APPENDIX S2 - DATA ROBUSTNESS

A. RELATIONSHIP BETWEEN GENUS AND SPECIES RICHNESS (FIG. S1-S2)

B. TEST OF GEOGRAPHICAL SAMPLING BIAS (FIG. S3)

C. ADDTIONAL REFERENCES

A. RELATIONSHIP BETWEEN GENUS AND SPECIES RICHNESS

The use of genus richness as a proxy for species richness is well-established tool when species level data are scarce, unreliable, or taxonomically unstable (Balmford *et al.*, 1996). Whilst there are limitations in the precision of estimates of species richness from genus richness, at broad scales this issue is unlikely to affect general conclusions. Moreover, the genus level appears to more closely represent species diversity than higher level taxa such as family (Balmford *et al.* 1996, Kallimanis *et al.*, 2012).

To ensure such assumptions were met for our data, we tested the relationship between species and genus richness empirically. We used two subsets of the data for which species level data were available to test the percent of variation within species richness across sites that is captured by genus richness. The first subset included all records within the Order Bryopsidales (*n* genera = 27; *n* species = 412) across all sites in the original database. This was the only order within the original data for which species lists were recorded. The second subset consisted of the sites newly added to the database in 2013 (*n* sites = 20). For both of these subsets, we used a linear regression to determine the proportion of the variation within species richness that could be explained by genus richness (i.e., the adjusted R^2 value). To ensure data met the assumptions of linear regression (i.e., normally distributed residuals) we square-root transformed both genus and species richness.

Within the Bryopsidales, genus richness captured a highly significant proportion of the variation within species richness (adj-R^2 = 0.84, *p* <0.001, Fig. S1). The amount of variation in species richness that was captured by genus richness was also highly significant amongst the new sites subset (adj-R^2 = 0.84, *p* <0.001; Fig. S2). Therefore, our use of the data at the genus level is a valid approach to understanding diversity in a broad sense.
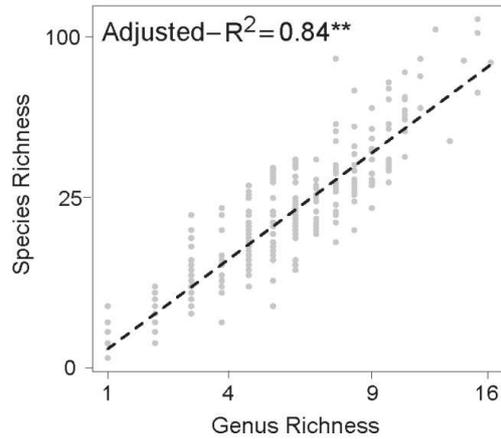
Fig. S1. Relationship between species and genus richness for original sites within the Order Bryopsidales. The linear regression is highly significant (p <0.001; dashed regression line plotted) and demonstrates the genus richness captures 84% of the variation in species richness amongst sites. ** = significant to 1% alpha level. Note that richness is plotted on a square-root scale.
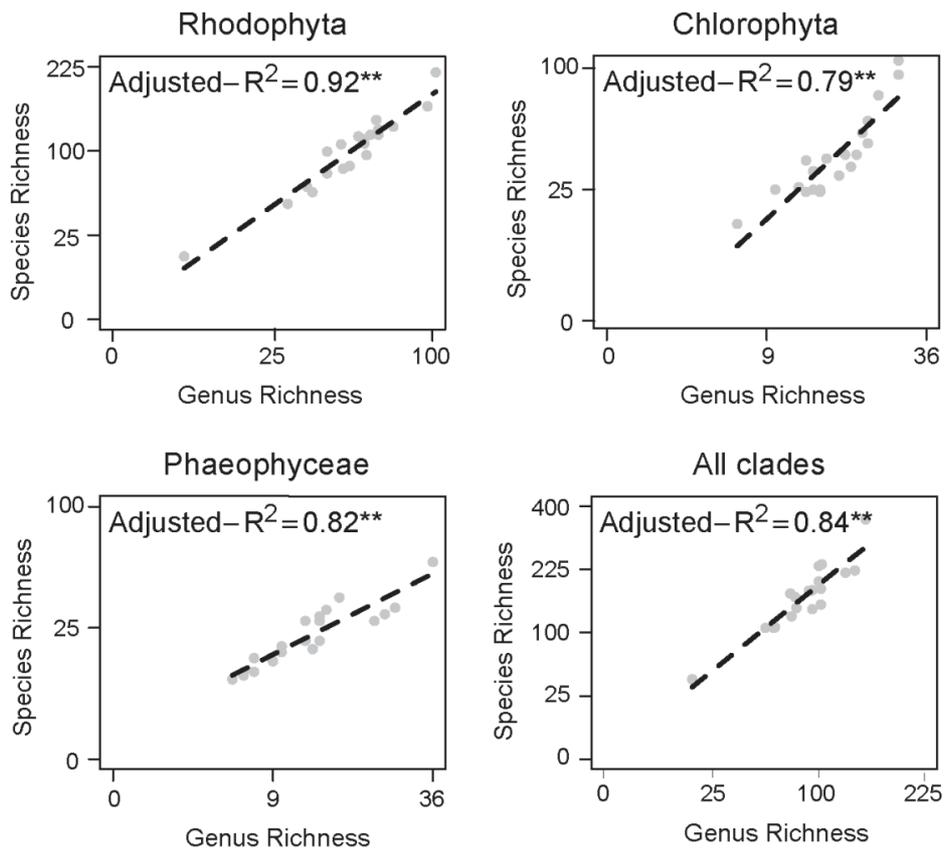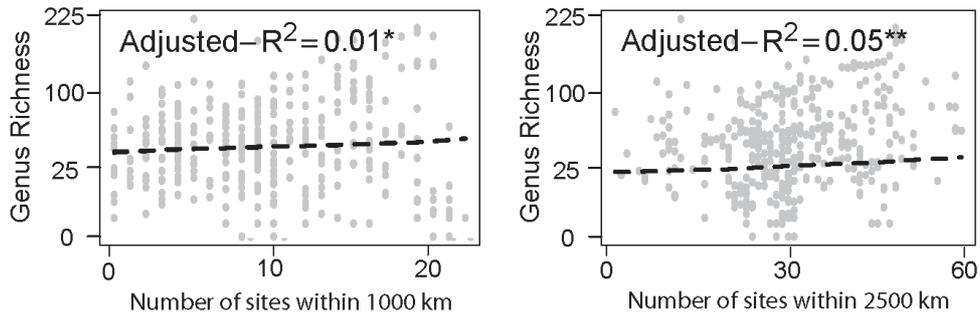

Fig. S2. Relationship between species and genus richness for new sites added to the database in 2013. The linear regression is highly significant for all clades separately and in combination ** = significant to 1% alpha level. Regression lines are plotted as dashed. Note that richness is plotted on a square-root scale.
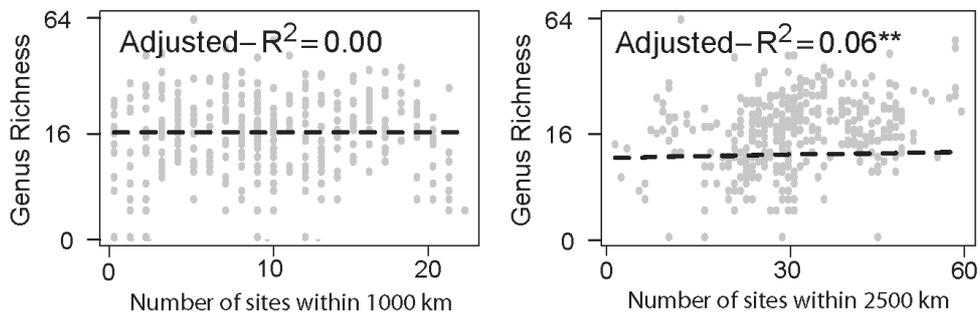
B. TEST OF GEOGRAPHICAL SAMPLING BIAS

Geographical sampling bias is an important issue that can confound analysis of large databases collated from multiple studies (Boakes *et al.*, 2010). The issue is two-fold. Firstly, the intensity of sampling tends to vary systematically through space due to accessibility of, or interest in, particular regions (Rondinini *et al.*, 2006). Secondly, different studies invariably use different methodologies to collect data, leading to small-scale disparity in sampling effort. Geographical variation in sampling intensity across regions can be associated with biased estimates in recorded richness (Gotelli & Colwell, 2001). This is more likely to lead to biased estimates when multiple surveys are pooled within a given area, or when estimated range overlap is used to calculate richness, both of which are common practice within biogeographical analyses. Kerswell (2006) used the range overlap method, and tested for this type of sampling bias within the data set by determining whether the richness estimates within a biogeographical region could be explained by the number of sites within that region: no relationship was found (see Kerswell 2006 Appendix B). Our analyses instead used only those genera recorded as present at that site; that is, we did not infer presences for genera whose ranges overlapped a site but that were not observed at that site. Consequently, we would expect our results to be less sensitive to sampling biases than studies using the approaches described above.

Nevertheless, to confirm this, we used linear regression to determine the extent to which genus richness for a given site could be predicted by the number of sites within a given proximity (1000 km and 2500 km radii). We find the influence of sampling intensity at this scale to be negligible despite statistical significance, accounting for no more than 6% (typically only 1%) of the total variation in genus richness amongst sites (Fig. S3). Moreover, this conclusion was not sensitive to the proximity used. These findings indicate that the patterns we have identified are unlikely to be sensitive to differential sampling intensity.

(a) Rhodophyta

Left panel: Adjusted−R^2=0.01*
Y-axis: Genus Richness (0, 25, 100, 225)
X-axis: Number of sites within 1000 km (0, 10, 20)

Right panel: Adjusted−R^2=0.05**
Y-axis: Genus Richness (0, 25, 100, 225)
X-axis: Number of sites within 2500 km (0, 30, 60)

(b) Chlorophyta

Left panel: Adjusted−R^2=0.00
Y-axis: Genus Richness (0, 16, 64)
X-axis: Number of sites within 1000 km (0, 10, 20)

Right panel: Adjusted−R^2=0.06**
Y-axis: Genus Richness (0, 16, 64)
X-axis: Number of sites within 2500 km (0, 30, 60)

(c) Phaeophyceae

Left panel: Adjusted−R^2=0.01*
Y-axis: Genus Richness (0, 25, 100)
X-axis: Number of sites within 1000 km (0, 10, 20)

Right panel: Adjusted−R^2=0.01**
Y-axis: Genus Richness (0, 25, 100)
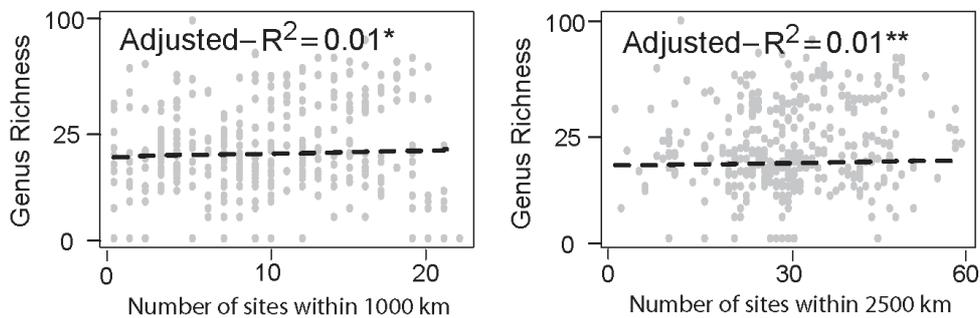X-axis: Number of sites within 2500 km (0, 30, 60)

Fig. S3. The effect of sampling bias presented as the relationship between the number of sites in a given proximity (1000 km, 2500 km) and the genus richness of each clade for a given site. Percent variation accounted for varies between 0% and 6%. * = significant to the 5% alpha level; ** = significant to the 1% alpha level.

C. ADDTIONAL REFERENCES

Balmford, A., Green, M.J.B. & Murray, M.G. (1996) Using Higher-Taxon Richness as a Surrogate for Species Richness: I. Regional Tests. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **263**, 1267-1274.

Boakes E.H., McGowan P.J.K., Fuller R.A., Chang-qing D., Clark N.E., O'Connor K., Mace G.M. (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biology,* **8**, e1000385.

Kallimanis, A.S., Mazaris, A.D., Tsakanikas, D., Dimopoulos, P., Pantis, J.D. & Sgardelis, S.P. (2012) Efficient biodiversity monitoring: Which taxonomic level to study? *Ecological Indicators*, **15**, 100-104.

Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H. & Possingham, H.P. (2006) Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, **9**, 1136-1145.