The R package divDyn for quantifying diversity dynamics using fossil sampling data

Kocsis, Á. T.[a,b,1], **Reddin, C. J.**[a], **Alroy, J.**[c], **Kiessling, W.**[a]

[a]*GeoZentrum Nordbayern, Universität Erlangen-Nürnberg, Loewenichstraße 28, D-91054 Erlangen, Germany*

[b]*MTA-MTM-ELTE Research Group for Paleontology, POB 137, H-1431 Budapest, Hungary*

[c]*Department of Biological Sciences, Macquarie University, Sydney NSW 2109, Australia*

[1]*Correspondence: Á. T. Kocsis, GeoZentrum Nordbayern, Universität Erlangen-Nürnberg, Loewenichstraße 28, D-91054 Erlangen, Germany. E-mail:* adam.kocsis@fau.de

1 SUMMARY

2 1. Unbiased time series of diversity dynamics are vital for quantifying the grand
3 history of life. Applications include identifying ancient mass extinctions and
4 inferring both biotic and abiotic controls on diversification rates.

5 2. We introduce divDyn, a new R package that facilitates the calculation of
6 taxonomic richness, extinction and origination rates from time-binned fossil
7 sampling data. State-of-the-art sampling completeness metrics, counting
8 protocols, and sampling standardisation functions permit the reconstruction
9 of biologically meaningful time series. Additional functions permit the
10 partitioning of turnover rates by trait and environmental affinity.

11 3. We display Phanerozoic-scale diversity dynamics of marine invertebrates
12 using the divDyn package. Using the core function and standard
13 subsampling options, we revisit earlier assessments of declining extinction
14 and origination rates over time and of equilibrial diversity dynamics to
15 assess their methodological dependency.

16 4. The modular and fast implementation of published methods ensures
17 traceability, reproducibility, and comparability of future studies.

Key-words: diversity dynamics, fossil record, Paleobiology Database, extinction, origination, subsampling

1 INTRODUCTION

2   Reconstructions of global diversity (Sepkoski *et al.* 1981; Alroy *et al.* 2001; Alroy *et al.* 2008; Alroy
3   2010a) as well as extinction and origination rates (Newell 1952; Carr & Kitchell 1980; Raup & Sepkoski
4   1982; Sepkoski 1993; Benton 1995; Foote 2000; Bambach, Knoll & Wang 2004; Alroy 2008; Alroy 2014;
5   Alroy 2015) have led to the recognition of ancient mass extinctions (Raup & Sepkoski 1982; Alroy 2008)
6   and major insights about the interplay of evolutionary crisis and recovery. All of these results depend on
7   the reconstruction of time series that characterise changes in the global biosphere. Calculating these time
8   series from an incomplete fossil record is a fundamental task, as they serve as the basis for the statistical
9   testing of grand questions in macroevolution.

10   Early studies largely relied on compendia of stratigraphic ranges, that is, they derived diversity metrics
11   from overlapping durations of taxa (e.g. Sepkoski 1984). Since the advent of the Paleobiology Database
12   (PaleoDB, https://paleobiodb.org) diversity dynamics have largely been inferred from large occurrence
13   datasets that incorporate hundreds of thousands of items. Occurrence data allow for alternative counting
14   methods and sampling standardisation, but implementing these methods in scripting languages is time-
15   consuming and can be challenging for students and researchers with little programming experience.
16   The algorithmic implementation of some procedures and the multiple steps of data filtering also permit
17   considerable analytical freedom, which potentially compromises the comparability and traceability of
18   results. Using a standardised toolkit will facilitate a fast and consistent workflow and allows researchers to
19   focus on scientific questions rather than losing time with repeated implementation of established methods
20   in computer scripts.

21   Here we present the R (2018) package divDyn, which facilitates the calculation of diversity dynamics
22   from fossil occurrence datasets along with additional methods of palaeobiological inference. Our purpose
23   is to establish a transparent, traceable and modular workflow from data acquisition to the calculation
24   of biologically meaningful diversity metrics. The primary application of the package is expected to be
25   for data from the Paleobiology Database (PaleoDB, https://palebiodb.org/), the largest fossil occurrence
26   dataset, which serves as a standard for palaeontological data analyses. However, any dataset for which
27   diversity metrics are to be assessed in temporal or spatial intervals can also be processed in divDyn. The
28   PaleoDB, Neptune (Lazarus 1994) and AMMON (Korn & Ilg 2007) databases are good examples of
29   alternative datasets. To demonstrate the advantages of divDyn, we revisit results from an earlier study on
30   Phanerozoic-scale diversity dynamics (Alroy 2008).

1   FEATURES

2   CALCULATION OF TIME SERIES

3   To prepare an occurrence dataset for analysis in divDyn, it must be formatted as a table with each row

4   representing a single occurrence of a taxon. Occurrences should be assigned to discrete time intervals

5   (bins). Alternatively, values in a continuous time-related dimension (e.g. years before present, or meters in

6   a section) can be used that will be translated to discreet bins by the package.

7   The core function `divDyn()` calculates taxonomic richness, extinction and origination rate estimates

8   using the presence-absence patterns of the bin-taxon matrix implied by the input dataset. It calculates a

9   large suite of diversity metrics and rates in one go, ranging from classical sampled-in-bin richness to the

10   recent second-for-third proportions (Alroy 2015, Table 1).

11   SAMPLING STANDARDISATION PROCEDURES

12   Sampling standardisation is a useful tool for smoothing out the effects of changing sampling intensity

13   on the available data. Subsampling, the most common group of methods in this field of research, uses

14   interpolation to answer the general question 'What would the results look like if fewer data were

15   available?'. The oldest and most widely used subsampling method is rarefaction (Sanders 1968; Miller &

16   Foote 1996) but there are many others (Alroy *et al.* 2001; Bush, Markey & Marshall 2004; Alroy 2010a;

17   Chao & Jost 2012). More implementations of these methods are available for estimating richness from

18   an incomplete sample (Hammer, Harper & Ryan 2001; Hsieh, Ma & Chao 2016), but their application is

19   more complicated in the context of time series reconstruction (Fig. 1), where information combined from

20   multiple samples is to be extracted.

21   We formalized the subsampling process in the wrapper function `subsample()`. The arguments of

22   the process are D, g, f, K, α_f and α_g, where D is the source dataset, g is the subsampling function, f is the

23   user-supplied function, K is the number of iterations, and α_f and α_g are sets of additional function argu-

24   ments for f and g, respectively. The desired result, based on all sampled data, is just a function of the data

25   and its own arguments

$$R = f(D, \alpha_f) \tag{1}$$

27   Given that the dataset is divisible into finite subsets, such as time bin-specific parts

$$D := \{D^1, D^2, ... D^U\} \tag{2}$$

29   each bin-specific part has an abstract subset that features the same desired sampling characteristics (for

30   instance, intensity), thus defining a set of data

$$^{sub}D := \{g(D^1, \alpha_g), g(D^2, \alpha_g), ..., g(D^u, \alpha_g)\} \tag{3}$$

32   With this ^{sub}D abstract representation of the data, the desired results (^{sub}R) would be just

1
$$^{sub}R = f(^{sub}D, \alpha_f) \tag{4}$$

2　　This can be approximated with Monte Carlo simulations, by generating multiple, randomly subsampled

3　instances of $^{sub}D_k$ (subsampled dataset in trial k out of K) and allowing the naturally emerging variation

4　to propagate to $^{sub}R_k$ (subsampled result in trial k). Thus, the simulation will lead to a vector of possible

5　results.

6　　　　From the point of implementation, f can be any function that is applicable to the original occurrence

7　dataset D and the class of $^{sub}R_k$ can be of any sort, such as vectors (time series) and complex structures, if

8　they can be vectorized in the form of a list-class R object. In the case of primitive trial results (vectors and

9　matrices), the averaging of time series can be automated. The current package version features Classical

10　Rarefaction (Sanders 1968; Miller & Foote 1996), Occurrence-Weighted By-List Subsampling (O^{x}W,

11　Alroy *et al.* 2001), and Shareholder Quorum Subsampling (SQS, Alroy 2010a).

12　**ADDITIONAL FUNCTIONALITY**

13　　To facilitate the stratigraphic assignment of collections, we compiled tables using the dynamic timescale

14　interpreter of Fossilworks (http://fossilworks.org/) that links entries to major geochronological intervals of

15　two predefined timescales (the level of geological stages and the 10-million-year time scale). The package

16　also includes additional tables to categorise the downloaded occurrences in terms of bathymetry (shallow,

17　deep), substrate (siliciclastic vs. carbonate) and reef vs. non-reef environments.

18　　Occurrence patterns of taxa allow the inference of preferred environments, which we can use to

19　compare diversity dynamics between groups with different affinities such as tropical vs. extra-tropical

20　or reef vs. non-reef (Alroy 2010a; Kiessling, Simpson & Foote 2010). As fossil sampling is biased and

21　heterogeneous, we implemented methods to calculate likely environmental preferences with sampling

22　patterns taken explicitly into account (Kiessling & Aberhan 2007), as well as test for selectivity of

23　extinctions (Kiessling & Kocsis 2015).

24　　The `divDyn()` function creates output in discretised time intervals. Results can be visualized effectively

25　with the additional plotting functions that display the used time scale (`tsplot()`), stratigraphic ranges

26　(`ranges()`), changes in composition (`parts()`) or distributions of values in time series (`shades()`).

27　　The basic functionality of the package is elaborated in the accompanying vignette (Handout to the R

28　package divDyn v0.6.0 for diversity dynamics from fossil occurrence data), with an example dataset of

29　Scleractinian corals used by Kiessling and Kocsis (2015).

1 EXAMPLE APPLICATION: PHANEROZOIC-SCALE DIVERSITY
2 DYNAMICS OF MARINE ANIMALS

3 Tracing diversity through the entire Phanerozoic (the last 541 million years of Earth history) has been
4 the focus of palaeobiological research ever since the first global diversity curves were published (Newell
5 1952). Some patterns have been surprisingly robust to scientific scrutiny, whereas others are more volatile.
6 For example, the temporal decline of turnover rates has largely gone unchallenged since its first observation
7 (Raup & Sepkoski 1982), whereas the original 'Big Five' mass extinction events of Raup and Sepkoski
8 (1982) have been repeatedly revisited, with different conclusions (Bambach, Knoll & Wang 2004; Alroy
9 2008). Much discussion has focused on the dramatic rise of marine biodiversity over the last 100 myr,
10 which is evident in older compilations (Valentine 1970; Sepkoski 1993) but much less so in sampling-
11 standardised analyses (Alroy *et al.* 2008). Not yet formally contested are Alroy's (2008, 2010b) analyses
12 of the temporal relationship between diversity and rates. If these results are robust at different temporal
13 resolutions, they strongly argue for equilibrial, diversity-dependent diversity dynamics (Sepkoski 1978;
14 Sepkoski 1984; Alroy 1996) through biotic interactions. With the continuous expansion of both fossil
15 occurrence data sets and the toolkit to analyse them, it is necessary to re-evaluate such scientific outcomes
16 on a periodical basis. The objective of this case study is to assess the robustness of previous results in
17 the face of the increase in the number of fossil occurrences and the number of choices we face when we
18 express diversity dynamics over deep time.

19 DATA PROCESSING AND APPLIED METHODS

20 The analyses presented in this section can be reproduced with the second vignette accompanying
21 the package (Phanerozoic-scale global marine biodiversity analysis with the R package divDyn v0.6;
22 http://github.com/adamkocsis/ddPhanero). The occurrences used here were downloaded on 14 September
23 2018, including all entries between the Ediacaran-Holocene interval. The data were filtered to marine taxa
24 and binned to geological stages as well the often used 10 myr bins (Alroy et al. 2008). As the procedural
25 treatment of stages in the Cambrian and Ordovician systems was influenced by considerable stratigraphic
26 error, they were resolved using biozone and formation entries (Ordovician), and with data from previous
27 analyses (Cambrian, Na & Kiessling 2015). In keeping with related literature (e.g. Sepkoski 1997), all
28 analyses were carried out at the genus level. However, species-level analyses can be conducted with
29 exactly the same procedures.

30 We computed diversity dynamics at both stratigraphic resolutions (stages and 10 myr), with three
31 different treatments of the data (raw, CR and SQS). Four different rate metrics were applied: per capita
32 rates (Foote 1999, used most frequently in studies), corrected three-timer rates (Alroy 2008), gap-filler
33 equations (Alroy 2014), and second-for-third substitution rates of Alroy (2015). This resulted in 24
34 different sets (2 timescales × 3 data treatments × 4 rate metrics) of richness, origination and extinction
35 rate series, each affected in a different way by the distorting effects of incomplete, heterogeneous sampling

1   and estimation error. Simple correlation and normality tests were applied to each of those sets to check

2   whether previous results are robust to varying protocols.

3   As indicated by Foote (2005), most taxonomic turnover was probably pulsed, likely at stage

4   boundaries. This assertion is supported by the infrequent correlations between interval durations and rate

5   values when the time dimension is excluded from the rate equations (Table 2 and Alroy 2008). For the

6   analysis of distributions, outliers and cross correlations, we detrended the rates and the richness by applying

7   LOESS with AICc-based smoothing parameters (Wang 2010) to describe long-term variation (Bambach,

8   Knoll & Wang 2004). Mass extinctions are defined as statistical outliers (using boxplot statistics) after the

9   LOESS trend has been removed.

RESULTS

11   The first order patterns of the time series acquired with the different methods match very well,

12   but the estimates for the individual time slices (Fig. 2) and thus the derived conclusions (Table 2) vary

13   considerably. Although the decline of extinction and origination rates can be confirmed if the Cambrian

14   and Ordovician is included in the dataset, the rates are unlikely to have featured a solid decline after the

15   Ordovician period. All detrended extinction-rate series feature the latest Permian value as a mass extinction,

16   which is consistently identified as the highest value in the series. The number of mass extinctions vary

17   considerably. Among the 'Big Five' mass extinctions (Raup & Sepkoski 1982), the end-Triassic and the

18   end-Cretaceous values also show up as outliers, as was indicated by Alroy (2008). The cross-correlations

19   that indicate equilibrial dynamics appear to be unstable and are likely affected by the estimation error of

20   the values in the time series.

DISCUSSION

22   Most major results derived from the Phanerozoic-scale analyses of the marine fossil animal record

23   are robust after the additional 10 years of data entry and methodological development since Alroy (2008).

24   Sepkoski (1993) found a similar pattern in analysing an older, range-based Phanerozoic diversity data set.

25   One notable exception is the decline of rates through time, which receives varying support if the first two

26   Phanerozoic periods are excluded. Likewise, changing data treatments either question equilibrial diversity

27   dynamics or point to the possibility that carrying capacities change through time (Sepkoski 1984; Alroy

28   2010b; Marshall & Quental 2016).

29   Whichever the case, using a standard toolkit like divDyn enhances our ability to reproduce previous

30   results and test the effect of added data or changing temporal resolution. We hope that our package will

31   spur large-scale diversity analyses beyond the still small group of trained peers, such that results can be

32   incorporated into broader evolutionary questions (Jablonski & Shubin 2015).

33   We intend to expand the set of output variables in the future, and the flexibility to write custom

1 subsampling functions will be incorporated into divDyn. Our package utilises methods developed mostly

2 by palaeobiologists. Simulations are currently being developed to compare results with other approach-

3 es such as PyRate (Silvestro, Salamin & Schnitzler 2014) or capture-mark-recapture methods (Liow &

4 Nichols 2010).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

12 ATK conceived the project, wrote the first manuscript draft and the software using code from WK and

13 JA as foundations. WK and CJR contributed to testing, interface and feature development, as well as the

14 debugging of code. All authors contributed to writing the manuscript.

DATA ACCESSIBILITY

16 The package is accessible from its GitHub repository (http://github.com/adamkocsis/divDyn) and

17 has been submitted to the CRAN servers. The occurrence data used here are freely available from the

18 Paleobiology Database. All files needed to reproduce the specific example are available on GitHub

19 (http://github.com/adamkocsis/ddPhanero).

REFERENCES

Alroy, J. (1996) Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **127,** 285-311.

Alroy, J. (2008) Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Science,* **105,** 11536-11542.

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

Alroy, J. (2010b) Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology,* **53,** 1211-1235.

Alroy, J. (2014) Accurate and precise estimates of origination and extinction rates. *Paleobiology,* **40,** 374-397.

Alroy, J. (2015) A more precise speciation and extinction rate estimator. *Paleobiology,* **41,** 633–639.

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. (2008) Phanerozoic trends in the global diversity of marine invertebrates. *Science,* **321,** 97-100.

Alroy, J., Marshall, C.R., Bambach, R.K., Bezusko, K., Foote, M., Fürsich, F.T., Hansen, T.A., Holland, S.M., Ivanyi, L.C., Jablonski, D., Jacobs, D.K., Jones, D.C., Kosnik, M.A., Lidgard, S., Low, S., Miller, A.I., Novack-Gottshall, P.M., Olszewski, T.D., Patzkowsky, M.E., Raup, D.M., Roy, K., Sepkoski, J.J.J., Sommers, M.G., Wagner, P.J. & Webber, A. (2001) Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proceedings of the National Academy of Science,* **98,** 6261-6266.

Bambach, R.K., Knoll, A.H. & Wang, S.C. (2004) Origination, extinction, and mass depletions of marine diversity. *Paleobiology,* **30,** 522-542.

Benton, M.J. (1995) Diversification and extinction in the history of life. *Science,* **268,** 52-58.

Bush, A.M., Markey, M.J. & Marshall, C.R. (2004) Removing bias from diversity curves: the effects of spatially organized biodiversity on sampling standardization. *Paleobiology,* **30,** 666-686.

Carr, T.R. & Kitchell, J.A. (1980) Dynamics of taxonomic diversity. *Paleobiology,* **6,** 427-443.

Chao, A. & Jost, L. (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology,* **93,** 2533-2547.

1 Foote, M. (1999) Morphological diversity in the evolutionary radiation of Paleozoic and post-Paleozoic
2   crinoids. *Paleobiology,* **25,** 1-115.

3 Foote, M. (2000) Origination and extinction components of taxonomic diversity: General problems. *Pa-*
4   *leobiology,* **26,** 74-102.

5 Foote, M. (2005) Pulsed origination and extinction in the marine realm. *Paleobiology,* **31,** 6-20.

6 Foote, M. & Miller, A.I. (2007) *Principles of Paleontology.* W. H. Freeman and Company, New York.

7 Hammer, O., Harper, D.A.T. & Ryan, P.D. (2001) PAST: Paleontological statistics software package for
8   education and data analsis. *Paleontologica Electronica,* **4,** 9.

9 Hsieh, T.C., Ma, K.H. & Chao, A. (2016) iNEXT: an R package for rarefaction and extrapolation of spe-
10   cies diversity (Hill numbers). *Methods in Ecology and Evolution,* **7,** 1451-1456.

11 Jablonski, D. & Shubin, N.H. (2015) The future of the fossil record: Paleontology in the 21st century.
12   *Proceedings of the National Academy of Sciences,* **112,** 4852-4858.

13 Kiessling, W. & Aberhan, M. (2007) Environmental determinants of marine benthic biodiversity dynamics
14   through Triassic-Jurassic time. *Paleobiology,* **33,** 414-434.

15 Kiessling, W. & Kocsis, Á.T. (2015) Biodiversity dynamics and environmental occupancy of fossil
16   azooxanthellate and zooxanthellate scleractinian corals. *Paleobiology,* **41,** 402-414.

17 Kiessling, W., Simpson, C. & Foote, M. (2010) Reefs as cradles of evolution and sources of biodiversity
18   in the Phanerozoic. *Science,* **327,** 196-198.

19 Korn, D. & Ilg, A. (2007) AMMON. http://www.wahre-staerke.com/ammon/

20 Lazarus, D. (1994) Neptune: A marine micropaleontology database. *Mathematical Geology,* **26,** 817-832.

21 Liow, L.H. & Nichols, J.D. (2010) Estimating rates and probabilities of origination and extinction using
22   taxonomic occurrence data: capture-mark-recapture (CMR) approaches. *Quantitative Methods in*
23   *Paleobiology* (eds J. Alroy & G. Hunt), pp. 81-94. The Paleontological Society.

24 Marshall, C.R. & Quental, T.B. (2016) The uncertain role of diversity dependence in species diversifi-
25   cation and the need to incorporate time-varying carrying capacities. *Phil. Trans. R. Soc. B,* **371,**
26   20150217.

27 Miller, A.I. & Foote, M. (1996) Calibrating the Ordovician radiation of marine lfe: Implications for Pha-
28   nerozoic diversity trends. *Paleobiology,* **22,** 304-309.

29 Na, L. & Kiessling, W. (2015) Diversity partitioning during the Cambrian radiation. *Proceedings of the*
30   *National Academy of Sciences,* **112,** 4702-4706.

31 Newell, N.D. (1952) Periodicity in invertebrate evolution. *Journal of Paleontology,* **26,** 371-385.

1 R Development Core Team (2018) R: A language and environment for statistical computing. R Foundation
2   for Statistical Computing, Vienna, Austria.

3 Raup, D.M. & Sepkoski, J.J.J. (1982) Mass extinctions in the marine fossil record. *Science,* **215,** 1501-
4   1503.

4 Sanders, H.L. (1968) Marine Benthic Diversity: A Comparative Study. *The American Naturalist,* **102,**
5   243-282.

6 Sepkoski, J.J., Bambach, R.K., Raup, D.M. & Valentine, J.W. (1981) Phanerozoic marine diversity and the
7   fossil record. *Nature,* **293,** 435-437.

8 Sepkoski, J.J., Jr. (1978) A Kinetic Model of Phanerozoic Taxonomic Diversity I. Analysis of Marine Or-
9   ders. *Paleobiology,* **4,** 223-251.

10 Sepkoski, J.J., Jr. (1984) A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families
11   and mass extinctions. *Paleobiology,* **10,** 246-267.

12 Sepkoski, J.J.J. (1993) Ten years in the library: new data confirm paleontological patterns. *Paleobiology,*
13   **19,** 43-51.

14 Sepkoski, J.J.J. (1997) Biodiversity: past, present, and future. *Journal of Paleontology,* **71,** 533-539.

15 Silvestro, D., Salamin, N. & Schnitzler, J. (2014) PyRate: a new program to estimate speciation and extin-
16   ction rates from incomplete fossil data. *Methods in Ecology and Evolution,* n/a-n/a.

17 Valentine, J.W. (1970) How many marine invertebrate fossil species? A new approximation. *Journal of*
18   *Paleontology,* **44,** 410-415.

19 Wang, X.-F. (2010) *fANCOVA: Nonparametric Analysis of Covariance*.

FIGURES

(a)

(b)

(c)
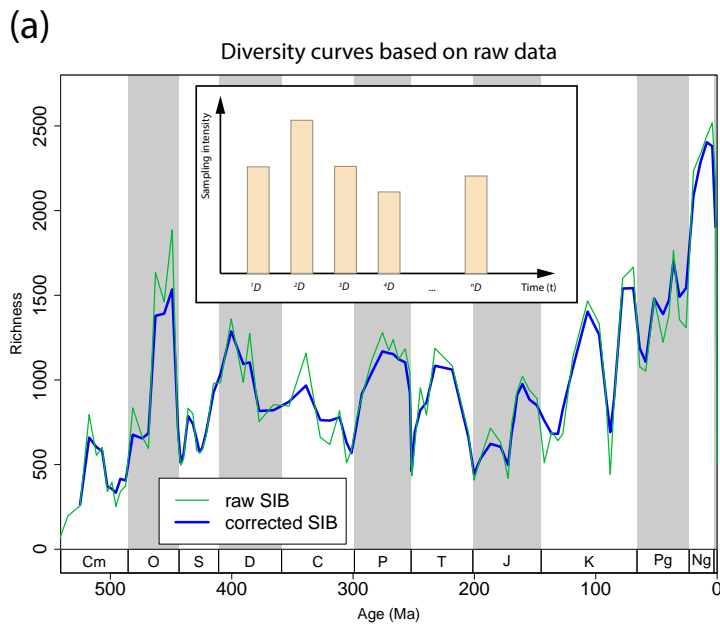
1 Figure 1. Demonstration of the procedures
2 implemented in the subsampling wrapper
3 function. (a) Calculation of raw results, (b)
4 calculation of results from a single subsampling
5 trial and (c) multiple trial results and averaging.
6 The curves show genus richness from the
7 Phanerozoic example dataset, standardized
8 with SQS at the stratigraphic resolution of
9 geologic stages. The number of iterations (K)
10 was 100, the quorum for SQS is 0.7.
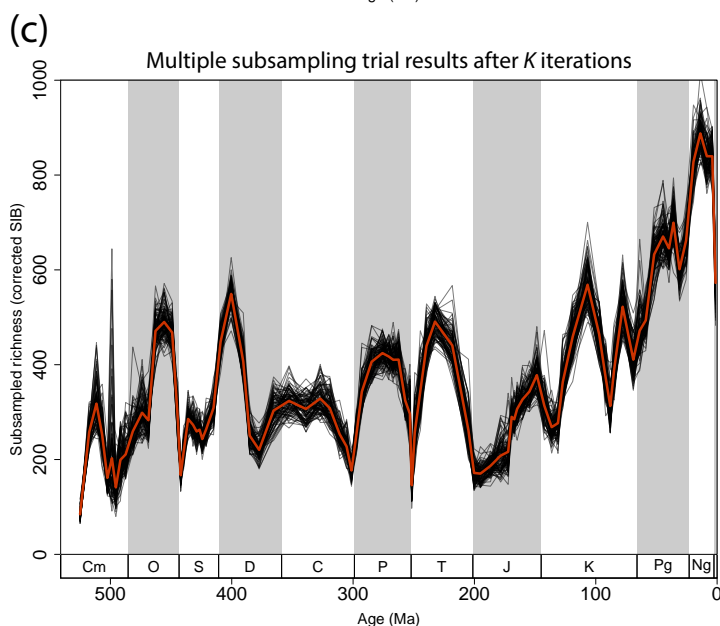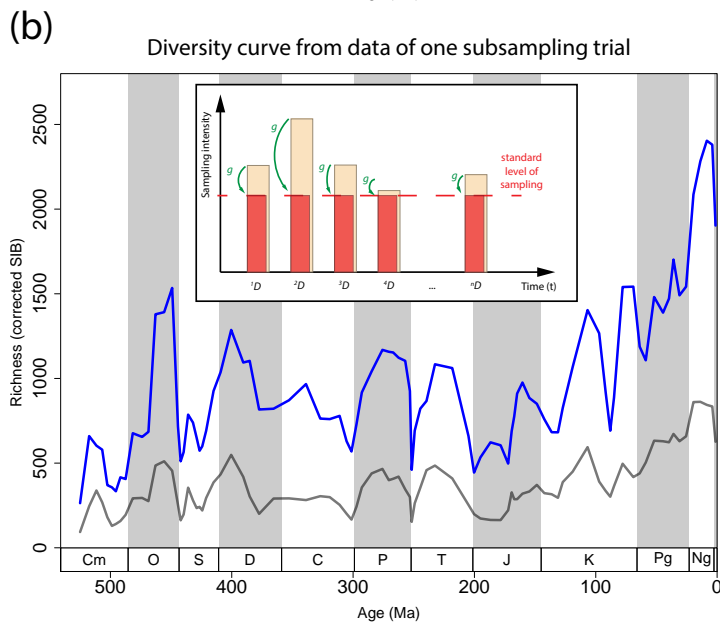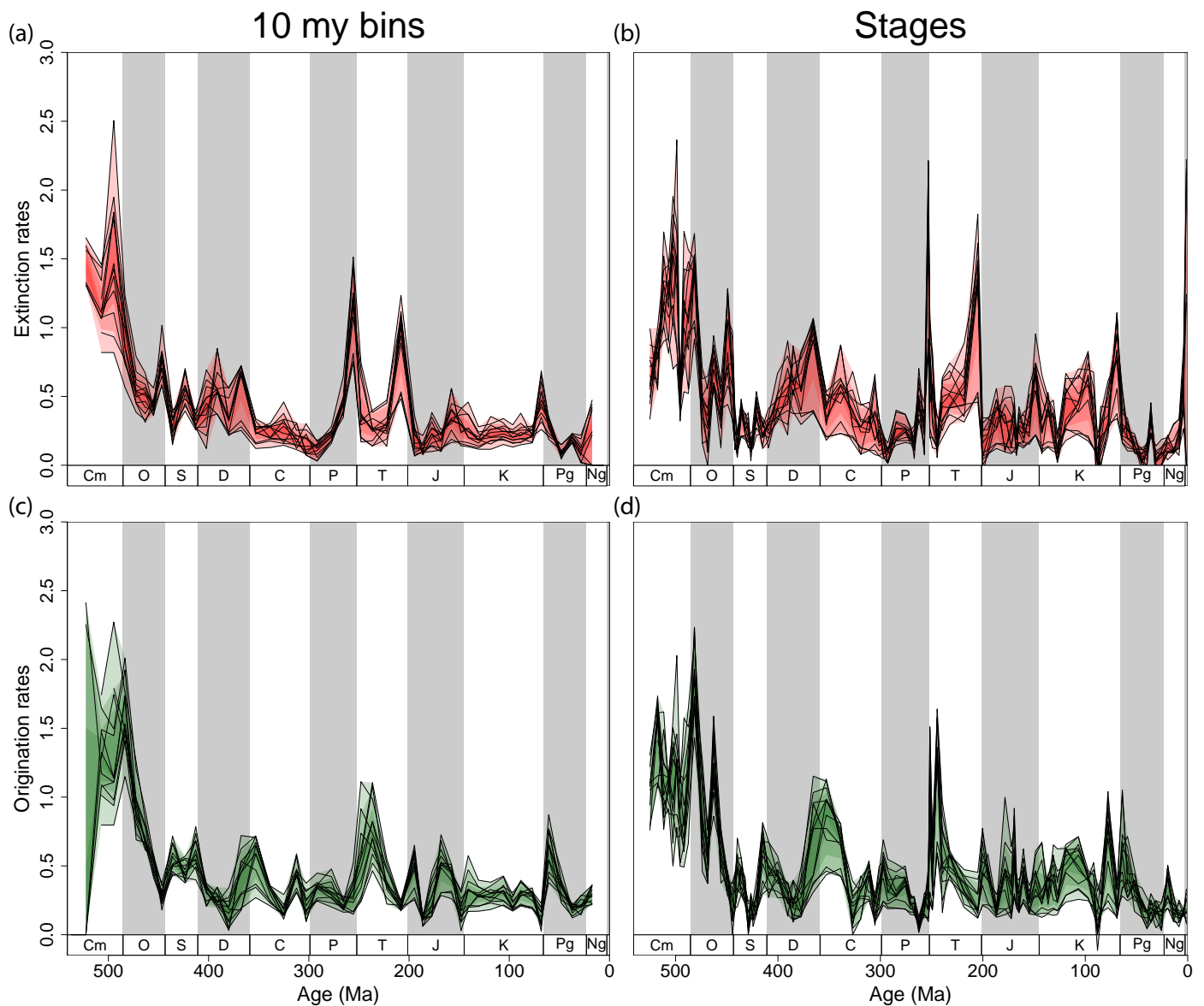
Figure 2. Genus-level Phanerozoic-scale extinction (a-b) and origination rates (c-d) calculated at the level of 10

million year bins and stratigraphic stages. Each panel features twelve (three treatment × four rate metrics) curves,

either raw, CR (quotas are 4,800 for the 10my bins and 1,100 for the stages) or SQS-standardized (quorum is 0.7

for both) average per capita rates, corrected three-timer rates, gap-filler rates or second-for-third substitution rates.

TABLES

1   Table 1. List of some of the variables output by the main function of the package.

Variable name	Metric name	Type	Reference
`tSing, tOri, tExt, tThrough`	range-based taxon patterns	counts	Foote, 1999
`t2d, t2u, t3, tPart, tGFu, tGFd`	occurrence-based taxon patterns	counts	Alroy, 2008; Alroy 2014
`extProp, oriProp`	proportional extinctions and originations	turnover	Newell, 1952
`extPC, oriPC`	per capita extinction and origination rates	turnover	Foote, 1999 (Alroy, 1996)
`ext3t, ori3t`	three-timer extinction and origination rates	turnover	Alroy, 2008
`extC3t, oriC3t`	corrected three-timer extinction and origination rates	turnover	Alroy, 2008
`extGF, oriGF`	gap-filler extinction and origination rates	turnover	Alroy, 2014
`ext2f3, ori2f3`	second-for-third substitution extinction and origination rates	turnover	Alroy, 2015
`divSIB`	sampled-in-bin diversity (SIB)	richness	Miller and Foote, 1996
`divRT`	range-through diversity (RT)	richness	Newell, 1952
`divBC`	boundary-crosser diversity (BC)	richness	Carr and Kitchell, 1980
`divCSIB`	corrected sampled-in-bin diversity	richness	Alroy et al., 2008
`samp3t`	three-timer sampling completeness	sampling completeness	Alroy, 2008
`sampRange`	range-based sampling completeness	sampling completeness	Foote and Miller, 2007

1 Table 2. Results of method-specific outcomes of the time series calculations for 10 myr bins (a) and stages (b).

2 Unless otherwise indicated, values are estimates of correlation coefficients. Significant results of Shapiro-Wilk tests

3 indicates deviation from log-normality. Formatting indicates significance (except for binary information): insignificant

4 values are not reported, regular entries indicate 0.05 ≥ *p* > 0.01, **bold** entries 0.01 ≥ *p* > 0.001 and **<u>underscored</u>**

5 **<u>bold</u>** entries denote *p* ≤ 0.001. Rate metrics are abbreviated: PC = the per capita rates (Foote 1999), C3t = corrected

6 three-timer rates (Alroy 2008), GF = gap-filler rates (Alroy, 2014) and 2f3 = second-for-third-substitution rates (Alroy,

7 2015).

(a) 10 MYR TIME SCALE

		raw PC	raw C3t	raw GF	raw 2f3	CR PC	CR C3t	CR GF	CR 2f3	SQS PC	SQS C3t	SQS GF	SQS 2f3
Pulsed/ continuous	ext. rates with durations												
	norm. ext. rates with durations	**<u>-0.49</u>**	**-0.44**	-0.34	-0.3	**<u>-0.56</u>**	**<u>-0.58</u>**	**<u>-0.55</u>**	**<u>-0.55</u>**	**<u>-0.47</u>**	**-0.46**	-0.32	-0.34
	orig. rates with durations			0.32									
	norm. orig. rates with durations	-0.37	**-0.42**			-0.36	**-0.4**				-0.35		
Declines	extinctions	**<u>0.55</u>**	**<u>0.58</u>**	**<u>0.61</u>**	**<u>0.51</u>**	**<u>0.55</u>**	**<u>0.61</u>**	**<u>0.62</u>**	**<u>0.57</u>**	**<u>0.55</u>**	**<u>0.61</u>**	**<u>0.65</u>**	**<u>0.57</u>**
	post-Ordovician extinctions		0.38	**0.42**			**0.45**	**0.44**	0.35		0.41	**0.44**	
	originations	**<u>0.5</u>**	0.45	**<u>0.5</u>**		**<u>0.55</u>**	0.43	**<u>0.52</u>**	0.41	**<u>0.53</u>**	0.43	**<u>0.53</u>**	
	post-Ordovician originations												
Mass extinctions	end-Permian ME	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
	end-Triassic ME	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
	end-Cretaceous ME	**yes**	**yes**	**yes**	**yes**	no	**yes**	**yes**	**yes**	no	**yes**	no	no
	end-Permian is highest	**yes**	no	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
	number of mass extinctions	**5**	**3**	**4**	**3**	**2**	**4**	**3**	**3**	**2**	**3**	**2**	**3**
Rate dist.	extinctions log-normal (p-values)						0.04						
	originations log-normal (p-values)			**<u><0.001</u>**								**0.01**	**0.01**
Equilibrial dyn.	origination and lagged diversity					0.35			0.32	**0.45**		0.31	0.35
	diversity and lagged extinction	**0.45**				0.35				**0.48**			
	diversity and lagged origination		**<u>0.55</u>**				**<u>0.67</u>**	0.48	0.42		**<u>0.54</u>**		

(b) STAGE-LEVEL TIME SCALE

		raw PC	raw C3t	raw GF	raw 2f3	CR PC	CR C3t	CR GF	CR 2f3	SQS PC	SQS C3t	SQS GF	SQS 2f3
Pulsed/ continuous	ext. rates with durations	0.23		0.22	**0.28**				0.22				**0.27**
	norm. ext. rates with durations	**<u>-0.45</u>**	**<u>-0.35</u>**	**-0.34**		**<u>-0.64</u>**	**<u>-0.53</u>**	**<u>-0.58</u>**	**-0.42**	**<u>-0.54</u>**	**<u>-0.51</u>**	**<u>-0.46</u>**	-0.28
	orig. rates with durations	**<u>0.35</u>**		0.32	0.33					0.24		0.24	0.23
	norm. orig. rates with durations	**<u>-0.52</u>**	**<u>-0.51</u>**	**<u>-0.4</u>**	-0.31	**<u>-0.66</u>**	**<u>-0.58</u>**	**<u>-0.55</u>**	**<u>-0.51</u>**	**<u>-0.59</u>**	**<u>-0.5</u>**	**<u>-0.41</u>**	**<u>-0.39</u>**
Declines	extinctions	**<u>0.52</u>**	**0.39**	**0.39**	**0.36**	**<u>0.59</u>**	**<u>0.43</u>**	**<u>0.46</u>**	**<u>0.46</u>**	**<u>0.59</u>**	**0.4**	**<u>0.39</u>**	**<u>0.37</u>**
	post-Ordovician extinctions					0.27				0.29			
	originations	**<u>0.44</u>**	**0.36**	**0.39**	**0.35**	**<u>0.58</u>**	**<u>0.39</u>**	**<u>0.43</u>**	**<u>0.43</u>**	**<u>0.52</u>**	**0.36**	**<u>0.39</u>**	**<u>0.38</u>**
	post-Ordovician originations					0.24							
Mass extinctions	end-Permian ME	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
	end-Triassic ME	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
	end-Cretaceous ME	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
	end-Permian is highest	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
	number of mass extinctions	**4**	**3**	**5**	**4**	**3**	**3**	**3**	**4**	**3**	**3**	**6**	**3**
Rate dist.	extinctions log-normal (p-values)			0.01	**<u><0.001</u>**			**<u><0.001</u>**				**<u><0.001</u>**	**<u><0.001</u>**
	originations log-normal (p-values)		0.05	**<u><0.001</u>**				**<u><0.001</u>**	**<u><0.001</u>**			**<u><0.001</u>**	0.02
Equilibrial dyn.	origination and lagged diversity	**<u>0.39</u>**	0.3			**0.34**				**<u>0.52</u>**	0.34	0.34	0.28
	diversity and lagged extinction	**<u>0.6</u>**	**0.41**	**0.39**	0.34	**<u>0.45</u>**				**<u>0.61</u>**	0.23	**0.37**	0.3
	diversity and lagged origination						**0.46**	**0.46**	0.38				