Supplement to "Punctuated non-equilibrium and niche conservatism explain biodiversity fluctuations through the Phanerozoic"

Andrew J. Rominger[1], Miguel A. Fuentes[1, 2, 3], and Pablo A. Marquet[1, 4, 5, 6, 7]

[1]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, US
[2]Instituto de Investigaciones Filosóficas, SADAF, CONICET, Bulnes 642, 1428 Buenos Aires, Argentin
[3]Facultad de Ingeniería y Tecnología, Universidad San Sebastián, Lota 2465, Santiago 7510157, Chile
[4]Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad de Chile, Alameda 340, Santiago, Chile
[5]Instituto de Ecología y Biodiversidad, Casilla 653, Santiago, Chile
[6]Laboratorio Internacional de Cambio Global (LINCGlobal), Pontificia Universidad Catlica de Chile, Alameda 340, Santiago, Chile
[7]Centro Cambio Global UC, Av. Vicuña Mackenna 4860, Campus San Vicuña, Santiago, Chile

1 Limit distribution of a time-averaged homogeneous origination-extinction process

Fossil taxa gain and lose taxa according to an origination-extinction process. We assume that most fossil occurrences of a taxon come from the period of its history when it is dominant and in steady state. In a time slice of duration τ during such a period of steady state the latent per capita rates of origination and extinction would be equal (i.e. $\lambda = \mu \equiv \rho$) and the number of origination or extinctions events (call such events Y) each follow an inhomogeneous Poisson process with rate ρN_t where N_t is the number of species or genera in the taxon of interest at time t. Allowing N_t to vary smoothly with time, and invoking the communicative property of the Poisson distribution, we arrive at the number Y of extinction *or* origination events in τ being distributed

$$Y \sim \mathrm{Pois}(\rho \int_{t=0}^{\tau} N(t)dt). \tag{1}$$

Under the steady state assumption we can approximate $N(t)$ by $\bar{N}$, the steady state diversity, leading to

$$Y \sim \mathrm{Pois}(\rho \bar{N} \tau). \tag{2}$$

Assuming the τ of each time period in the Paleobiology Database or Sepkoski's compendium to be approximately equal (i.e. equal durations of major stratographic units) then the distribution of fluctuations within taxa will be asymptotically Gaussian.

The Gaussian assymptotics of time-avereaged birth-death processes have been proven and explored elswhere as well[1,2].

2 Additional super-statistical analyses

To evaluate the sensitivity of our super-statistical analysis on the particular data used and we tested our predictions on different data sets (see below). The fact that it works in all different applications indicates that it is robust to vagaries of different recording strategies and bias corrections in paleobiology. This could mean that much of the raw signal in massive fossil datasets, at least signals regarding fluctuations, are not artifacts of sampling, as has been proposed before[3].

2.1 Raw PBDB data

We calculated the super-statistical prediction at the order level from raw genus diversity recorded in the PBDB without correcting for taphonomic or sampling bias (Fig. 3). The super-statistical calculation also closely fits the raw data as in the case of sampling and publication bias-corrected data.

2.2 Different taxonomic ranks in PBDB data

As noted in the main text, the super-statistical prediction predictably breaks down at higher taxonomic scales. In Figure 4 we present this worsening fit graphically using class level data with three-timer and publication corrected PBDB data

2.3 Sepkoski's compendium

Sepkoski's compendium[4] provided the first hypothesis of Phanerozoic diversification. As such, it has served as a benchmark for further investigation into large-scale paleobiological patterns[5]. We conducted the same super-statistical analysis as in the main text and find comparable results. Specifically, the super-statistical prediction far out preforms the null Gaussian model (Fig. 5) and worsens with increasing taxonomic scale (Fig. 5), again implying the uniqueness of orders.

References

1. Keilson, J. & Rao, S. S. A process with chain dependent growth rate. *Journal of Applied Probability* **7,** 699–711 (1970).
2. Grassmann, W. K. The asymptotic variance of a time average in a birth-death process. *Annals of Operations Research* **8,** 165–174 (1987).
3. Hannisdal, B. & Peters, S. E. Phanerozoic Earth System Evolution and Marine Biodiversity. *Science* **334,** 1121–1124 (2011).
4. Sepkoski, J. J. *A compendium of fossil marine animal families* (Milwaukee Public Museum, Milwaukee, WI, 1992).
5. Alroy, J. *et al.* Phanerozoic Trends in the Global Diversity of Marine Invertebrates. *Science* **321,** 97–100 (2008).
6. Alroy, J. The Shifting Balance of Diversity Among Major Marine Animal Groups. *Science* **329,** 1191–1194 (2010).

Supplemental Figures

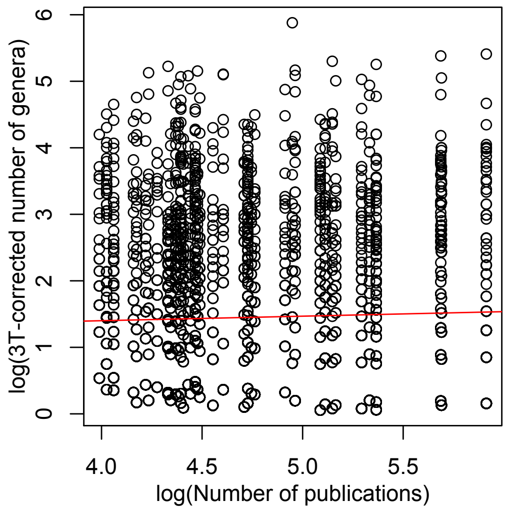

Figure 1: Relationship between number of publications and genus diversity as recorded by the PBDB.

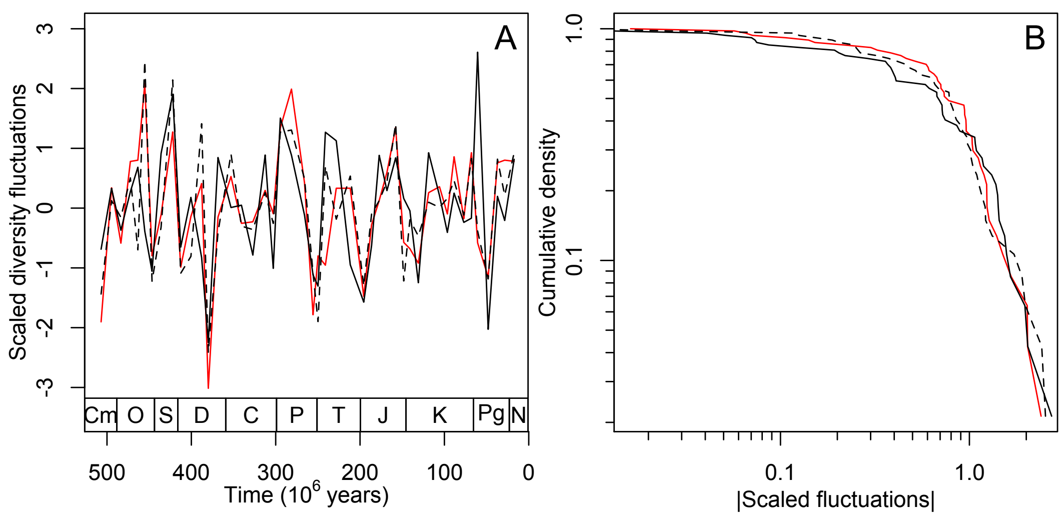

Figure 2: Comparison of SQS method[6] (solid black line) with the raw data (dashed black) and our three-timer and publication bias correction method (red). The time-series of all marine invertebrate genera shows general agreement with the only major deviations toward the modern (A). Despite these differences the distribution of fluctuations in genus diversity across all marine invertebrates show good agreement (B).

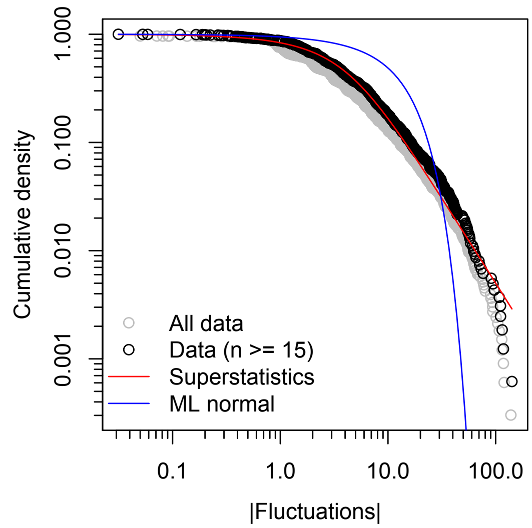

Figure 3: Super-statistical prediction of raw (i.e. not bias corrected) order-level fluctuations in genus diversity recorded in the PBDB. Grey dots are the full data of orders, while black ones are orders with more than 15 points. The red line is our theoretical prediction and the blue line the best Gaussian fit to the data.

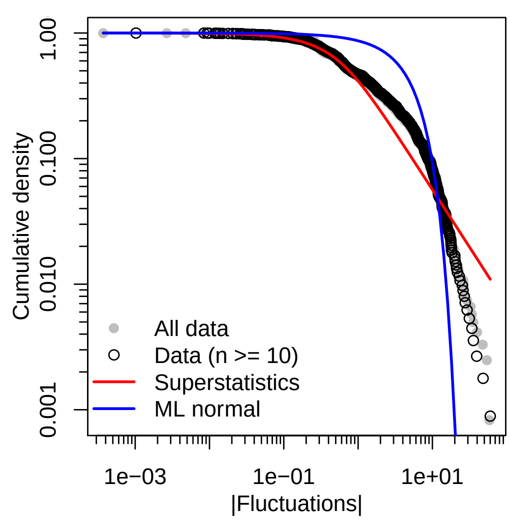

Figure 4: Super-statistical prediction of bias corrected class-level fluctuations in genus diversity recorded in the PBDB. Grey dots are the full data of orders, while black ones are orders with more than 15 points. The red line is our theoretical prediction and the blue line the best Gaussian fit to the data. Note at the class level the fit is predictably worse, see main text for discussion.

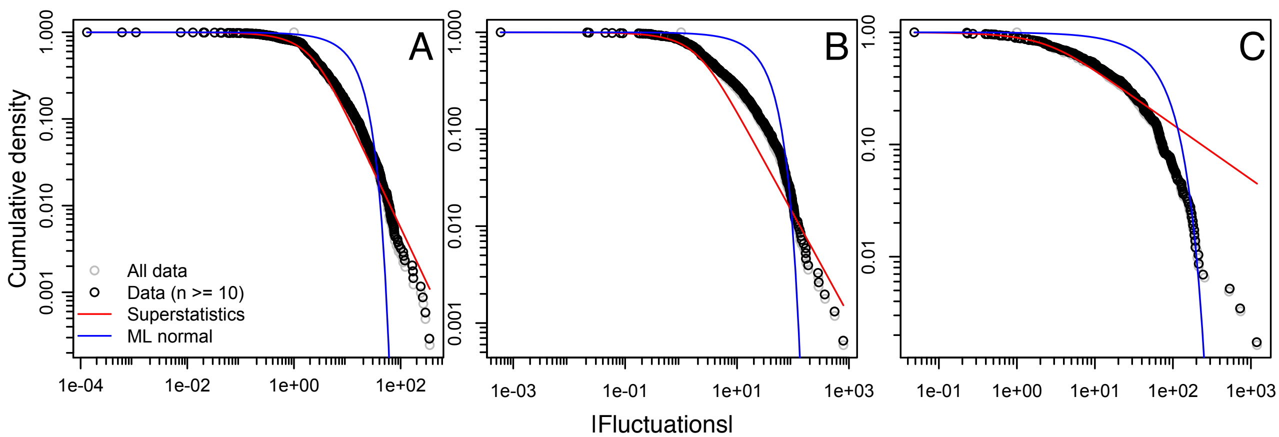

Figure 5: Super-statistical prediction (red line) of fluctuations in genus diversity recorded in Sepkoski's compendium of marine invertebrates compared to maximum likelihood normal distribution (blue line). Super-statistical theory explains order level fluctuations well (A) with increasingly poorer fits at the class (B) and phylum (C) levels.