Category	**Quantity**	**Zipf distribution**	**Goodness-of-fit test**				**Likelihood-ratio test**					
			R^2	**RMSE**	p_{KS}	p_{CM}	LR_{EXP}	p_{EXP}	$LR_{\mathrm{S\text{-}EXP}}$	$p_{\mathrm{S-EXP}}$	LR_{LN}	p_{LN}
PCDR	Atmospheric CO_2 level	$N \sim \lvert\mathrm{PCDR}_{\mathrm{CO_2}}\rvert^{-1.77}$	0.97	0.23	0.96	0.92	5.93	2.08×10^{-05}	7.92	1.45×10^{-27}	12.93	5.63×10^{-03}
	Global average temperature	$N \sim \lvert\mathrm{PCDR}_{\mathrm{T}}\rvert^{-1.75}$	0.95	0.19	0.91	0.82	9.78	1.94×10^{-14}	1.89	8.52×10^{-05}	0.09	0.026
	Genus richness	$N \sim \lvert\mathrm{PCDR}_{\mathrm{GR}}\rvert^{-1.58}$	0.90	0.14	0.98	0.99	12.84	1.49×10^{-46}	3.25	6.29×10^{-27}	7.36	11.6710^{-16}
	Body size	$N \sim \lvert\mathrm{PCDR}_{\mathrm{BS}}\rvert^{-1.75}$	0.92	0.19	0.99	0.92	16.63	6.38×10^{-03}	10.83	6.11×10^{-21}	9.12	7.41×10^{-24}
PCIR	Atmospheric CO_2 level	$N \sim \lvert\mathrm{PCIR}_{\mathrm{CO_2}}\rvert^{-2.03}$	0.85	0.15	0.87	0.77	5.91	1.10×10^{-24}	5.91	7.46×10^{-12}	13.58	9.83×10^{-12}
	Global average temperature	$N \sim \lvert\mathrm{PCIR}_{\mathrm{T}}\rvert^{-1.58}$	0.94	0.24	0.94	0.89	9.05	2.83×10^{-08}	4.75	0.176	4.96	3.28×10^{-05}
	Genus richness	$N \sim \lvert\mathrm{PCIR}_{\mathrm{GR}}\rvert^{-1.64}$	0.85	0.24	0.98	0.92	19.56	4.67×10^{-36}	10.82	4.65×10^{-12}	22.98	5.92×10^{-07}
	Body size	$N \sim \lvert\mathrm{PCIR}_{\mathrm{BS}}\rvert^{-1.43}$	0.93	0.15	1.00	0.99	9.18	7.33×10^{-04}	5.36	0.017	5.87	4.58×10^{-04}
PCVR	Atmospheric CO_2 level	$N \sim \lvert\mathrm{PCVR}_{\mathrm{CO_2}}\rvert^{-1.63}$	0.94	0.19	0.93	0.91	8.21	5.92×10^{-08}	4.91	6.12×10^{-35}	2.94	8.76×10^{-14}
	Global average temperature	$N \sim \lvert\mathrm{PCVR}_{\mathrm{T}}\rvert^{-1.74}$	0.84	0.17	0.94	0.97	11.57	9.34×10^{-17}	14.77	1.79×10^{-04}	8.47	0.038
	Genus richness	$N \sim \lvert\mathrm{PCVR}_{\mathrm{GR}}\rvert^{-1.77}$	0.96	0.13	0.99	0.97	17.28	3.75×10^{-29}	6.53	1.35×10^{-18}	16.52	7.13×10^{-25}
	Body size	$N \sim \lvert\mathrm{PCVR}_{\mathrm{BS}}\rvert^{-1.63}$	0.98	0.09	1.00	0.99	4.05	4.83×10^{-85}	8.94	9.33×10^{-03}	6.78	0.354

Table S2. Summary of statistical analyses for each pair of quantities (including the Phanerozoic atmospheric CO_2 level (CO_2), global average temperature (T), genus richness (GR), and body size (BS)), of the per-capita change rates (PCCRs) of these quantities, and of the absolute values of these PCCRs. The time step between two consecutive data points in the interpolated time series of the four quantities is 0.1 Myr (Section 2). Reported statistics include the Pearson correlation coefficient (Pearson's r) with its p-value (p_r) and the Spearman's rank correlation coefficient (Spearman's ρ) with its p-value (p_ρ).

Quantity	**Quantity pair**	**Pearson's *r***	p_r	**Spearman's *ρ***	p_ρ
Environmental variables & biological metrics	CO_2 level vs. Temperature	0.33	$\ll 0.05$	0.43	$\ll 0.05$
	CO_2 level vs. Genus richness	-0.27	$\ll 0.05$	-0.25	$\ll 0.05$
	CO_2 level vs. Body size	0.001	0.83	0.01	0.64
	Temperature vs. Genus richness	-0.29	$\ll 0.05$	-0.25	$\ll 0.05$
	Temperature vs. Body size	0.06	$\ll 0.05$	0.08	$\ll 0.05$
	Genus richness vs. Body size	0.42	$\ll 0.05$	0.39	$\ll 0.05$
PCCRs of environmental variables & biological metrics	$PCCR_{CO_2}$ vs. $PCCR_T$	0.03	$\ll 0.05$	-0.01	$\ll 0.05$
	$PCCR_{CO_2}$ vs. $PCCR_{GR}$	0.002	0.69	0.03	0.21
	$PCCR_{CO_2}$ vs. $PCCR_{BS}$	-0.07	$\ll 0.05$	-0.04	$\ll 0.05$
	$PCCR_T$ vs. $PCCR_{GR}$	-0.09	$\ll 0.05$	-0.06	$\ll 0.05$
	$PCCR_T$ vs. $PCCR_{BS}$	0.04	0.27	0.02	0.67
	$PCCR_{GR}$ vs. $PCCR_{BS}$	0.01	0.002	0.02	$\ll 0.05$
Absolute PCCRs of environmental variables & biological metrics	$\lvert PCCR_{CO_2}\rvert$ vs. $\lvert PCCR_T\rvert$	-0.04	0.03	-0.06	0.01
	$\lvert PCCR_{CO_2}\rvert$ vs. $\lvert PCCR_{GR}\rvert$	0.04	$\ll 0.05$	0.01	0.07
	$\lvert PCCR_{CO_2}\rvert$ vs. $\lvert PCCR_{BS}\rvert$	-0.02	$\ll 0.05$	0.04	$\ll 0.05$
	$\lvert PCCR_T\rvert$ vs. $\lvert PCCR_{GR}\rvert$	0.12	$\ll 0.05$	0.07	0.12
	$\lvert PCCR_T\rvert$ vs. $\lvert PCCR_{BS}\rvert$	0.05	0.13	0.02	0.37
	$\lvert PCCR_{GR}\rvert$ vs. $\lvert PCCR_{BS}\rvert$	0.05	$\ll 0.05$	0.07	$\ll 0.05$