GSA DATA REPOSITORY 2012046 Lloyd et al.

Additional Details of Methodology

The Database

As it was not tractable to enter data from the world's oceans as a whole we limited ourselves to the North Atlantic, which we define as 90°N to 20°S and including both the Mediterranean and Caribbean, and its margins. This area offers the key advantage of being relatively densely sampled by the DSDP and ODP, compared to say, the Pacific. Similarly the margins, particularly North America and Western Europe, are also heavily studied.

The use of deep-sea data (i.e. cores) limits the number of fossil groups that can be studied to calcareous micro- and nannofossils (although siliceous microfossils would work just as well). For the present study we chose the Coccolithophores as they are abundant, diverse and commonly recorded as they are major zone fossils and often used as a first means of dating sediments. They have been recorded from numerous land-based sections and cores and are essentially cosmopolitan (having very little endemism), making them ideal candidates for this study.

Picking a suitable rock record measure in the deep sea is not as easy as it is on the land. For this reason Lloyd *et al.* (in press) ultimately chose the number of different DSDP and ODP sites with coccolithophore-bearing rock of that age. In order to make our land collections as similar as possible we use the number of localities (bearing a unique latitude-longitude) with coccolithophore-bearing rock. These can either be outcrop data, where sediment was systematically collected against a logged section, or borehole (core) data more similar to the deep-sea approach. In all cases we required the primary species occurrence data to come from distribution (and not range) charts in order that it reflect *sampled* diversity, and hence be suitable for the subsampling and modelling approaches used here. Similarly, in order for a reference to be included it must incorporate a full microfloral list and not just zone fossils that would artificially under represent biodiversity.

In order to further reduce the amount of data entry required we chose to make the fundamental unit of out database a biozone at a site or locality rather than tabulate the much more numerous individual samples. Here a biozone is either a nannofossil or planktic foraminifera zone. Our dates come from Ogg *et al.* (2008), and specifically the TimeScale Creator program (https://engineering.purdue.edu/Stratigraphy/tscreator/).

Analytical Methods

1. Taxonomic standardisation. Before analysing our data we standardised our taxonomy using a new list of valid, invalid and synonymised taxa originally based on the NEPTUNE database (Lazarus 1994; Spencer-Cervato 1999), but significantly overhauled

by one of us (JRY). In the process of manual data entry we have additionally uncovered many names not included in the NEPTUNE list making our global nannofossil synonymy list the most comprehensive and up-to-date presently available. This list is stored in the main database, allowing data entry to proceed using the original names from the distribution charts. This procedure thus allows for a different future taxonomy to still be used should opinions on synonymy etc. change. For data analysis we adopt the following procedure: 1) synonyms are replaced with their senior counterparts, 2) any resulting duplicate occurrences are removed and, 3) invalid taxa, questionable occurrences, taxa whose status is presently considered unknown and cf. or aff. taxa are removed.

2. Creating time bins of equal duration and calculating error bars. Units are given numerical dates based on Ogg *et al.* (2008) and TimeScale Creator as follows. If only a nannofossil or planktic foraminiferal zonation is known then the top of the youngest and bottom of the oldest are used. If both zone types are present then the dates of the overlap are used, conferring greater precision. In some cases, however, the two zonations do not overlap (implying uncertainty). When this happens then the maximum possible age range is used. Finally, if the uncertainty between the maximum and minimum possible dates is large (>15 million years) then we remove that unit and its constituent taxa from the analyses.

As we are interested in counts of species richness through time an appropriate time binning approach is required. However, nannofossil or foraminiferal zones are problematic to use as they vary considerably in length and are thus likely to give misleading results, with more taxa likely to accumulate in a longer bin than a shorter one. This problem was identified by Sepkoski & Koch (1996) who recommended using time bins of roughly equal length. Alroy *et al*. (2008) adopted such an approach by combining geologic stages to get roughly 11 million-year time bins. Although this is appropriate for an overview of Phanerozoic macrofossils, or poorly time-constrained taxa such as dinosaurs (Lloyd *et al.* 2008) such coarse binning is unnecessary for the data used here. Instead we adopt the Alroy *et al*. (2008) approach, but combine biozones, instead of geologic stages, to make roughly 6 million-year time bins. We made an additional modification to this approach however, which is to enforce the inclusion of major geologic boundaries (the Jurassic-Cretaceous, Cretaceous-Palaeogene, Eocene-Oligocene and Palaeogene-Neogene). This is because these are often associated with major turnover events and a bin spanning such a boundary is thus likely to have artificially inflated diversity because of an extinction and recovery flora being time-averaged together. In application we ended up with time bins of mean length and standard deviation of 5.87 and 1.22 m.y., respectively (boundaries at: 0, 6.14, 11.90, 17.95, 23.03, 30.04, 33.90, 42.42, 48.60, 54.09, 59.99, 65.50, 72.35, 77.38, 83.99, 89.63, 96.01, 101.49, 109.06, 114.30, 120.70, 125.11, 131.70, 140.24, 145.50, 149.95, 156.30, 161.61, 168.24, 174.18, 181.35, 185.30, 191.87, 196.69 and 199.60 Ma).

Even after clumping zones together it is inevitable that some units will lack the precise dating required to assign them to a single time bin. Previous workers have had diametrically opposed solutions to this quandary. For example, Alroy *et al*. (2008) simply ignore taxa that cannot be assigned to a single bin and don't count them. By

contrast, vertebrate workers have tended to treat uncertainty instead as the range of a taxon, counting it in all bins it could *possibly* be in (e.g., Benton 1995). Here we regard both solutions to be somewhat extreme and prefer instead a method that is intended to better quantify this uncertainty. Firstly we assume that each unit really does belong to a single bin and assign it based on a randomisation approach. This is done by picking a random number from a uniform distribution between the oldest and youngest possible dates for the unit. We then assign the unit to a time bin based on this single date and perform all of the counts outlined below. We then repeat this procedure 1,000 times and record the resulting mean and 95% confidence intervals for our counts.

3. Picking a sampling proxy. Here we use the number of DSDP/ODP sites or land-based localities that have yielded sediments dated to a specific time bin as a measure of sampling. Sites can be considered a good measure of sampling, as they are decided on a priori by the researchers on the DSDP/ODP leg. Localities are thus the obvious equivalent proxy on land, although in many cases these may be determined by other factors such as present exposure at the surface. In any case there are limited alternatives, as there is no such thing as deep-sea formations and map area would be similarly inappropriate.

4. Modelling. Assuming that sampling is a major factor in producing observed taxonomic richness curves an interesting follow up question is: how much of the observed richness is unexplained by sampling? Smith & McGowan (2007) introduced a procedure to tackle this question that starts from the notion that sampling perfectly predicts observed richness. In other words, the smallest sample is matched up with the lowest observed richness, the second smallest with the second lowest and so on. A simple linear model is then fitted to this new data from which a function can be derived that allows us to predict the richness for a given sample.

Here we extend the method of Smith & McGowan (2007) by considering non-linearity in the data by fitting other models: logarithmic, exponential, hyperbolic, sigmoidal and polynomial (Lloyd in press). The best model is then chosen using the sample size corrected Akaike information Criterion (AIC; Akaike 1973), the AIC_c (Johnson & Omland 2004). The AIC_c weighs both a model's fit (a close fit being best) and its complexity (a simple model being best).

5. Subsampling. An alternative way to remove a sampling bias from a species richness curve is to rarefy or subsample (e.g., Alroy *et al.* 2001; Alroy *et al.* 2008). Here we use Alroy's (2010a,b,c) method of 'shareholder quorum subsampling' (SQS) that solves some of the problems with classical rarefaction. As our data is literature based we use the single publication occurrence correction (Alroy2010a,c). This is employed using a modified version of Alroy's R function (http://www.nceas.ucsb.edu/~alroy/SQS.html) that employs this correction and is available from GTL on request. Here we report results for $q = 0.4$ only as higher values mean some time bins do not return a result and lower values tend towards a flatter curve equivalent to a null hypothesis. Unlike classical rarefaction SQS does not require multiple replicates and hence does not produce

confidence intervals. However, due to the way uncertainty of dating is handled (see 2 above) we do report confidence intervals in Figs. 2 and 3.

6. Median unit count. The final way in which sampling bias is addressed is to tabulate the median number of taxonomic entities described at a particular time and place. For less cosmopolitan groups this would be analogous to alpha diversity. Here this value is calculated as follows: 1) For each unit in our database (Lloyd *et al.* in press) we first total the list of different taxa – regardless of level (e.g., genus, species, 'group') or current validity (i.e., synonyms) – which are recorded as present. 2) When the unit list for a time bin is compiled following randomisation (see 2 above) the median value is taken and stored. 3) The median and 95% confidence intervals of this value are then plotted (Fig. 3*C*). (NB: Similar results are obtained if a mean value is used.)

References

Akaike H (1973) in *2nd International Symposium on Information Theory*, eds Petrov BN, Csaki F (Akadémiai Kiadó, Budapest), pp 267-281.

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

Alroy J (2010b) in *Quantitative Methods in Paleobiology*, eds Alroy J, Hunt G (The Paleontological Society, New Haven), pp 55-80.

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

Alroy J, Marshall CR, Bambach RK, Bezusko K, Foote M, Fursich FT, Hansen TA, Holland SM, Ivany LC, Jablonski D, Jacobs DK, Jones DC, Kosnik MA, Lidgard S, Low S, Miller AI, Novack-Gottshall PM, Olszewski TD, Patzkowsky ME, Raup DM, Roy K, Sepkoski JJ, Sommers MG, Wagner PJ, Webber A. (2001) Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proc Natl Acad Sci USA* 98:6261-6266.

Alroy J, Aberhan M, Bottjer DJ, Foote M, Fursich FT, Harries PJ, Hendy AJW, Holland SM, Ivany LC, Kiessling W, Kosnik MA, Marshall CR, McGowan AJ, Miller AI, Olszewski TD, Patzkowsky ME, Peters SE, Villier L, Wagner PJ, Bonuso N, Borkow PS, Brenneis B, Clapham ME, Fall LM, Ferguson CA, Hanson VL, Krug AZ, Layou KM, Leckey EH, Nurnberg S, Powers CM, Sessa JA, Simpson C, Tomasovych A, Visaggi CC (2008) Phanerozoic trends in the global diversity of marine invertebrates. *Science* 321:97–100.

Benton MJ (1995) Diversity and extinction in the history of life. *Science* 268:52-58.

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. *Trend Ecol Evol* 19:101-108.

Lazarus D (1994) Neptune: a marine micropaleontology database. *Math. Geol.* 26:817-832.

Lloyd GT (in press) A refined modelling approach to assess the influence of sampling on palaeobiodiversity curves: new support for declining Cretaceous dinosaur richness. *Biol Lett* XX:XX-XX.

Lloyd GT, Smith AB, Young JR (in press) in *Comparing the Rock and Fossil Records: Implications for Biodiversity*, eds McGowan AJ, Smith AB (Geological Society, London), pp XX-XX.

Lloyd GT, Davis KE, Pisani D, Tarver JE, Ruta M, Sakamoto M, Hone DWE, Jennings R, Benton MJ (2008) Dinosaurs and the Cretaceous Terrestrial Revolution. *Proc Roy Soc Lond B* 275:2483-2490.

Ogg JG, Ogg G, Gradstein FM (2008) *The Concise Geologic Time Scale*, (Cambridge University Press, Cambridge).

Sepkoski JJ, Koch CF (1996) in *Global events and event stratigraphy in the Phanerozoic*, ed Walliser OH (Springer, Berlin), pp 21–34.

Smith AB, McGowan AJ (2007) The shape of the Phanerozoic diversity curve. How much can be predicted from the sedimentary rock record of Western Europe? *Palaeontology* 50:765-777.

Spencer-Cervato C (1999) The Cenozoic deep-sea microfossil record: explorations of the DSDP/ODP sample set using the Neptune database. *Palaeo. Electronica* 2.

Table DR1. Comparative statistical correlation (Spearman Rank) between land-based and deep-sea sampling and fossil records. Significant *p*-values (at $\alpha = 0.05$) are in bold.

Comparison:		Raw data		Generalised differences	
		ρ	*p*	ρ	*p*
Sampling	Land-based vs. deep-sea	0.27	0.13	0.11	0.52
Species	Land-based richness vs. deep-sea richness (Total)	0.44	**<0.01**	0.02	0.93
	Land-based richness vs. deep-sea richness (Mesozoic)	0.81	**<0.01**	0.01	0.97
	Land-based richness vs. deep-sea richness (Cenozoic)	-0.01	0.99	0.23	0.51
	Land-based richness vs. land sampling	0.84	**<0.01**	0.48	**<0.01**
	Land-based richness vs. deep-sea sampling	0.37	**0.03**	0.02	0.92
	Deep-sea richness vs. deep-sea sampling	0.95	**<0.01**	0.40	**0.02**
	Deep-sea richness vs. land sampling	0.34	0.05	0.19	0.29
Genera	Land-based richness vs. deep-sea richness (total)	0.80	**<0.01**	0.26	0.15
	Land-based richness vs. deep-sea richness (Mesozoic)	0.86	**<0.01**	0.04	0.86
	Land-based richness vs. deep-sea richness (Cenozoic)	0.42	0.20	0.58	0.09
	Land-based richness vs. land sampling	0.79	**<0.01**	0.46	**<0.01**
	Land-based richness vs. deep-sea sampling	0.24	0.16	0.01	0.93
	Deep-sea richness vs. deep-sea sampling	0.53	**<0.01**	0.36	**0.04**
	Deep-sea richness vs. land-based sampling	0.65	**<0.01**	0.36	**0.04**

Table DR2. Comparative statistical correlation (Spearman Rank) between and within shareholder quorum subsampling (Alroy, 2010; SQS) and model-corrected (Lloyd, 2011; MC) diversity estimates. Significant *p*-values (at $\alpha = 0.05$) are in bold.

	Richness comparison:	Raw data		Generalised differences	
		ρ	*p*	ρ	*p*
Species	SQS land-based v MC land-based	0.64	**<0.01**	0.54	**<0.01**
	SQS deep-sea v MC deep-sea	0.71	**<0.01**	0.68	**<0.01**
	SQS combined v MC combined	0.75	**<0.01**	0.62	**<0.01**
	SQS land-based v SQS deep-sea	0.64	**<0.01**	0.37	0.07
	MC land-based v MC deep-sea	0.13	0.46	-0.02	0.89
Genera	SQS land-based v MC land-based	0.55	**<0.01**	0.37	**0.04**
	SQS deep-sea v MC deep-sea	0.96	**<0.01**	0.75	**<0.01**
	SQS combined v MC combined	0.75	**<0.01**	0.68	**<0.01**
	SQS land-based v SQS deep-sea	0.76	**<0.01**	0.17	0.40
	MC land-based v MC deep-sea	0.33	0.06	-0.14	0.42

Table DR3. Comparative statistical correlation (Spearman Rank) between shareholder quorum subsampling (Alroy, 2010; SQS) and model-corrected (Lloyd, 2011; MC) diversity curves with sampled diversity. Significant *p*-values (at $\alpha = 0.05$) are in bold.

	Richness comparison:	Raw data		Generalised differences	
		ρ	*p*	ρ	*p*
Species	SQS deep-sea v sampled deep-sea	0.06	0.78	0.38	0.06
	SQS land-based v sampled land-based	0.82	**<0.01**	0.56	**<0.01**
	SQS combined v sampled deep-sea	0.35	**0.04**	0.05	0.76
	SQS combined v sampled land-based	0.87	**<0.01**	0.52	**<0.01**
	MC deep-sea v sampled deep-sea	0.32	0.06	0.64	**<0.01**
	MC land-based v sampled land-based	0.28	0.11	0.11	0.55
	MC combined v sampled deep-sea	0.00	0.98	0.17	0.34
	MC combined v sampled land-based	0.64	**<0.01**	0.53	**<0.01**
Genera	SQS deep-sea v sampled deep-sea	0.60	**<0.01**	0.52	**<0.01**
	SQS land-based v sampled land-based	0.91	**<0.01**	0.77	**<0.01**
	SQS combined v sampled deep-sea	0.76	**<0.01**	0.37	**0.05**
	SQS combined v sampled land-based	0.90	**<0.01**	0.71	**<0.01**
	MC deep-sea v sampled deep-sea	0.50	**<0.01**	0.70	**<0.01**
	MC land-based v sampled land-based	0.35	**0.04**	0.16	0.37
	MC combined v sampled deep-sea	0.31	0.07	0.22	0.21
	MC combined v sampled land-based	0.60	**<0.01**	0.66	**<0.01**

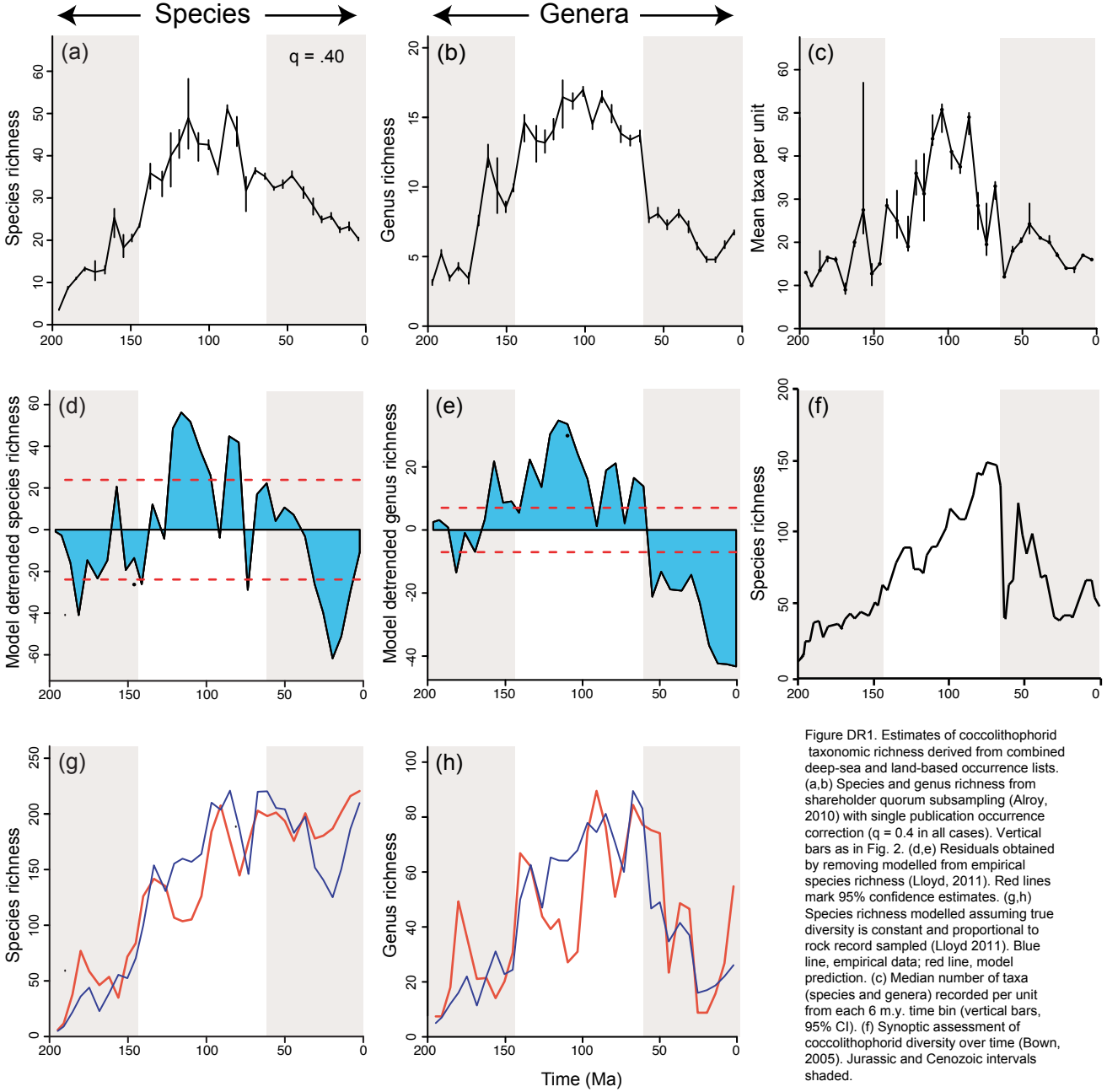

Figure DR1. Estimates of coccolithophorid taxonomic richness derived from combined deep-sea and land-based occurrence lists. (a,b) Species and genus richness from shareholder quorum subsampling (Alroy, 2010) with single publication occurrence correction (q = 0.4 in all cases). Vertical bars as in Fig. 2. (d,e) Residuals obtained by removing modelled from empirical species richness (Lloyd, 2011). Red lines mark 95% confidence estimates. (g,h) Species richness modelled assuming true diversity is constant and proportional to rock record sampled (Lloyd 2011). Blue line, empirical data; red line, model prediction. (c) Median number of taxa (species and genera) recorded per unit from each 6 m.y. time bin (vertical bars, 95% CI). (f) Synoptic assessment of coccolithophorid diversity over time (Bown, 2005). Jurassic and Cenozoic intervals shaded.