



Frontiers of Biogeography 15.4

Article: Body size and abundance are decoupled from species richness in Australian marine bivalves

Authors: Matthew R. Kerr and John Alroy

Supplementary Text, Figures & Tables

Summary:

Text S1. A description of the richness estimator and analytical tests.

Figure S1. Diagrammatic representation of valve measurements taken for the study.

Figure S2. Plots of estimated richness against true richness in analytical tests performed on the geometric series richness index described in both the main text and the supplement.

Figure S3. Latitudinal body size trends for bivalve families in this study.

Figure S4. Relationships between the alpha gamin parameter and richness, total abundance, and latitude.

Figure S5. Results of a nearest neighbour analysis comparing the true median body size to generated null distributions for three individual sites.

Figure S6. Results of a nearest neighbour analysis comparing the true median body size to generated null distributions for three individual sites.

Figure S7. Results of a nearest neighbour analysis comparing the true median body size to generated null distributions for three individual sites.

Figure S8. Results of a nearest neighbour analysis comparing the true median body size to generated null distributions for three individual sites.

Figure S9. Latitudinal trend in true median body size generated by a nearest neighbour analysis.

Table S1. The strength of the latitudinal gradient in marine bivalves on the eastern coastline of Australia for several diversity metrics.

Table S2 The results of a factor analysis of the ten environmental variables used in the study, showing which latent variable each one is associated with.

Table S3. Latitudinal trends in body size for marine bivalve families on the eastern coastline of Australia.

Text S1

As discussed in the main text, we estimated species richness by assuming the geometric series distribution. To test the accuracy of this protocol, we carried out a simulation that assumed a finite species pool with counts generated from a highly randomised geometric series distribution. We chose the geometric series for three reasons. First, there is some evidence that the log series distribution might govern many data sets (Alroy 2015, Baldridge et al. 2016). However, it is not directly possible to simulate a count of unseen species by assuming the log series because the distribution does not directly specify the relative size of this class (meaning, it assumes community-wide richness is infinite). Second, there is also some evidence that the Poisson log normal distribution applies to real-world data sets (Connolly et al. 2014). This distribution approaches a randomised geometric series given a realistic sigma (standard deviation) parameter. Third, we wished to show that the new index is accurate when its assumptions are met.

We created 1,000 simulated data sets as follows. First, we drew a random log normal variate with a mean of $\ln 20$, exponentiated it, and rounded up to set the number of species in the pool (the richness value to be estimated). Second, for each species we drew a random number ranging from zero one to define the underlying p parameter that governs the geometric series. Third, for each species we used this p statistic to draw a count of individuals using the rgeom function in the R package stats.

We noted large and highly visible differences in performance amongst the four estimators. We projected lines of unity on plots of estimated against true richness in each case (Fig. S9). The residual standard error of the data points relative to the lines was 0.7096 for raw richness; 0.3130 for the geometric series index; 0.4842 for the jackknife, which is usually an underestimate; 0.5624 for Chao 1, which often provides an even lower underestimate but may randomly overestimate; and 0.5665 for the Poisson log normal estimator, which often produces high outliers. In other words, the new method is almost twice as accurate as two of the methods and also considerably better than the jackknife. Results for ACE and iNEXT are not detailed because estimates produced by both methods are very strongly correlated with Chao 1.

Bibliography

- Alroy, J. (2015). The shape of terrestrial abundance distributions. *Science Advances*, **1**, e1500082. <https://doi.org/10.1126/sciadv.1500082>
- Baldridge, E., Harris, D. J., Xiao, X. & White, E. P. (2016). An extensive comparison of species-abundance distribution models. *PeerJ*, **4**, e2823. <https://doi.org/10.7717/peerj.2823>
- Bulmer, M. G. (1974). On fitting the Poisson Lognormal Distribution to species-abundance data. *Biometrics*, **30**, 101-110. <https://doi.org/10.2307/2529621>
- Burnham, K. P. & Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, **65**, 625-633.
<https://doi.org/10.2307/2335915>
- Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, **11**, 265-270. <https://doi.org/10.2307/4615964>
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783-791. <https://doi.org/10.2307/2531532>
- Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K. & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45-67.
<https://doi.org/10.1890/13-0133.1>
- Chao, A. & Lee, S.-M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210-217.
<https://doi.org/10.1080/01621459.1992.10475194>
- Connolly, S. R., MacNeil, M. A. & Wilson, R. S. (2014). Commonness and rarity in the marine biosphere. *Proceedings of the National Academy of Sciences USA*, **111**, 8524-8529.
<https://doi.org/10.1073/pnas.1406664111>
- Fisher, R. A., Corbet, A. S. & Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42–58. <https://doi.org/10.2307/1411>
- Grøtan, V. & Engen, S. (2008). poilog: Poisson lognormal and bivariate Poisson lognormal distribution. R package version 0.4.

Oksanen, J., Guillaume Blanchet, F., Friendly, M. Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. & Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-6.
<https://CRAN.R-project.org/package=vegan>

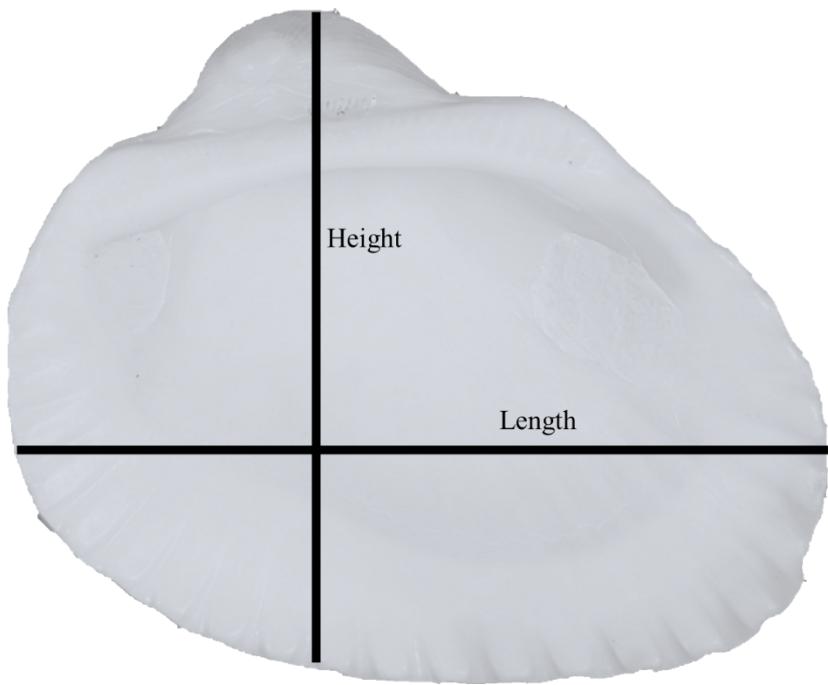


Figure S1. Shell diagram indicating measurements of length and height taken for this study. Shell is shown in lateral orientation.

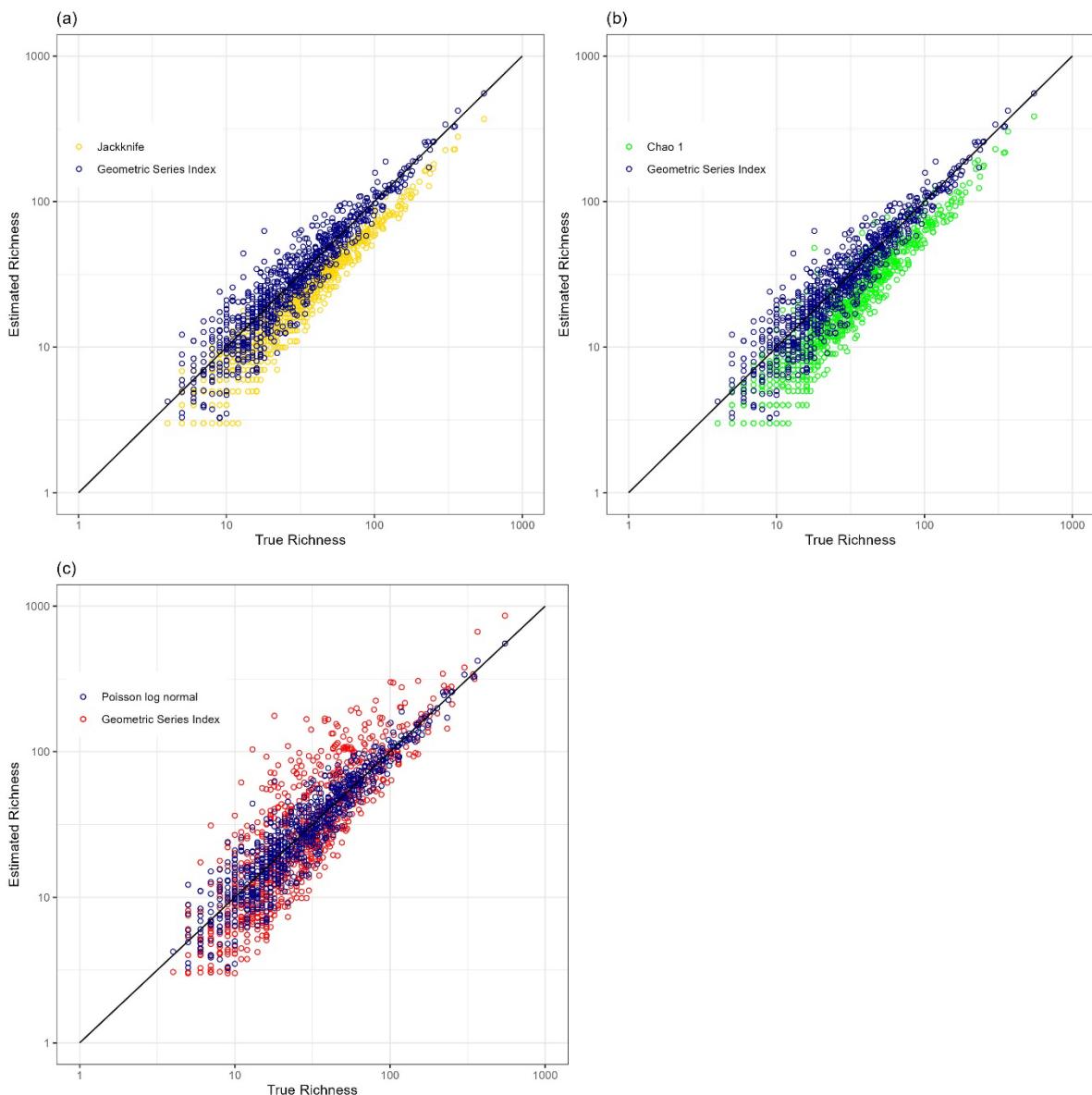


Figure S2. Relationships between estimated and true richness in species inventories generated by Monte Carlo simulation. The geometric series index (blue points) is generally accurate and the other estimators are not. See Text S1 for details of the procedure. Axes are logged and lines of unity are illustrated. (a) Additional estimates (orange points) are based on the first-order jackknife (Burnham and Overton 1978). (b) Additional estimates (green points) are based on Chao 1 (Chao 1984, 1987). (c) Additional estimates (red points) are based on fitting the Poisson log normal distribution (Bulmer, 1974).

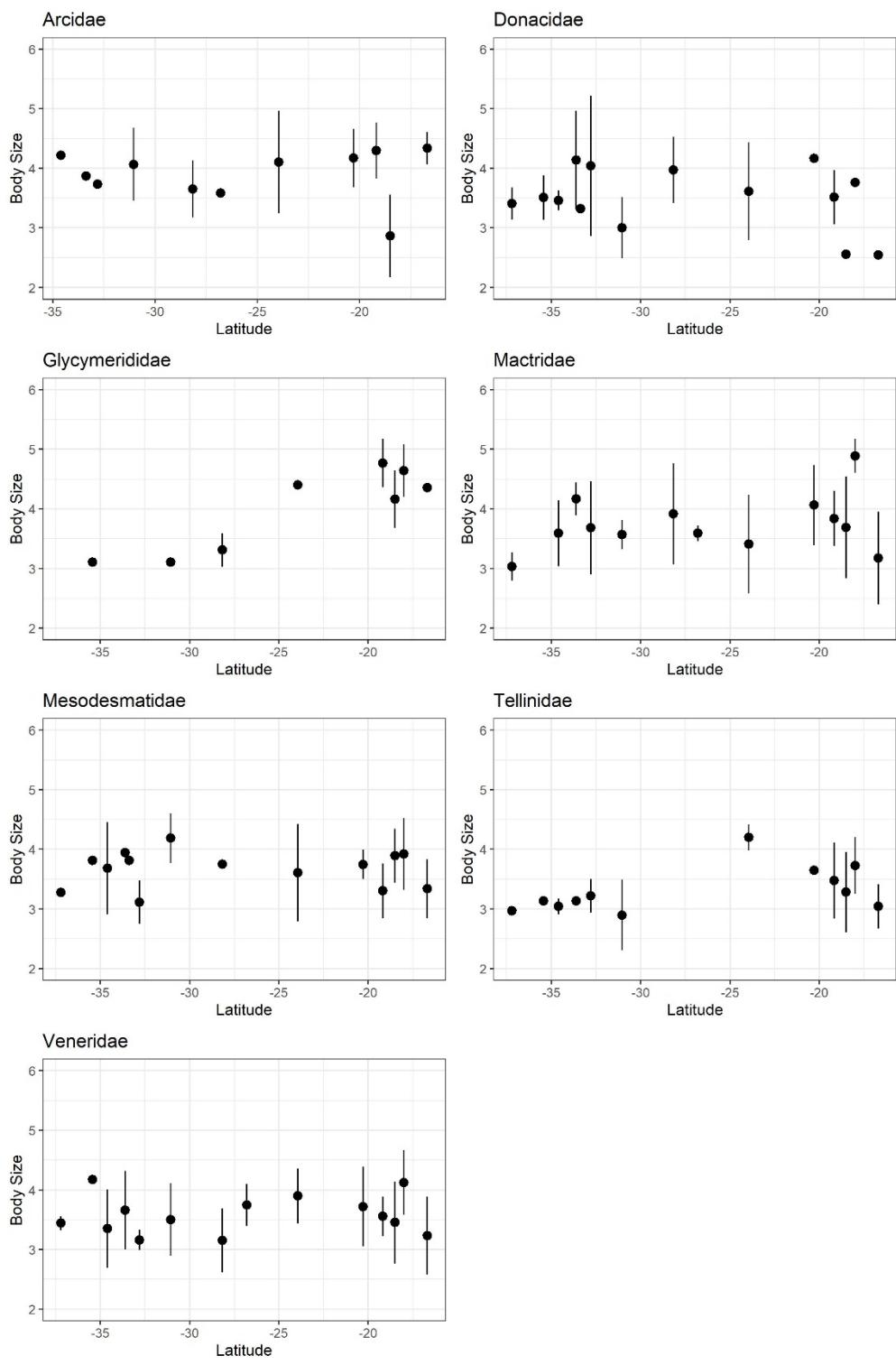


Figure S3. Latitudinal trends in mean body size along the eastern coastline of Australia for the seven largest families of marine bivalves. Body size is calculated as the geometric mean of length and height (see main text). No trend is statistically significant (Table S2).

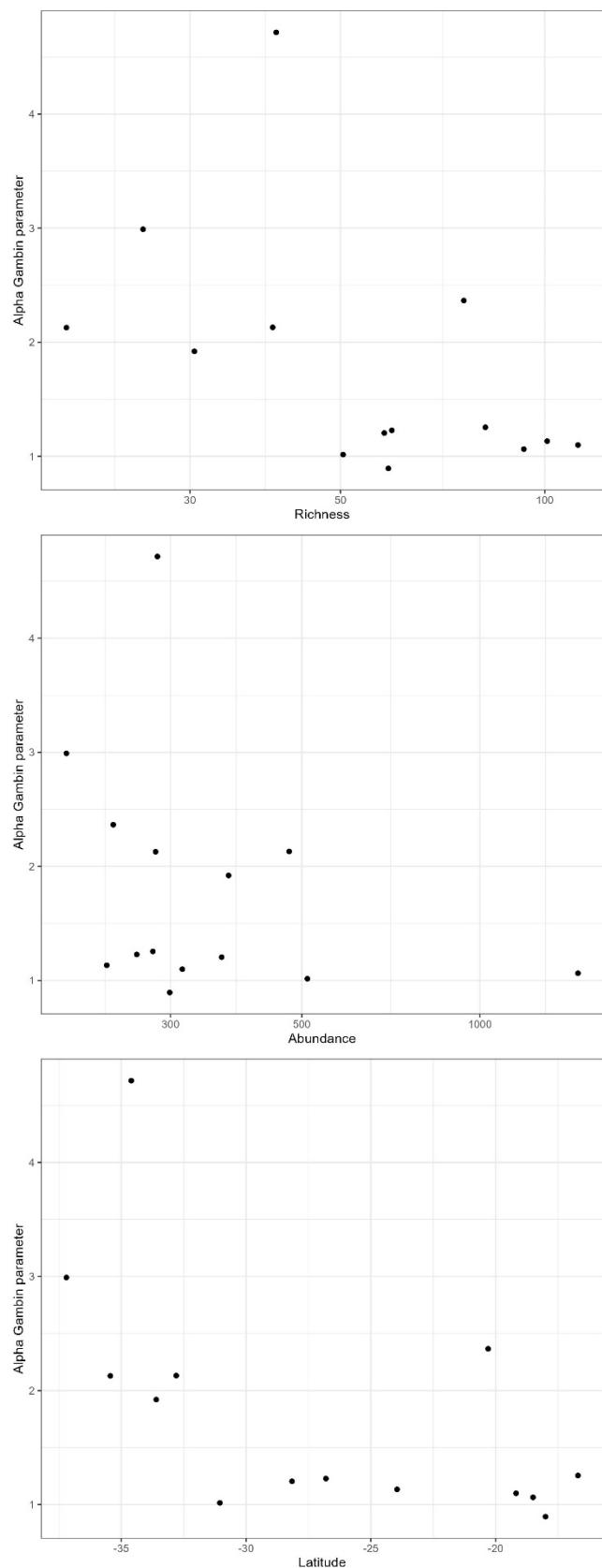
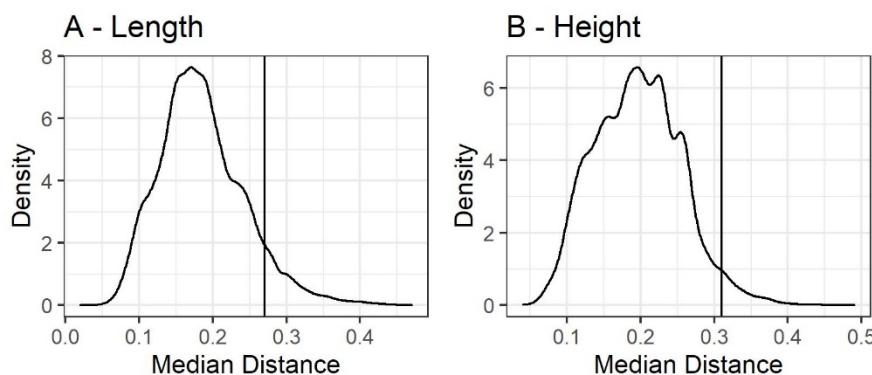
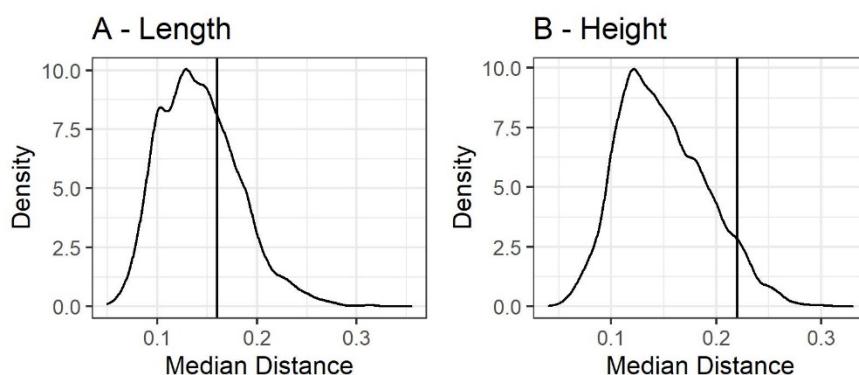


Figure S4. Relationship between the alpha gambin parameter and richness (a), the total number of individuals (b) and latitude (c). No trend is statistically significant.

Cannonvale (-20.3)



Cape Pall (-19.18)



Hat Head (-31.05)

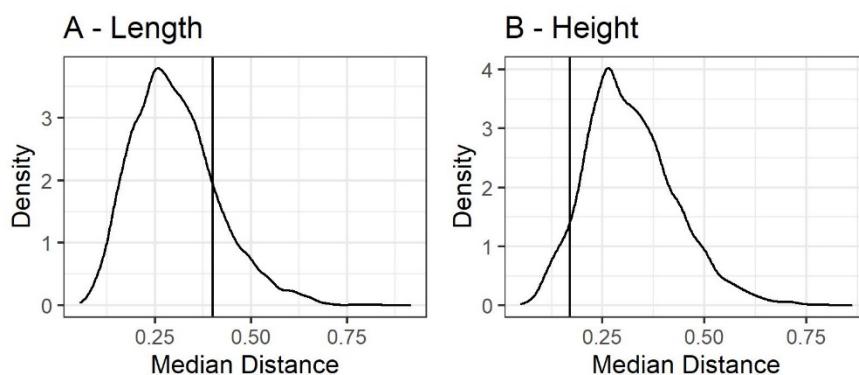
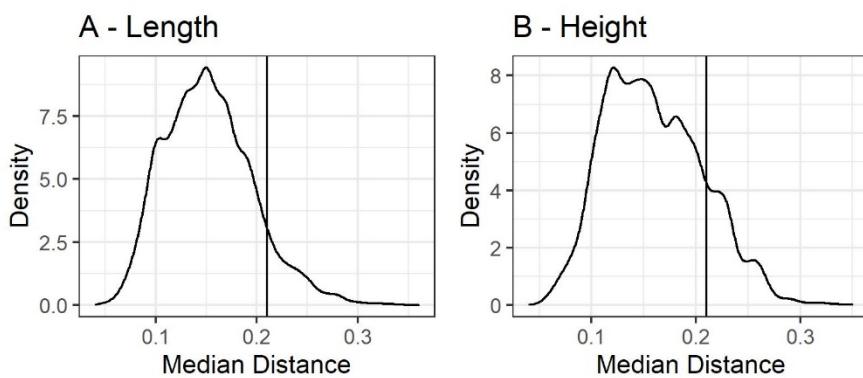
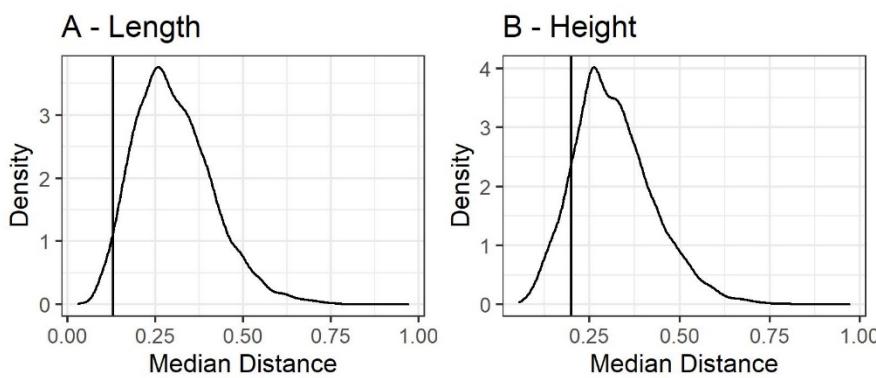


Figure S5. Results of a nearest-neighbour analysis of valve dimensions (A = length and B = height) of marine bivalves sampled from along the eastern coastline of Australia. The vertical line represents the true median of nearest-neighbour distances for the site. The probability density curve represents values resulting from a subsampling procedure and is generated from all bivalve measurements drawn from across the coastline. See text for full details. Site latitude is shown in parentheses.

Lucinda (-18.5)



One Mile (-32.8)



Palm Beach (-33.6)

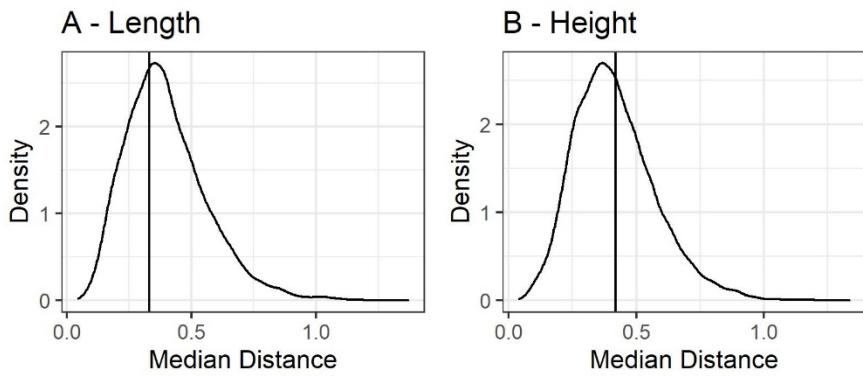
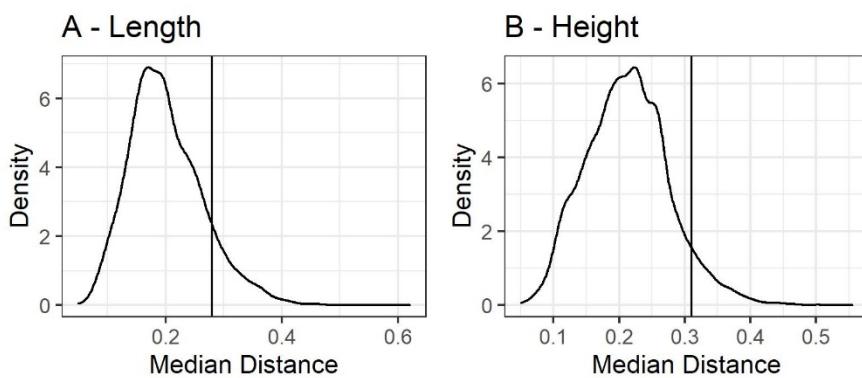
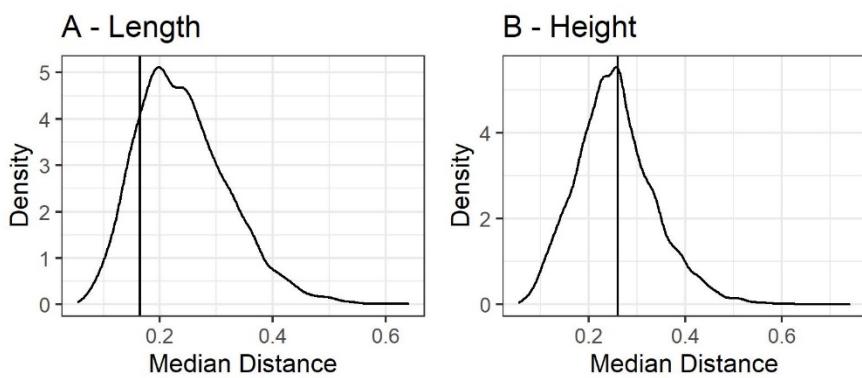


Figure S6. Results of a nearest-neighbour analysis of valve dimensions (A = length and B = height) of marine bivalves sampled from along the eastern coastline of Australia. The vertical line represents the true median of nearest-neighbour distances for the site. The probability density curve represents values resulting from a subsampling procedure and is generated from all bivalve measurements drawn from across the coastline. See text for full details. Site latitude is shown in parentheses.

Palm Cove (-16.7)



Rainbow (-28.16)



Saltwater (-37.2)

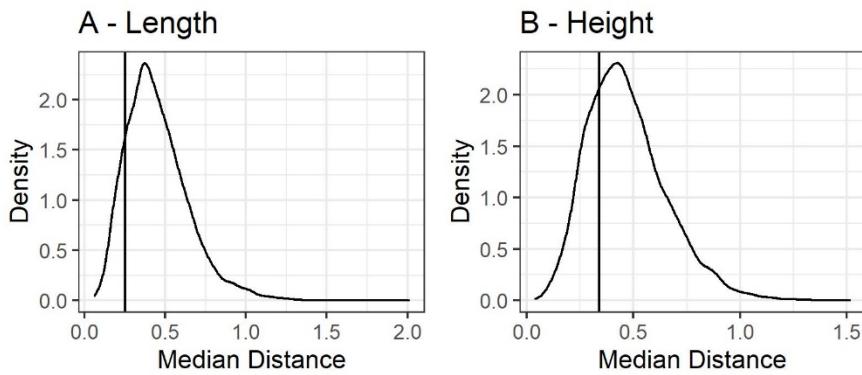
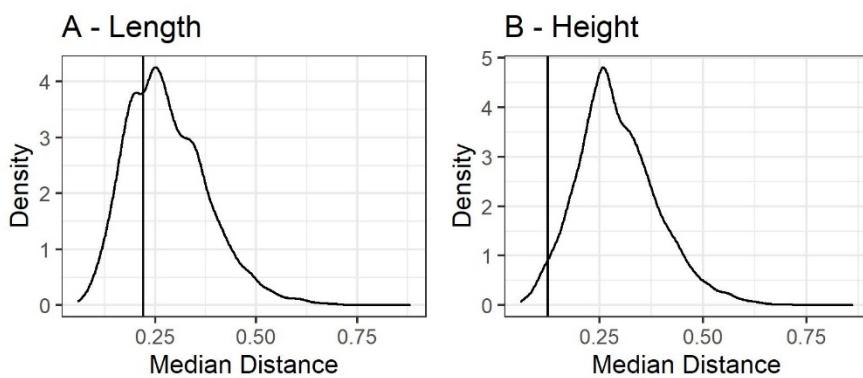
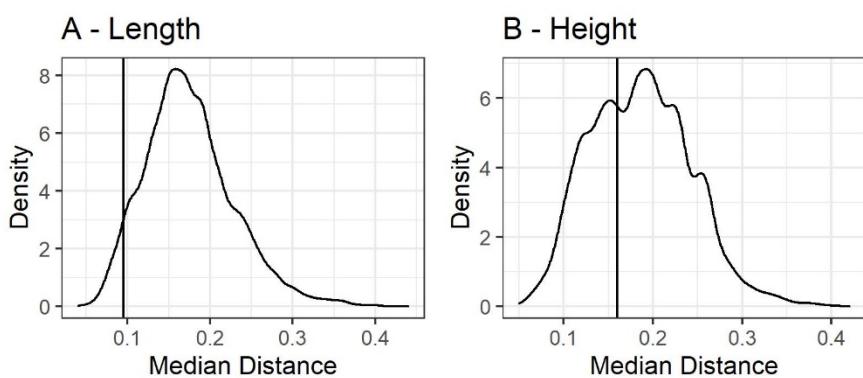


Figure S7. Results of a nearest-neighbour analysis of valve dimensions (A = length and B = height) of marine bivalves sampled from along the eastern coastline of Australia. The vertical line represents the true median of nearest-neighbour distances for the site. The probability density curve represents values resulting from a subsampling procedure and is generated from all bivalve measurements drawn from across the coastline. See text for full details. Site latitude is shown in parentheses.

Shellharbour (-34.6)



Tannum (-23.95)



Tully (-18)

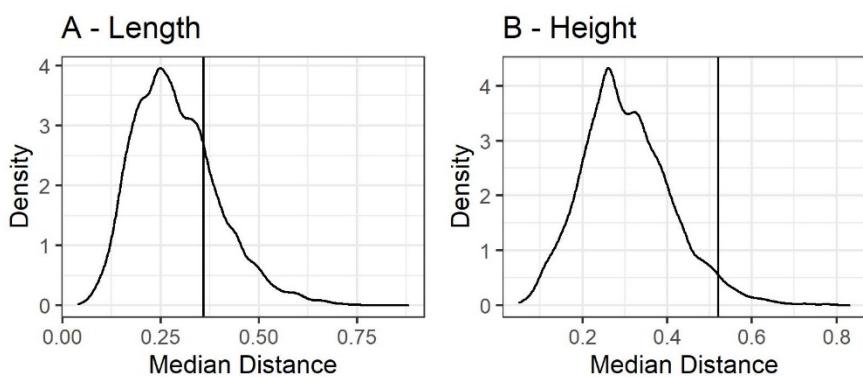


Figure S8. Results of a nearest-neighbour analysis of valve dimensions (A = length and B = height) of marine bivalves sampled from along the eastern coastline of Australia. The vertical line represents the true median of nearest-neighbour distances for the site. The probability density curve represents values resulting from a subsampling procedure and is generated from all bivalve measurements drawn from across the coastline. See text for full details. Site latitude is shown in parentheses.

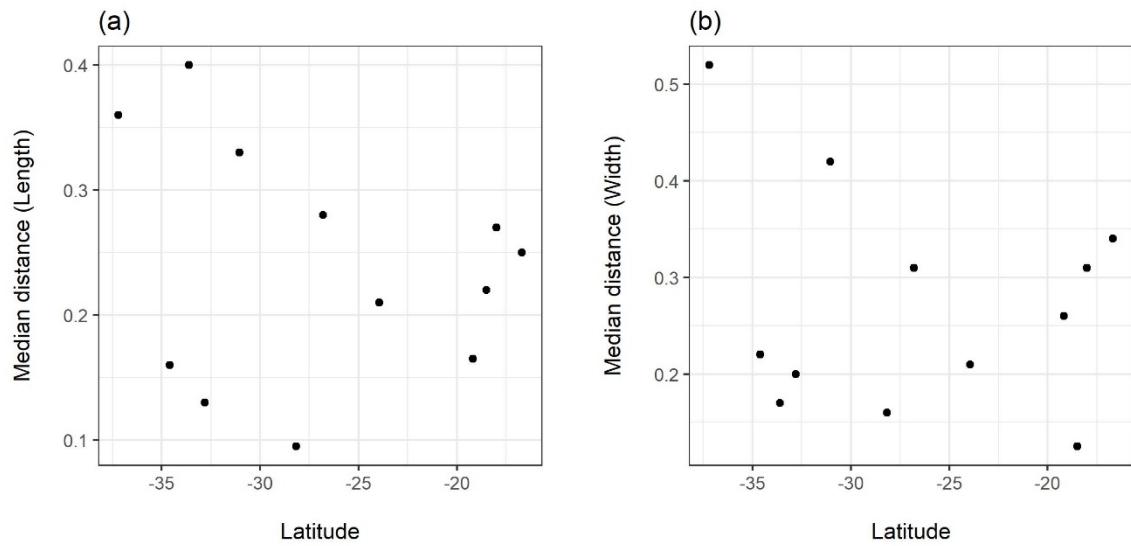


Figure S9. Latitudinal distribution of the true median retained in a nearest-neighbour analysis. Smaller values represent sites with a lower median distance between a shell measurement and the closest measurement from other shells at the same site. Results are shown for two valve dimensions (A = length and B = height).

Table S1. The strength of the latitudinal gradient in marine bivalves on the eastern coastline of Australia. Shown are Spearman's rank correlations (ρ) and slopes of logged diversity against latitude. Five different diversity measures are shown: Chao1 (Chao, 1984), S^2/m (Alroy, 2020), Fisher's α (Fisher *et al.*, 1934), Simpson's D (Simpson, 1949) and raw richness. **Bolded** values are significant ($p < 0.01$).

| Method | ρ | Slope |
|-------------------|--------------|--------------|
| Chao1 | 0.764 | 2.543 |
| S^2/m | 0.818 | 2.180 |
| Fisher's α | 0.764 | 0.058 |
| Simpson's D | 0.571 | 0.809 |
| Raw richness | 0.776 | 1.441 |

Table S2. The results of a factor analysis of the ten environmental variables used in the study. Each factor was assigned to a latent variable based on which factor it loaded most strongly onto (bolded).

| | Latent1 | Latent2 | Latent3 |
|-----------------------|---------------|--------------|--------------|
| Mean Temperature | -0.971 | -0.0257 | 0.0733 |
| SD Temperature | 0.38 | 0.622 | -0.185 |
| Mean Salinity | 0.00152 | 0.811 | -0.0248 |
| SD Salinity | -0.621 | -0.245 | 0.441 |
| Mean Dissolved Oxygen | 0.966 | -0.212 | -0.0815 |
| SD Dissolved Oxygen | -0.156 | 0.727 | 0.213 |
| Mean Nitrate | 0.824 | 0.157 | -0.188 |
| SD Nitrate | 0.76 | 0.1 | -0.23 |
| Mean Phosphate | 0.84 | -0.296 | 0.395 |
| SD Phosphate | 0.927 | 0.124 | -0.0306 |
| Mean Silicate | -0.103 | 0.059 | 0.967 |
| SD Silicate | 0.711 | 0.424 | 0.35 |

Table S3. Spearman's rank correlations (ρ) between latitude and median body size for the largest seven families of marine bivalves on the eastern coastline of Australia, with associated p -values. Body size is calculated as the geometric mean between length and height (see Materials and Methods). Family trends are shown visually in Fig. S2.

| Family Name | ρ | p value |
|----------------|--------|-----------|
| Arcidae | 0.182 | 0.595 |
| Donacidae | -0.090 | 0.762 |
| Glycymerididae | 0.683 | 0.062 |
| Mactridae | 0.236 | 0.437 |
| Mesodesmatidae | 0.048 | 0.869 |
| Tellinidae | 0.504 | 0.094 |
| Veneridae | 0.086 | 0.773 |