Appendix 3. Variation in grazer effect sizes with experimental methods.

Method of excluding grazers

While all experiments involved the exclusion of grazers *in situ*, they did differ in their starting conditions, methods of grazer exclusion, plot sizes and duration. Across all habitats, the effect of excluding grazers was lower in the 38% of experiments in which plots had been cleared of primary producers at the start of the experiment, in contrast to those that left the producers untouched (Fig. S4a, $F_{1,616} = 8.31$, $P = 0.004$). The starting conditions however, were unevenly distributed among habitats ($\chi^2 = 25.6$, df = 3, $P < 0.001$), with the experiments in the rocky intertidal being more likely to clear the substrate prior to commencement, and differences among starting conditions only evident within this habitat (Fig S4a).

The majority of exclusion experiments (74%) used cages or physical barriers to manipulate the density of herbivores. 16 % of experiments used physical removal, while 10% used chemical deterrents (barriers of copper-based paint or biocides). The methods of exclusion were unevenly distributed amongst types of herbivores. For experiments that manipulated molluscs, chemical deterrents and exclusion cages resulted in larger effect sizes than did physical removal (Fig. S4b, $F_{2,291} = 12.08$, $P < 0.002$). These results could arise from differences in densities among the treatment types (e.g., partial recolonisation following removal by hand) or differences in the species for which each treatment was applied (e.g., physical removals more likely for large species). For urchins, there was no difference in effect sizes between experiments using exclusion cages and those using physical removal ($F_{1,81} = 2.31$, $P = 0.13$), and for crustaceans, there was no difference between experiments using chemical deterrents and exclusion cages ($F_{1,48} = 0.52$, $P = 0.47$) (Fig. S4b).

We were unable to compare the efficiency of exclusion methods as these were not uniformly reported across habitats or taxa (e.g., assumed to be 100% for cages excluding large grazers). We were also unable to test for the possibility that the exclusion of fish from plots resulted in compensatory increases in the abundance of small herbivores within cages. We did, however, exclude experiments where the authors identified clear experimental artefacts from their exclusion techniques.

Thirty-one % of experiments used procedural controls in an attempt to estimate artefacts associated with exclusion methods (e.g., half-caged plots that allow access to grazers but mimic the possible artefacts due to shading or changes to water flow). With these experiments, the effect size calculated by contrasting exclusion plots with procedural controls was a strong predictor of the effect size calculated with exclusion plots and un-manipulated controls ($F_{1,398} = 1147$, $P < 0.001$).

The slope of this relationship (0.95) did not differ significantly from one (95% confidence interval: 0.89–1.01) suggesting that artefacts associated with exclusion cages have little effect on the outcome of grazer exclusion experiments (if we assume that herbivores had full access to the procedural controls).

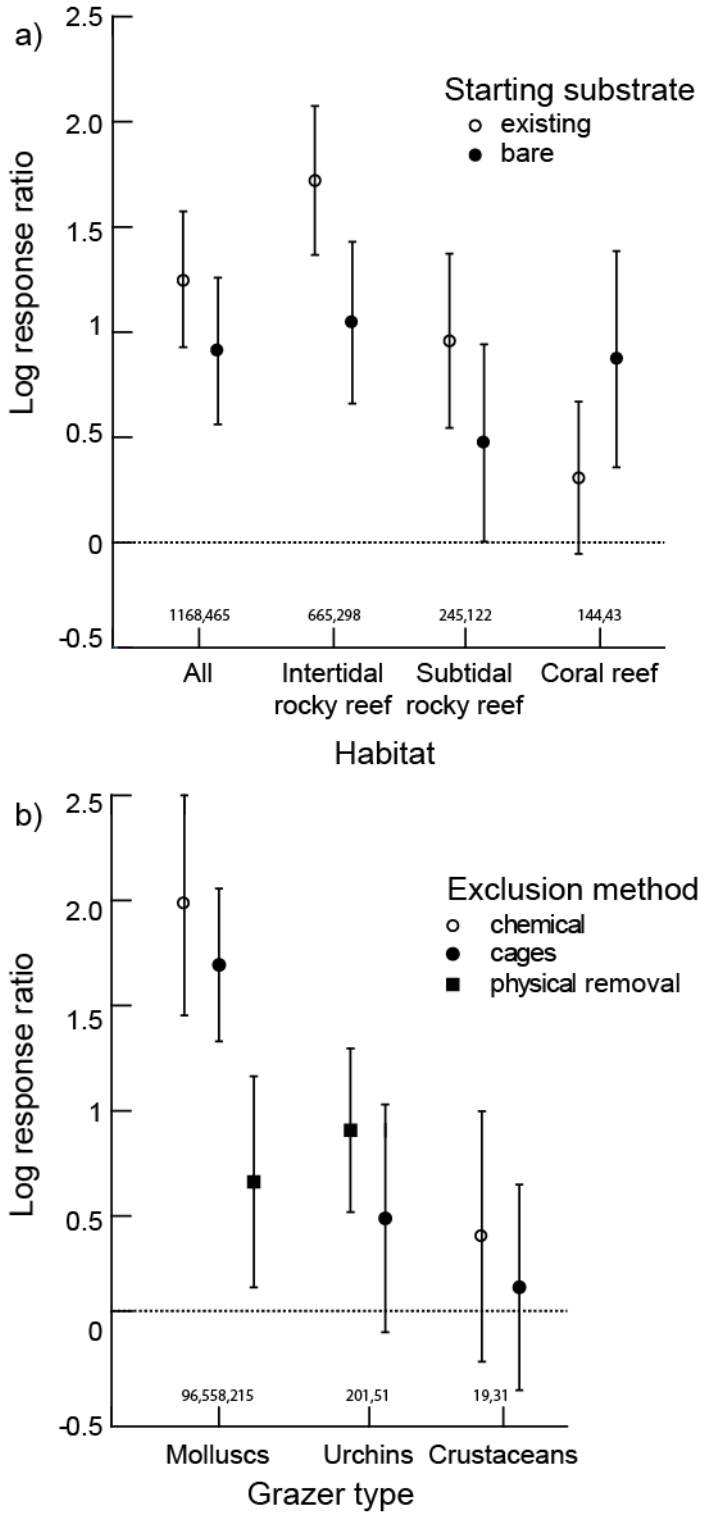


Fig. S4 The variation in the effects of excluding grazers with differences in experimental methodology. In a), for each major habitat type, effects sizes are contrasted between experiments that were conducted with existing communities and those where the substrate was cleared prior to the experiment starting. In b), for three major groups of grazers, effects sizes are contrasted among methods of exclusion (cages vs chemical deterrents vs physical removal). Means and 95% confidence intervals are REML estimates from linear mixed models of predictor variables with experiment as a random factor. The dotted line at zero is the effect size expected if there is no effect of removing herbivores. The number of observations per mean is given above the x axis.

The plot size used in experiments did not vary significantly among the four major habitats in the data set, ANOVA, $F_{3,561} = 2.02$, $P = 0.11$, ln transformed data). Across all habitats, there was a very weak positive relationship between effect size and plot size (Fig. S5, Table S1, $F_{1,386} = 5.17$, $P = 0.02$, $R^2 = 0.01$).

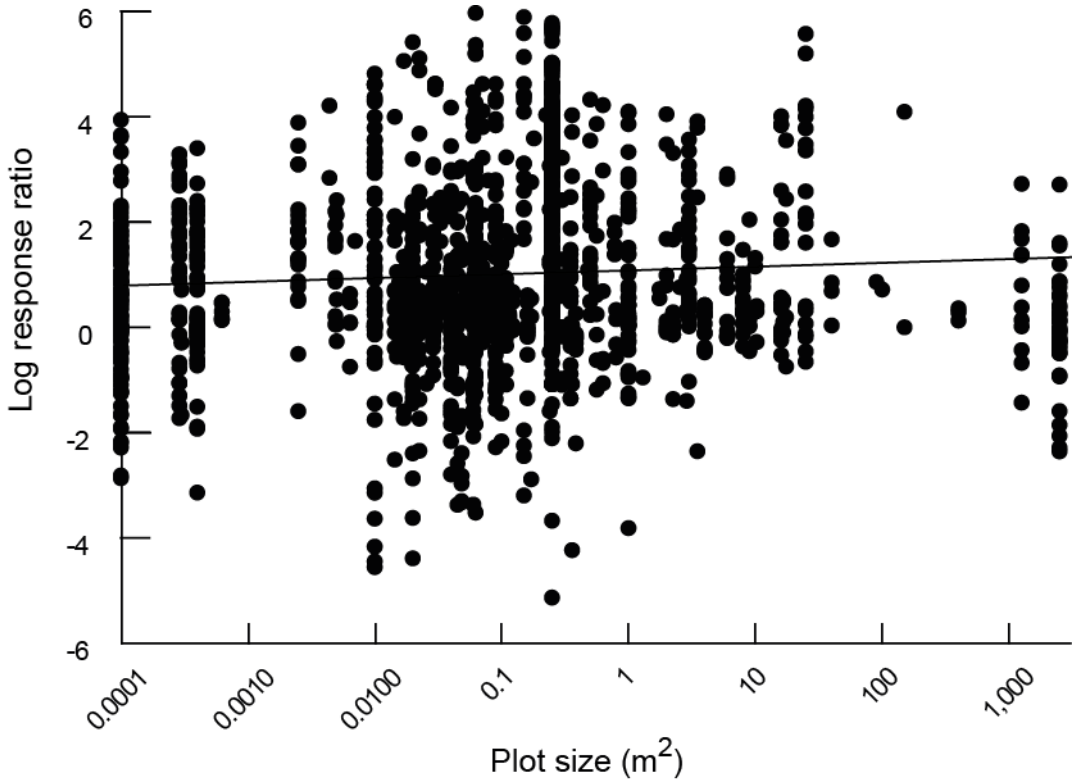


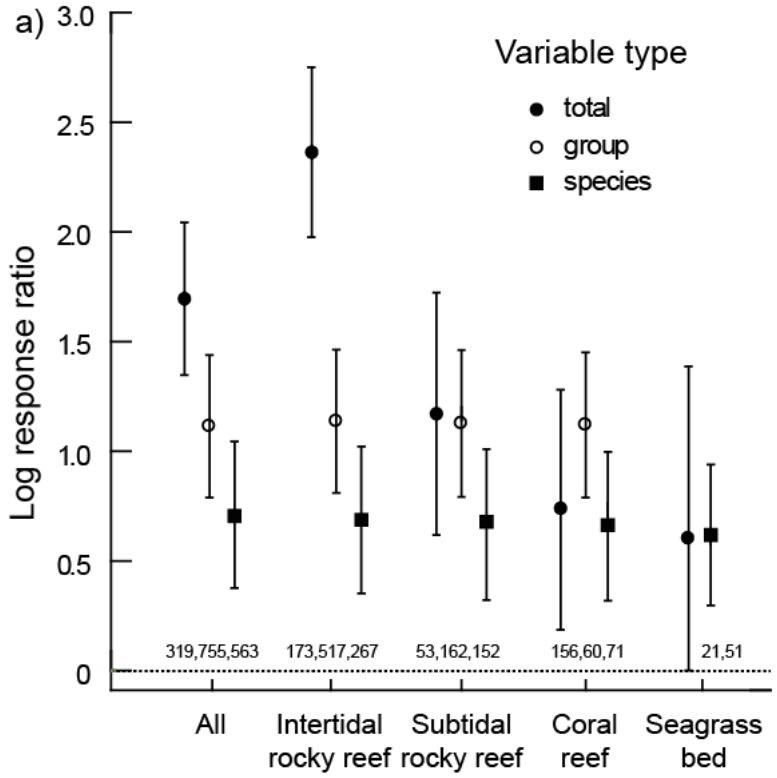
Fig. S5. The relationship between the effect of excluding grazers and plot size.

Methods of measuring producer biomass

Differences between exclusion and control plots were largest when using measures of total abundance in contrast to measures from single species or groups of species (Fig. S6a, $F_{2,1522} = 31.24$, $P < 0.001$). This effect was due to the high impact seen in measures of total abundance from intertidal rocky reefs only, with little difference among types of measurement in the other habitats (Fig. S6a).

A majority of studies used percentage cover to measure the abundance of primary producers (67 % of observations), with 13 % using measures of biomass, 12 % counts per unit area, and 5 % measures of size. With the types of measures used being unevenly distributed among habitats (e.g., nearly all seagrass studies used biomass and most intertidal studies used % cover), we contrasted the types of measures within habitats for those where there were more than five replicate

experiments per type. Within intertidal and subtidal rocky reefs, measures of % cover and biomass were higher than counts, and within subtidal rocky and coral reefs, measures of biomass were higher than % cover and counts (Fig. S6b,Table S1).



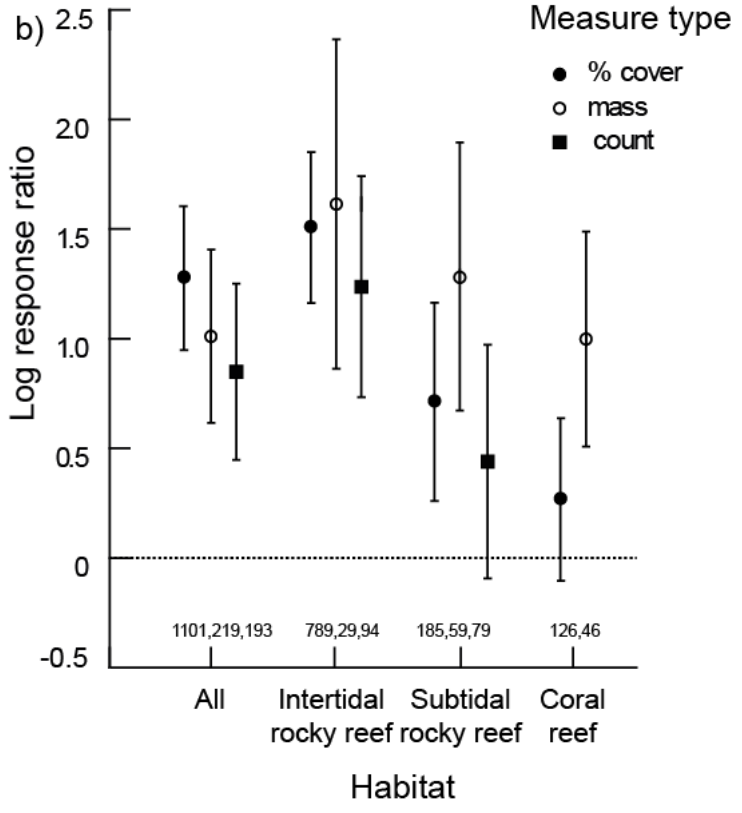


Fig. S6. Variation in the effects of excluding grazers among the different ways in which primary producers were measured. In a) variables are categorised as total (e.g., all biomass in a plot), group (a subset of producers in a plot, e.g., crustose coralline algae), and individual species. In b), variables are categorised as measures of percentage cover, measures of biomass, and densities of individuals per plot. Error bars, dotted line and sample sizes are as explained in Fig. S4.

Experimental duration

Across all habitats, the effect of the duration of the experiment was dependent on the type of measure taken (Fig. S7). A longer experimental duration was associated with a larger effect of excluding grazers when measures were of total abundance and of groups of species, but not for single species (Fig S7, Table S1). For individual species, a longer experimental duration was associated with larger variance in effect sizes (positive or negative) (Fig S7b).

The significant effects of latitude (Fig 1a), temperature (Fig 1b), nitrate (Fig 1c) and phosphate on the effect of excluding grazers (Table S2) were unaltered by the addition of duration as a covariate in the linear mixed models ($P < 0.02$ for all analyses).

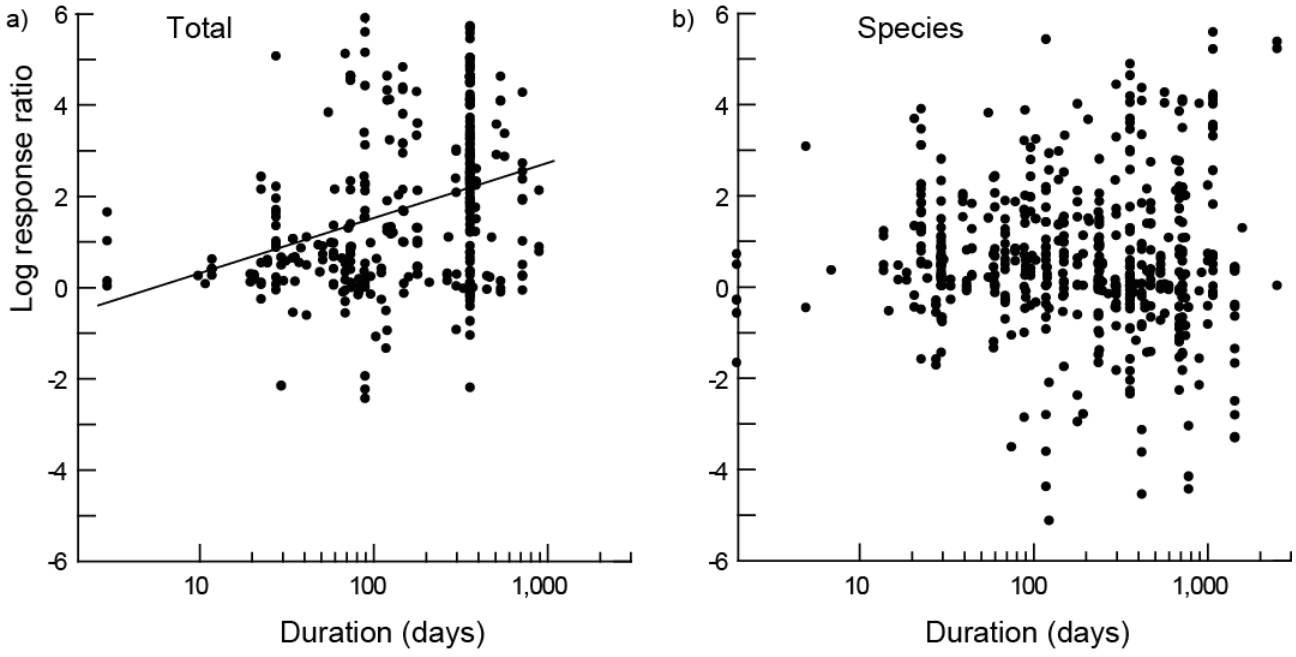


Fig. S7 The relationships between the effect of excluding grazers and experimental duration for measures of total producer abundance per plot (a) and measures of individual species (b).