SUPPLEMENTARY MATERIAL 1
Spatial sampling heterogeneity limits the detectability of deep time latitudinal biodiversity gradients

Lewis A. Jones, Christopher D. Dean, Philip D. Mannion, Alexander Farnsworth, and Peter A. Allison

Author for correspondence: Lewis A. Jones (LewisA.Jones@outlook.com)

This PDF includes:

Supplementary Tables

Table S1. Summary statistics of virtual species range sizes for a unimodal-type latitudinal biodiversity gradient simulation for the present day ................................................................. 2
Table S2. Kolmogorov-Smirnov two-sample test results from empirical range size frequency distributions and those simulated ............................................................................................................ 2
Table S3. Kolmogorov-Smirnov two-sample test results from empirical occurrence frequency distributions and those simulated ............................................................................................................ 3
Table S4. Summary table of the percentage of correctly reconstructed palaeolatitudinal bins and zones (tropics, temperate, polar) with peak diversity ............................................................................................................ 3

Supplementary Figures

Figure S1. A graphic depicting the major steps in the simulations used in this study .............................................. 4
Figure S2. A graphic depicting the types of latitudinal biodiversity gradients simulated in this study ................. 4
Figure S3. Probability grids used in simulations to ‘force’ occurrence data to conform to the specified type of latitudinal biodiversity gradient ......................................................................................... 5
Figure S4. Range size frequency distributions of species within major marine invertebrate groups .................. 6
Figure S5. Example of range size frequency distributions from a unimodal-type latitudinal biodiversity gradient simulation for the present day ............................................................................................ 7
Figure S6. Example occurrence frequency distributions for empirical and simulated data .................................. 7
Figure S7. Global spatial sampling coverage and extent for different types of latitudinal biodiversity gradients ...... 8
Figure S8. Heat maps of residual palaeolatitudinal richness between simulated and sampled/sampling-standardised latitudinal biodiversity gradients ................................................................. 9
Figure S9. Total displacement between proportional simulated and sampled/sampling-standardised diversity curves. 10 Figure S10. Total displacement between pair-wise combinations of proportional simulated, sampled and sampling-standardised diversity curves, for different types of latitudinal biodiversity gradient ................................................................. 11
Figure S11. Pearson’s correlation coefficient between simulated and sampled/sampling-standardised diversity curves for each type of latitudinal biodiversity gradient ......................................................................................................... 12
Figure S12. Pearson’s correlation coefficient between pairwise combinations of simulated, sampled and sampling-standardised diversity curves, for different types of latitudinal biodiversity gradient ......................................................................................................... 12
Figure S13. Kolmogorov-Smirnov two-sample tests between simulated and sampled/sampling-standardised diversity curves for each type of latitudinal biodiversity gradient ......................................................................................................... 13
Figure S14. Asselian to Piacenzian plot of the palaeolatitudinal bin with peak diversity for simulated flat-type (a, b), unimodal-type (c, d) and bimodal-type (e, f) latitudinal biodiversity gradients ......................................................................................................... 14
Figure S15. Relationship between palaeolatitudinal richness and spatial sampling coverage ................................ 15
Figure S16. Relationship between palaeolatitudinal richness and spatial sampling extent .................................... 16
Figure S17. Global raw counts of sampled diversity from simulations of species’ distributions and richness .......... 17
Figure S18. Relationship between global sampled richness and spatial sampling coverage and extent ............... 18
Supplementary Tables

Table S1. Summary statistics of virtual species range sizes for a unimodal-type latitudinal biodiversity gradient simulation for the present day. Values are mean scores for 1000 species across the 100 simulation replicates.

| Metric                          | Mean  | SD   | Minimum | Maximum |
|---------------------------------|-------|------|---------|---------|
| Latitude (°)                    | 11.38 | 11.43| 0       | 148     |
| Longitude (°)                   | 11.70 | 11.53| 0       | 163     |
| Great Circle Distance (km)      | 1,785 | 1,486| 0       | 16,433  |

Table S2. Kolmogorov-Smirnov two-sample test results from empirical range size frequency distributions and those simulated. Significant results ($P < 0.05$) are identified with an asterisk and indicate statistically significant differences between groups.

| Comparison                      | Metric                          | D Statistic | P       |
|---------------------------------|---------------------------------|-------------|---------|
| Simulated and Porifera          | Latitude                        | 0.625       | 0.088   |
| Simulated and Porifera          | Longitude                       | 0.75        | 0.002*  |
| Simulated and Porifera          | Great Circle Distance           | 0.667       | 0.143   |
| Simulated and Echinoidea        | Latitude                        | 0.5         | 0.027   |
| Simulated and Echinoidea        | Longitude                       | 0.583       | 0.034*  |
| Simulated and Echinoidea        | Great Circle Distance           | 0.5         | 0.474   |
| Simulated and Brachiopoda       | Latitude                        | 0.375       | 0.627   |
| Simulated and Brachiopoda       | Longitude                       | 0.333       | 0.518   |
| Simulated and Brachiopoda       | Great Circle Distance           | 0.333       | 0.893   |
| Simulated and Bivalvia          | Latitude                        | 0.75        | 0.022*  |
| Simulated and Bivalvia          | Longitude                       | 0.75        | 0.002*  |
| Simulated and Bivalvia          | Great Circle Distance           | 0.667       | 0.143   |
| Simulated and Anthozoa          | Latitude                        | 0.75        | 0.022*  |
| Simulated and Anthozoa          | Longitude                       | 0.75        | 0.002*  |
| Simulated and Anthozoa          | Great Circle Distance           | 0.667       | 0.143   |
| Porifera and Echinoidea         | Latitude                        | 0.625       | 0.087   |
| Porifera and Echinoidea         | Longitude                       | 0.667       | 0.01*   |
| Porifera and Echinoidea         | Great Circle Distance           | 0.833       | 0.026*  |
| Brachiopoda and Porifera        | Latitude                        | 0.875       | 0.004*  |
| Brachiopoda and Porifera        | Longitude                       | 0.917       | < 0.01* |
| Brachiopoda and Porifera        | Great Circle Distance           | 1           | 0.002*  |
| Brachiopoda and Echinoidea      | Latitude                        | 0.75        | 0.022*  |
| Brachiopoda and Echinoidea      | Longitude                       | 0.75        | 0.002*  |
| Brachiopoda and Echinoidea      | Great Circle Distance           | 0.833       | 0.026*  |
| Bivalvia and Porifera           | Latitude                        | 0.375       | 0.66    |
| Bivalvia and Porifera           | Longitude                       | 0.333       | 0.518   |
| Bivalvia and Porifera           | Great Circle Distance           | 0.5         | 0.474   |
| Bivalvia and Echinoidea         | Latitude                        | 0.75        | 0.019*  |
| Bivalvia and Echinoidea         | Longitude                       | 0.75        | 0.002*  |
| Bivalvia and Echinoidea         | Great Circle Distance           | 0.833       | 0.026*  |
| Bivalvia and Brachiopoda        | Latitude                        | 0.875       | 0.004*  |
| Bivalvia and Brachiopoda        | Longitude                       | 0.917       | < 0.01* |
| Bivalvia and Brachiopoda        | Great Circle Distance           | 1           | 0.002*  |
| Anthozoa and Porifera           | Latitude                        | 0.375       | 0.66    |
| Anthozoa and Porifera           | Longitude                       | 0.333       | 0.518   |
| Anthozoa and Porifera           | Great Circle Distance           | 0.333       | 0.931   |
| Anthozoa and Echinoidea         | Latitude                        | 0.75        | 0.019*  |
| Anthozoa and Echinoidea         | Longitude                       | 0.833       | < 0.01* |
| Anthozoa and Echinoidea         | Great Circle Distance           | 0.833       | 0.026*  |
| Anthozoa and Brachiopoda        | Latitude                        | 0.875       | 0.004*  |
| Anthozoa and Brachiopoda        | Longitude                       | 0.917       | < 0.01* |
| Anthozoa and Brachiopoda        | Great Circle Distance           | 1           | 0.002*  |
| Anthozoa and Bivalvia           | Latitude                        | 0.25        | 0.98    |
| Anthozoa and Bivalvia           | Longitude                       | 0.167       | 0.996   |
| Anthozoa and Bivalvia           | Great Circle Distance           | 0.333       | 0.931   |
Table S3. Kolmogorov-Smirnov two-sample test results from empirical occurrence frequency distributions and those simulated. Significant results ($P < 0.05$) are identified with an asterisk and indicate statistically significant differences between groups.

| Comparison                  | $D$ Statistic | $P$  |
|-----------------------------|---------------|------|
| Simulated and Anthozoa     | 0.556         | 0.124|
| Simulated and Porifera     | 0.333         | 0.73 |
| Simulated and Bivalvia     | 0.444         | 0.336|
| Simulated and Brachiopoda  | 0.444         | 0.336|
| Simulated and Echinoidea   | 0.333         | 0.699|
| Anthozoa and Porifera      | 0.556         | 0.126|
| Anthozoa and Bivalvia      | 0.222         | 0.989|
| Anthozoa and Brachiopoda   | 0.889         | 0.002*|
| Anthozoa and Echinoidea    | 0.556         | 0.126|
| Bivalvia and Porifera      | 0.556         | 0.126|
| Bivalvia and Brachiopoda   | 0.889         | 0.002*|
| Bivalvia and Echinoidea    | 0.556         | 0.126|
| Brachiopoda and Porifera   | 0.667         | 0.037*|
| Brachiopoda and Echinoidea | 0.667         | 0.037*|
| Porifera and Echinoidea    | 0.222         | 0.979|

Table S4. Summary table of the percentage of correctly reconstructed palaeolatitudinal bins and zones (tropics, temperate, polar) with peak diversity, for each type of simulated latitudinal biodiversity gradient (flat, unimodal, and bimodal). Zones are broadly defined as 30°S to 30°N for the tropics, 30°S to 60°S and 30°N to 60°N for temperate regions, and 60°S to 90°S and 60°N to 90°N for polar regions. Percentages are based on all 56 stages (Asselian–Piacenzian).

| Type of latitudinal gradient | Reconstruction type | Correct palaeolatitudinal bin (%) | Correct zone (%) |
|------------------------------|---------------------|----------------------------------|------------------|
| Flat                         | Sampled             | 17.86                            | 30.36            |
| Unimodal                     | Sampled             | 48.21                            | 73.21            |
| Bimodal                      | Sampled             | 69.64                            | 96.43            |
| Flat                         | Sampling-standardised| 17.86                            | 30.36            |
| Unimodal                     | Sampling-standardised| 50                               | 75               |
| Bimodal                      | Sampling-standardised| 73.21                            | 98.21            |
Supplementary Figures

Figure S1. A graphic depicting the major steps in the simulations used in this study. Simulations are not intended to reconstruct genuine latitudinal biodiversity gradients in deep time. The purpose of these simulations is to emulate the distribution of biodiversity given three different types of gradient (flat, unimodal and bimodal). Diagram depicts only one simulation run. However, simulations were run 100 times for each type of latitudinal biodiversity gradient (=3) and stratigraphic stage (=56). In this study, 1,000 species were generated for each simulation run. Shallow marine area was defined as 0–200 m depth.

Figure S2. A graphic depicting the types of latitudinal biodiversity gradients simulated in this study.
Figure S3. Probability grids used in simulations to ‘force’ occurrence data to conform to the specified type of latitudinal biodiversity gradient. (a) A latitudinal biodiversity gradient with no specified peak in biodiversity (flat-type); (b) a tropical peak in biodiversity with a sharp decline towards the poles (unimodal-type); and (c) temperate peaks in biodiversity that decline towards the poles and equator (bimodal-type). The probability grid for a flat-type latitudinal biodiversity gradient with all cells equal to 1. The unimodal-type probability grid had a maximum value of 1 at the equator, declining towards the poles ($\mu = 0, \sigma = 20$). The bimodal-type probability grid had a maximum value of 1 at 45°N and 45°S, declining towards both the poles and equator ($\mu = 45, \sigma = 10$ for each hemisphere).
Figure S4. Range size frequency distributions of species within the marine invertebrate groups Anthozoa (a, b, c), Porifera (d, e, f), Bivalvia (g, h, i), Brachiopoda (j, k, l) and Echinoidea (m, n, o). (a, d, g, j, m) Latitude; (b, e, h, k, n) longitude; (c, f, i, l, o) Great Circle Distance. Range size is calculated as the maximum latitudinal, longitudinal or Great Circle Distance distance between occurrences of each species. Great Circle Distance was calculated using the spDists() function from the R package ‘sp’ [1].
Figure S5. Example of range size frequency distributions from a unimodal-type latitudinal biodiversity gradient simulation for the present day. (a) Latitude; (b) longitude; (c) Great Circle Distance. Range size is calculated as the maximum latitudinal, longitudinal or Great Circle Distance distance between occurrences of each species. Great Circle Distance was calculated using the spDists() function from the R package ‘sp’ [1].

Figure S6. Example occurrence frequency distributions. (a) Anthozoa; (b) Bivalvia; (c) Brachiopoda; (d) Echinoidea; (e) Porifera; (f) simulated. (f) Occurrence frequency distribution example for a unimodal-type latitudinal biodiversity gradient simulation for the present day.
Figure S7. Global spatial sampling coverage and extent (summed minimum-spanning tree [MST] length). (a) Spatial sampling coverage is calculated as the percentage of cells sampled from the total amount of available cells, where available of cells is the number of cells within shallow marine masks. (b) Summed MST length is calculated between all sampled cells within shallow marine masks.
Figure S8. Heat maps of residual palaeolatitudinal richness between simulated and sampled/sampling-standardised latitudinal biodiversity gradients. Heat maps range from the Asselian (298.9 Myr ago) to Piacenzian (2.58 Myr ago) at stage level, and show data within 15° palaeolatitudinal bins. (a–c) Residual proportional richness between mean simulated and sampled richness; (d–f) residual proportional richness change between mean simulated and sampling-standardised richness. For each type of latitudinal biodiversity gradient, simulated richness values were based on the mean bin value from the 100 simulation replications. Sampling-standardised palaeolatitudinal richness estimates (d–f) are based on the mean of means of the 1,000 bootstrap replicates for each replication of latitudinal biodiversity gradient simulations (= 100). Proportional latitudinal richness values were calculated by dividing each palaeolatitudinal bin by the total of all palaeolatitudinal bins within each stage. Residual richness was calculated by subtracting proportional simulated latitudinal biodiversity gradients from sampled/sampling-standardised counterparts. Residual richness scores of >0 indicate the relative diversity of the palaeolatitudinal bin is over-represented whereas scores of <0 indicate the relative diversity of the palaeolatitudinal bin is under-represented. Residual richness scores indicate that the relative richness of palaeolatitudinal bin 30–45° is over-represented for most of the Jurassic to Neogene (a–f). Conversely, relative richness in tropical palaeolatitudinal bins are consistently under-represented in unimodal-type gradients (b, c). Grey tiles illustrate palaeolatitudinal bins with insufficient data to compute sampling-standardised (classical rarefaction) richness estimates.
Figure S9. Total displacement ($D$; see methods) between proportional simulated and sampled/sampling-standardised (orange/purple line) diversity curves (a: flat-type, b: unimodal-type, c: bimodal-type). The smaller the difference, the more similar the curves.
Figure S10. Total displacement ($D$; see methods) between pair-wise combinations of proportional simulated (green line), sampled (orange line) and sampling-standardised (purple line) diversity curves, for different types of latitudinal biodiversity gradient (a: unimodal- and bimodal-type; b: unimodal- and flat-type; c: flat- and bimodal-type). The smaller the difference, the more similar the curves.
Figure S11. Pearson’s correlation coefficient ($r$) between simulated and sampled/sampling-standardised diversity curves for each type of latitudinal biodiversity gradient: (a, d) flat-type; (b, e) unimodal-type; (c, f) bimodal-type. The greater Pearson’s $r$, the more similar sampled/sampling-standardised diversity curves are to their genuine (simulated) curve. Filled circles are statistically significant results ($P < 0.05$).

Figure S12. Pearson’s correlation coefficient ($r$) between pair-wise combinations of proportional simulated, sampled and sampling-standardised diversity curves, for different types of latitudinal biodiversity gradient: (a, b, c) unimodal- and flat-type; (d, e, f) unimodal- and bimodal-type; (g, h, i) flat- and bimodal-type. The greater Pearson’s $r$, the more similar the diversity curves. Filled circles are statistically significant results ($P < 0.05$).
Figure S13. Kolmogorov-Smirnov two-sample tests ($D$) between simulated and sampled/sampling-standardised diversity curves for each type of latitudinal biodiversity gradient: (a, d) flat-type; (b, e) unimodal-type; (c, f) bimodal-type. The greater the $D$ value, the more different sampled/sampling-standardised diversity curves are to their genuine (simulated) curve. Filled circles are statistically significant results ($P < 0.05$).
Figure S14. Asselian to Piacenzian (298.9–2.58 Ma) plot of the palaeolatitudinal bin with peak (maximum) diversity for simulated flat-type (a, b), unimodal-type (c, d) and bimodal-type (e, f) latitudinal biodiversity gradients. (a–f) Green points indicate the simulated known palaeolatitudinal bin with peak diversity; (a, c, e) orange points depict the sampled peak diversity palaeolatitudinal bin; (b, d, f) purple points indicate the sampling-standardised (classical rarefaction) peak diversity palaeolatitudinal bin.
Figure S15. Relationship between palaeolatitudinal richness and spatial sampling coverage. Relationship between simulated palaeolatitudinal richness and spatial sampling coverage for a flat- (a), unimodal- (d) and bimodal-type (g) latitudinal biodiversity gradient. Relationship between sampled palaeolatitudinal richness and spatial sampling coverage for a flat- (b), unimodal- (e) and bimodal-type (h) latitudinal biodiversity gradient. Relationship between sampling-standardised palaeolatitudinal richness and spatial sampling coverage for a flat- (c), unimodal- (f) and bimodal-type (i) latitudinal biodiversity gradient. The coefficient of determination ($R^2$) with associated $P$-values are depicted within each plot. Each point represents an individual stage and palaeolatitudinal bin, and is coloured by their respective geological period.
Figure S16. Relationship between palaeolatitudinal richness and spatial sampling extent (summed minimum-spanning tree [MST] length). Relationship between simulated palaeolatitudinal richness and spatial sampling extent for a flat- (a), unimodal- (d) and bimodal-type (g) latitudinal biodiversity gradient. Relationship between sampled palaeolatitudinal richness and spatial sampling extent for a flat- (b), unimodal- (e) and bimodal-type (h) latitudinal biodiversity gradient. Relationship between sampling-standardised palaeolatitudinal richness and spatial sampling extent for a flat- (c), unimodal- (f) and bimodal-type (i) latitudinal biodiversity gradient. The coefficient of determination ($R^2$) with associated $P$-values are depicted within each plot. Each point represents an individual stage and palaeolatitudinal bin, and is coloured by their respective geological period.
Figure S17. Global raw counts of sampled diversity from simulations of species’ distributions and richness (= 1000), conforming to a flat-type (a), unimodal-type (b), and bimodal-type (c) latitudinal biodiversity gradient. (a–c) Green points depict mean sampled richness at stage level based on the known sampling window (i.e. collections). The green ribbon depicts the 95% quantile value of sampled richness from simulations. Each stage, for each type of latitudinal biodiversity gradient, has a simulated richness of 1000 species, and any variation in the diversity curve is a result of the stage-level sampling window.
Figure S18. Relationship between global sampled richness and spatial sampling coverage and extent (summed minimum-spanning tree [MST] length) for different types of latitudinal biodiversity gradients. Relationship between sampled global richness and spatial sampling coverage for a flat- (a), unimodal- (b) and bimodal-type (c) latitudinal biodiversity gradient. Relationship between sampled global richness and global spatial sampling extent for a flat- (d), unimodal- (e) and bimodal-type (f) latitudinal biodiversity gradient. The coefficient of determination ($R^2$) with associated $P$-values are depicted within each plot. Each point represents an individual stage, and is coloured by their respective geological period.

References

1. Bivand RS, Pebesma EJ, Gómez-Rubio V. 2013 *Applied spatial data analysis with R*. Second edition. New York, U.S.A.: Springer. See http://www.asdar-book.org/.