

Supplementary Materials

Figure S1. Species rarity. In the main body of the manuscript we present results for how rarity (log-modulo of skewness; **right**) scales with sample abundance, i.e., the number of individual organisms or gene reads in a sample (N) (see Fig. 1a). The log-modulo transformation adds a value of one to each measure of skewness and converts negative values to positive values, making them all positive and able to be log-transformed. The analysis showed similar scaling but a greater intercept for microbes, revealing greater rarity. We also quantified rarity as a logarithmically transformed measure of skewness, as derived in Magurran and McGill (2011)³² (Measuring Species diversity) (**left**), however, this relationship which also showed increasing rarity (as decreasing log-skew) was substantially weaker than the relationship based on the log-modulo transformation of skewness. Consequently, we used the log-modulo measure in the main body and for the main result.

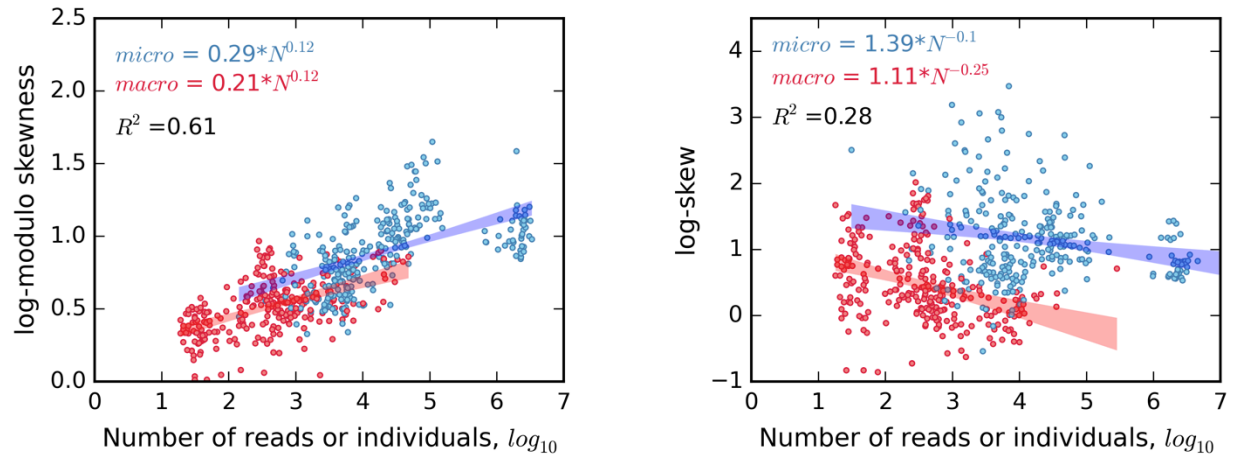


Figure S2. Dominance. In the main body of the manuscript we present results for the abundance or number of individuals or gene reads belonging to the most abundant species (N_{max}) or species-level taxonomic group scales with sample abundance, i.e., number of individual organisms or gene reads detected (N) (see Fig. 1b). We observed strong and largely similar scaling slopes for microbes and macrobes, and because N_{max} is an absolute measure of dominance and because the relationship is nearly isometric (i.e. nearly 1:1), we would expect no relationship for relative measures of dominance ($D \sim N_{max}/N$), which is supported here by regressions having little statistical strength, but where the data are not homogenously spread across the y-axis.

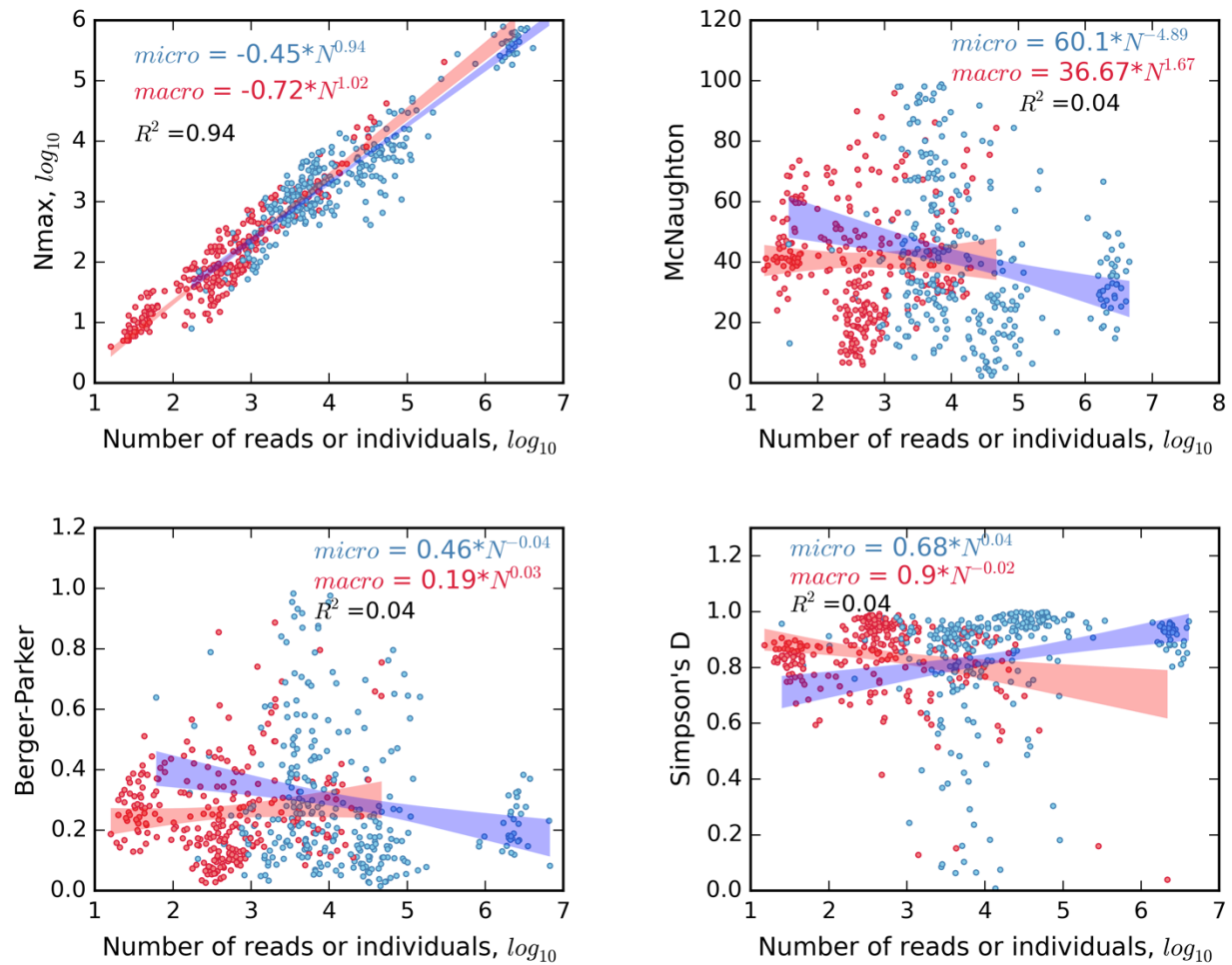


Figure S3. Species evenness. In the main body of the manuscript we present results for how similarity in abundance (i.e. evenness) relates to sample abundance, i.e., number of individual organisms or gene reads detected (N) (see Fig. 1c). We observed similar slopes microbes than macrobes, as also seen here, when using Heip's evenness index and Smith and Wilson's O evenness index. Slopes differ more greatly when using Smith and Wilson's evenness index, which gives greater weight to rare species than does Simpson's evenness.

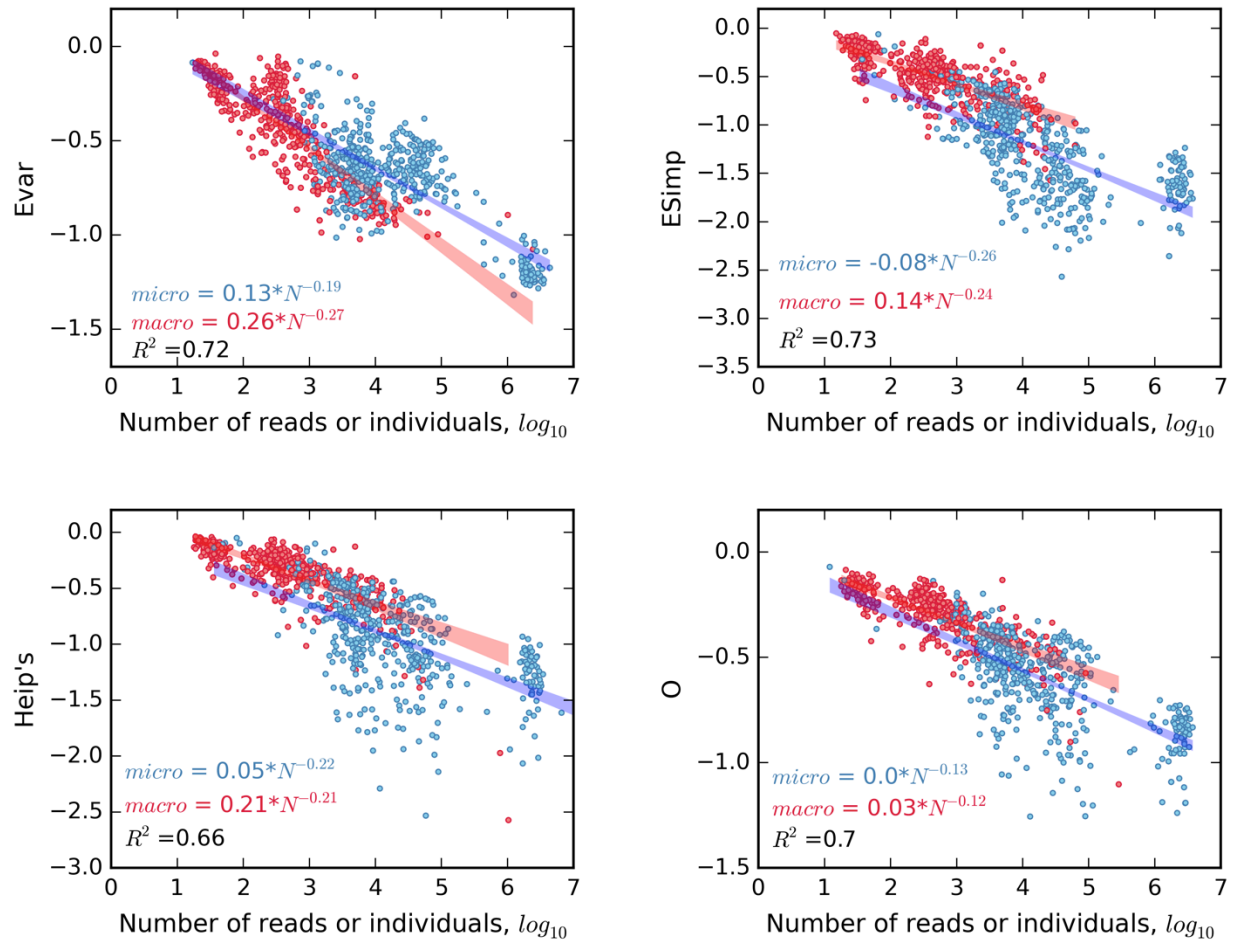


Figure S4. Species Richness. In the main body of the manuscript we present results for how observed numbers of species or species level taxonomic units (for microbes) relate to sample abundance, i.e., number of individual organisms or gene reads detected (N) (see Fig. 1d). We observed a steeper relationship and stronger scaling for microbes than macrobes. These results were qualitatively similar to estimates of richness: Chao1, ACE, Jackknife1, and Margalef's. These additional results reveal the same qualitative pattern and for all but Margalef's index, the same quantitative result.

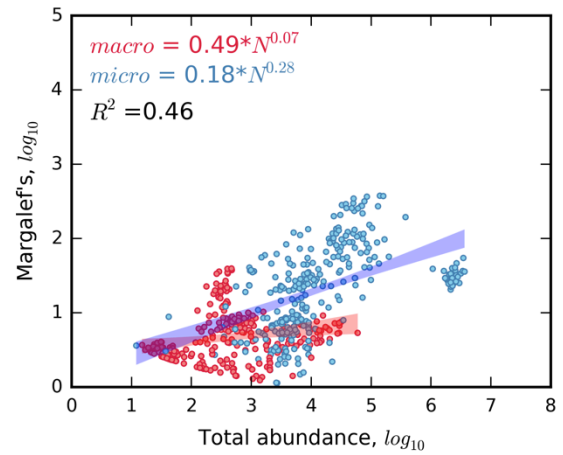
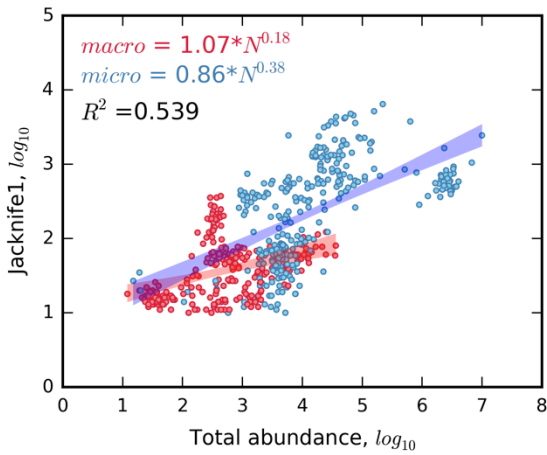
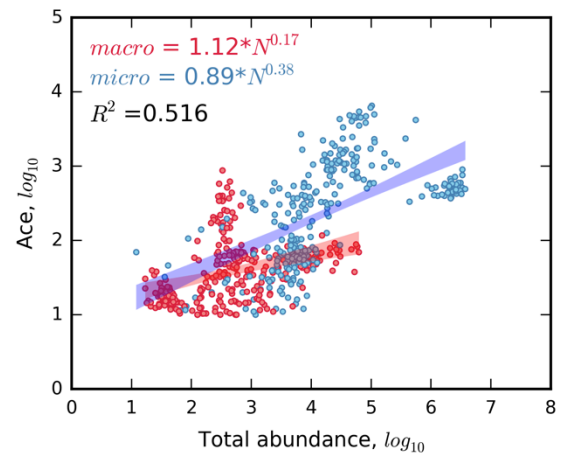
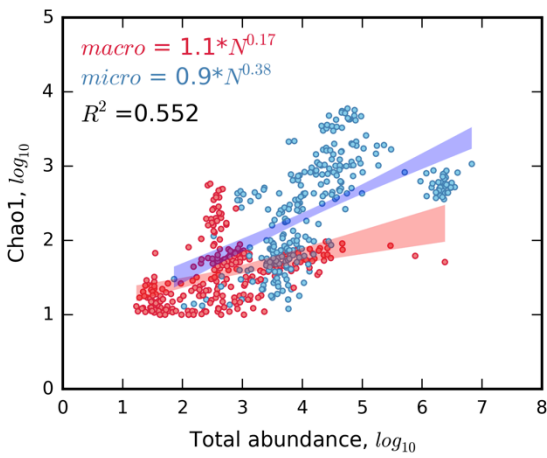


Figure S5. Robust responses to samples size. Our analysis revealed that while passing parametric tests depended on sample size, where larger samples resulted in p -values less than the alpha value (0.05) the regression model coefficients and the coefficients of determination (R^2) were independent of sample size. More specifically, ordinary least squares regression includes several assumptions, not all of which are fatal when violated. These include linearity, normally distributed error terms, and no serial correlation in error terms. We tested assumptions of linearity, normality, homoscedasticity (no change in error structure across the x -axis), and serial correlation, across a range of sample sizes because larger samples are more likely to uncover a real difference (greater statistical power) but are more likely to fail parametric tests of regression assumptions.

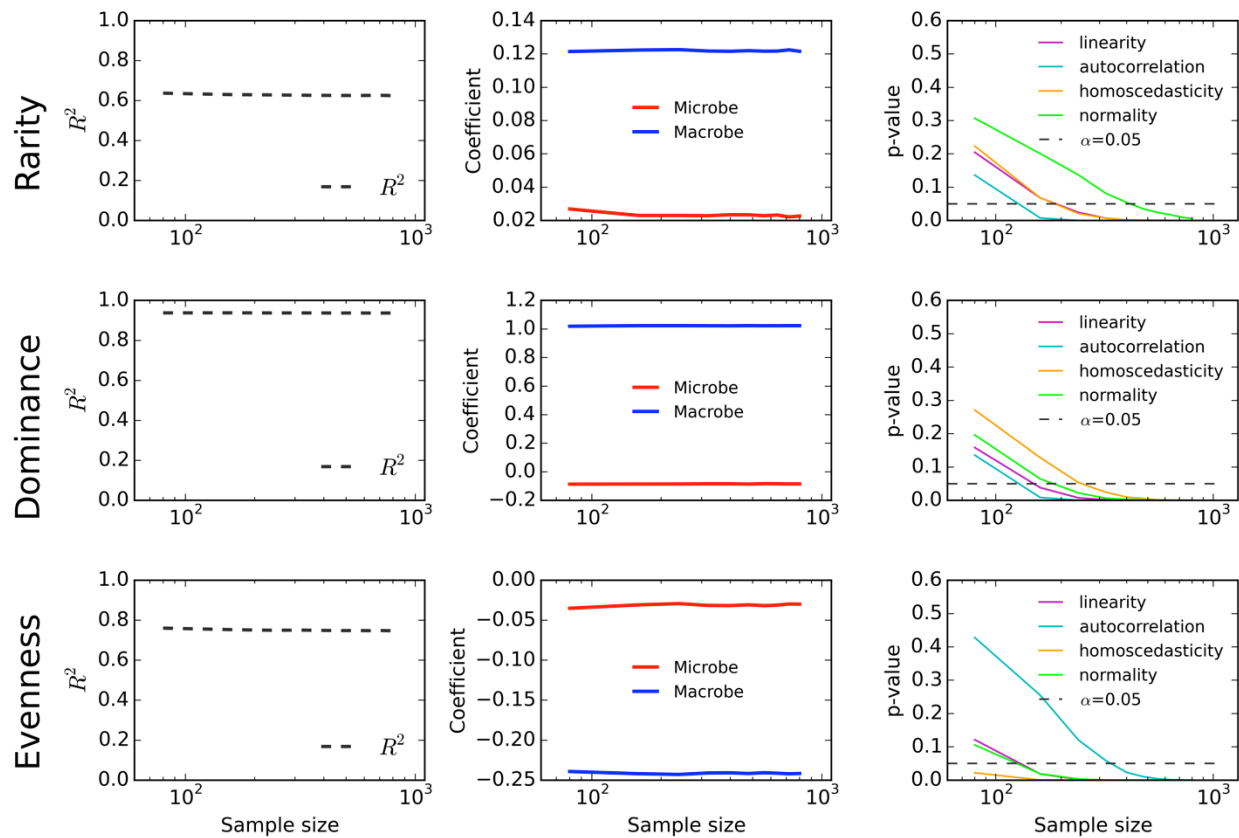


Figure S6. Testing the effect of categorical variable through random reassignment. Here, we randomly reassigned sites to the microbe/macrobe categorical variable and reveal **1.)** that identical model parameters can be obtained and **2.)** the general form of the relationships when the categorical variable is basically ignored. The plots of data in each subfigure represent a single random sample from microbe and macrobe data compilations. The model formulas represent average coefficient values from 1,000 random resamplings (with reassignment of the microbe/macrobe category).

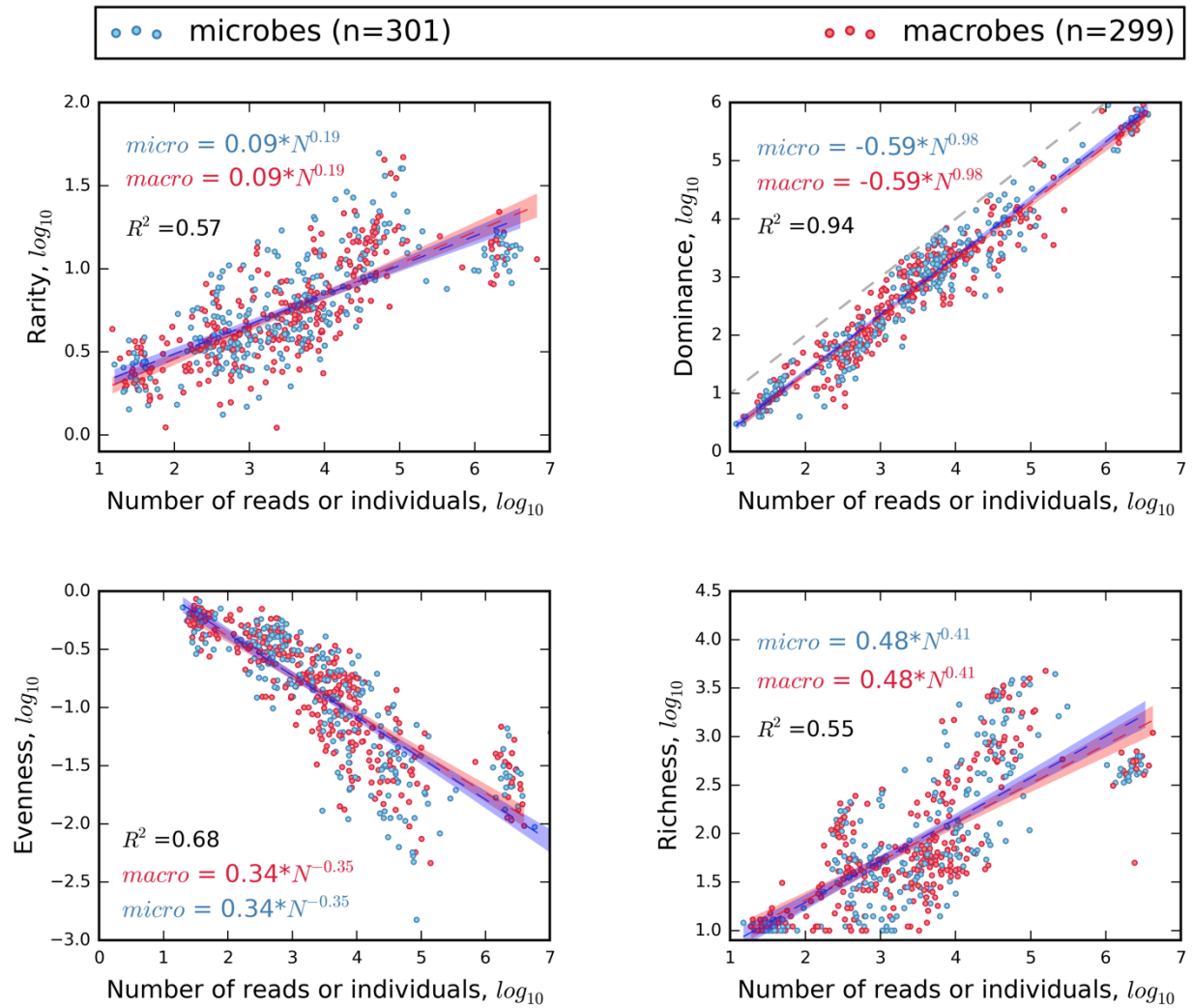
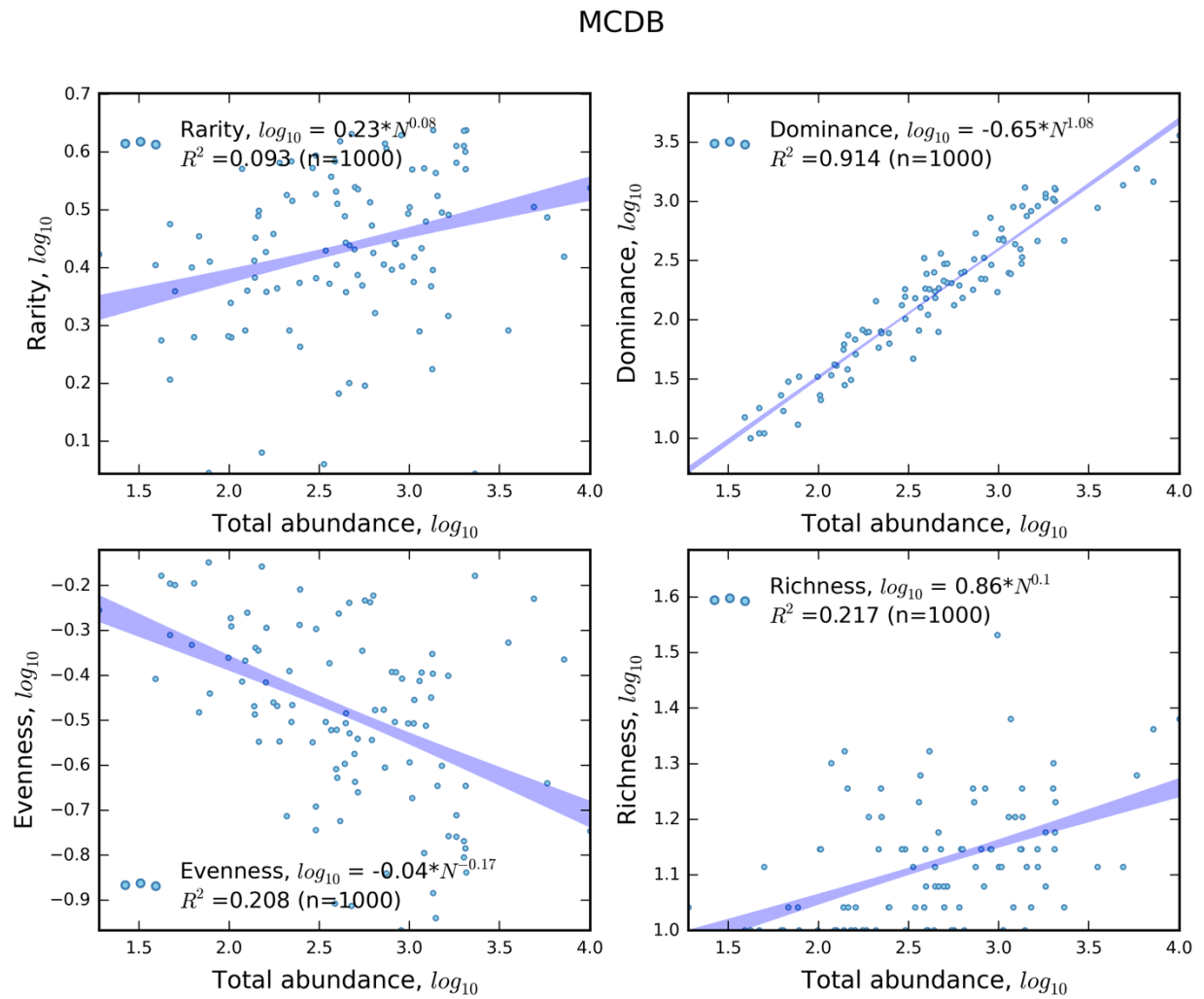


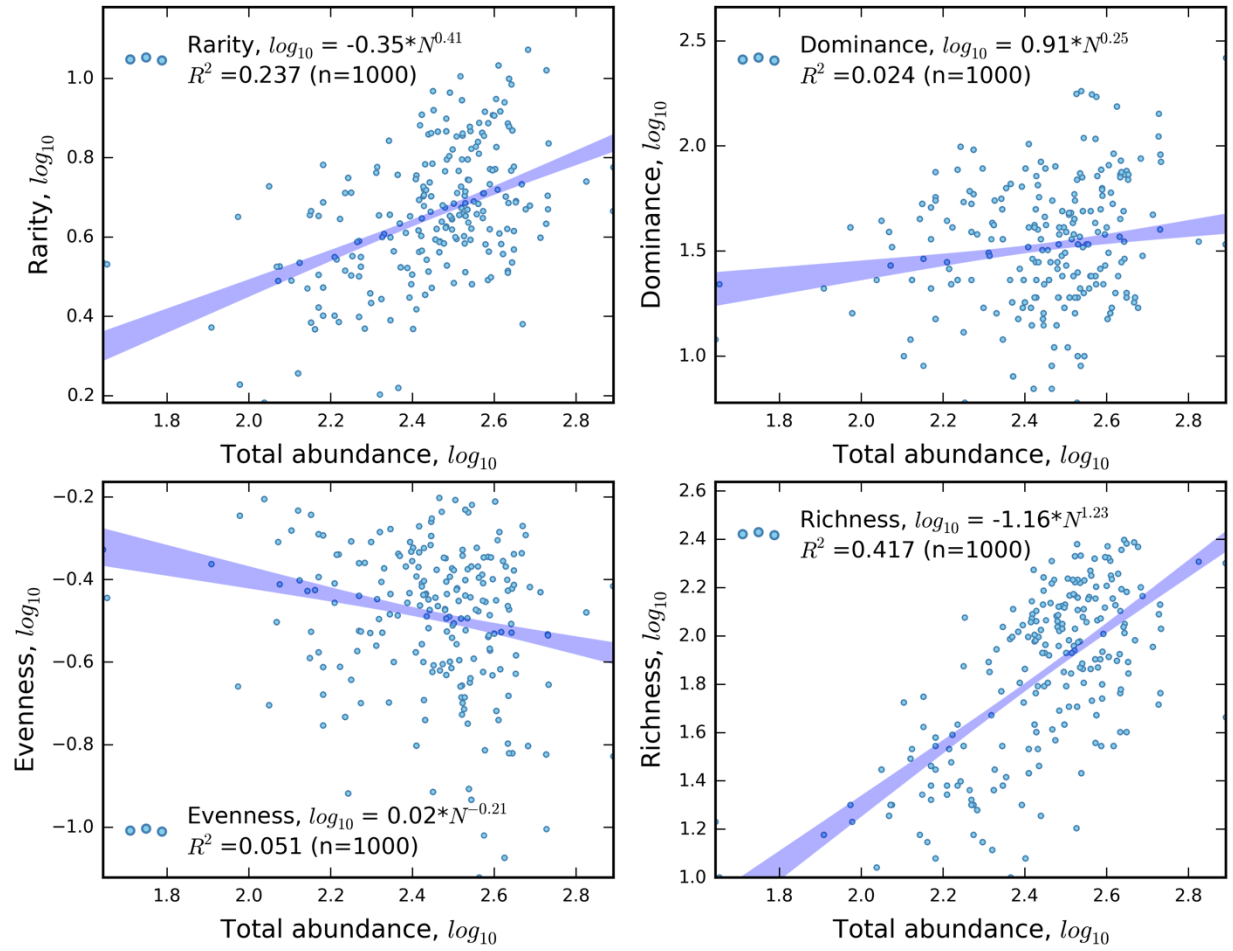
Figure S7. A-H. Results per dataset. The following figures (each with four subplots) show how aspects of diversity relate to sample abundance (N), i.e., the number of individual organisms or gene reads detected. The metrics are the same as those used in Fig. 1 in the main body, that is rarity (log-modulo skewness), dominance (N_{max}), Simpson's evenness metric, and observed richness (S). While the exact form and strength of the relationships vary, each dataset follows the same direction, i.e., for each relationship: increasing for rarity, dominance, and richness, and decreasing evenness. The only exception is no relationship of rarity to N for the Forest Inventory and Analysis.

Sub-figure A. Mammal Community Database (MCDB)



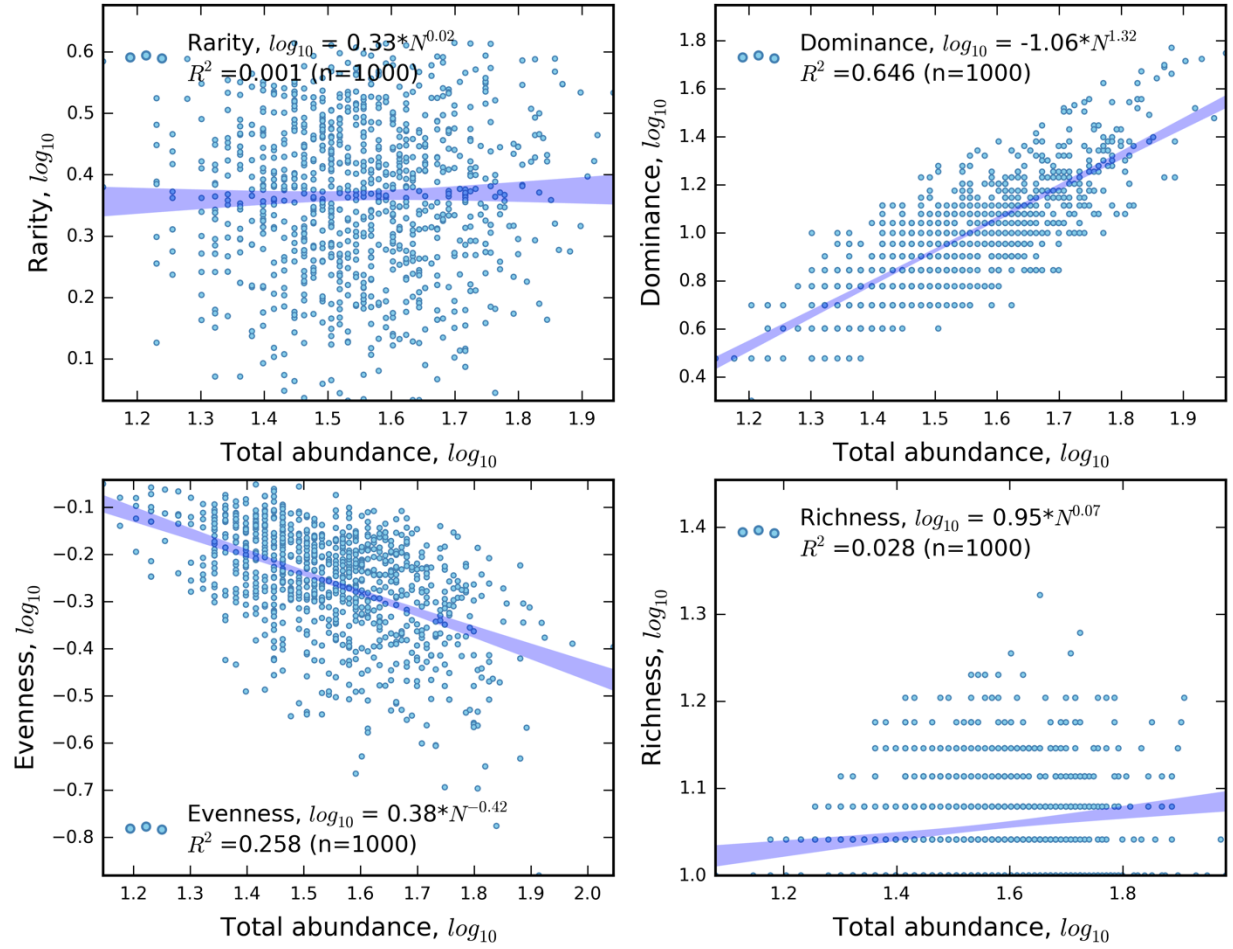
Sub-figure B. Alwyn Gentry's Forest Transects (GENTRY)

GENTRY



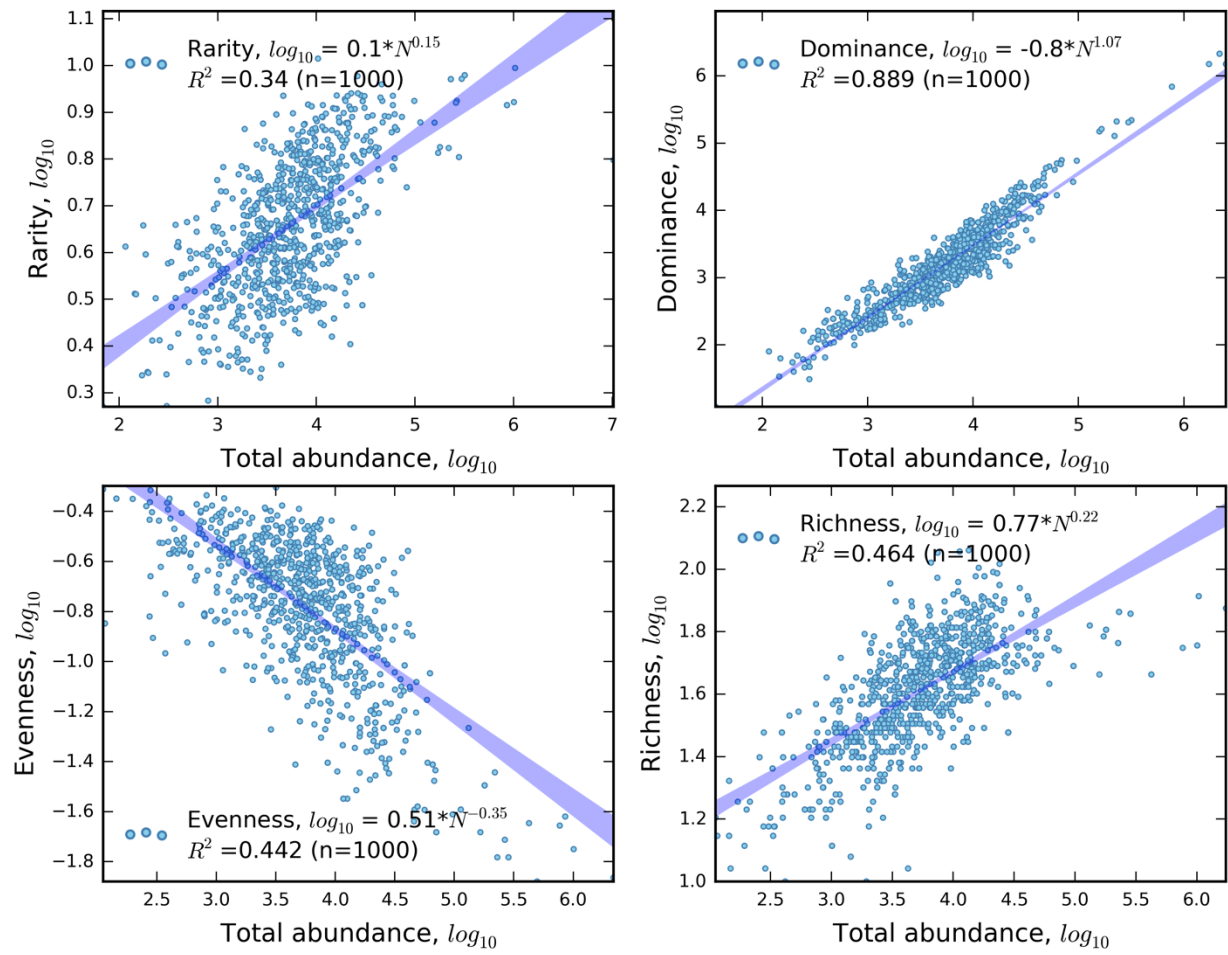
Sub-figure C. USDA Forest Inventory and Analysis dataset (FIA)

FIA



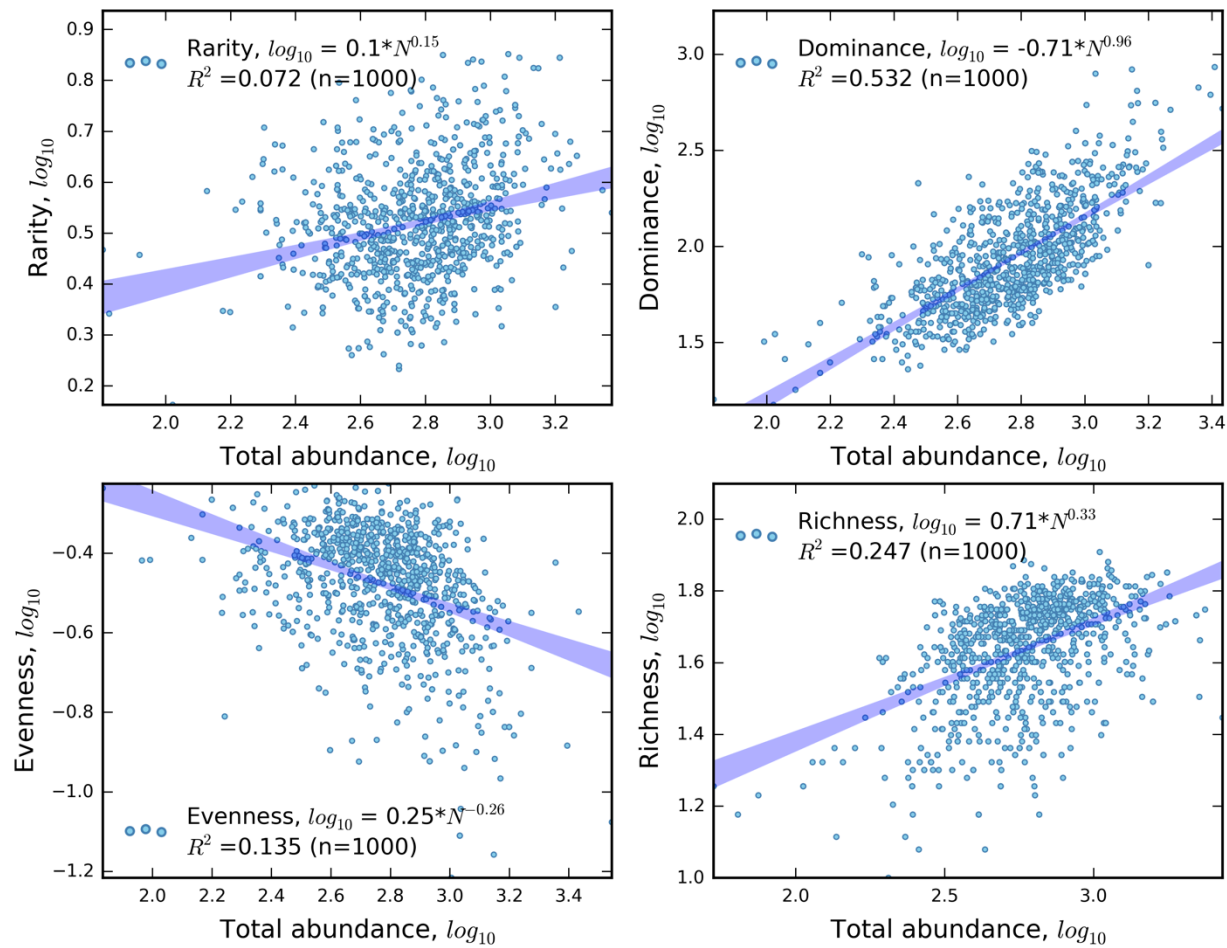
Sub-figure D. National Audubon Society’s Christmas Bird Count (used with permission) (CBC)

CBC



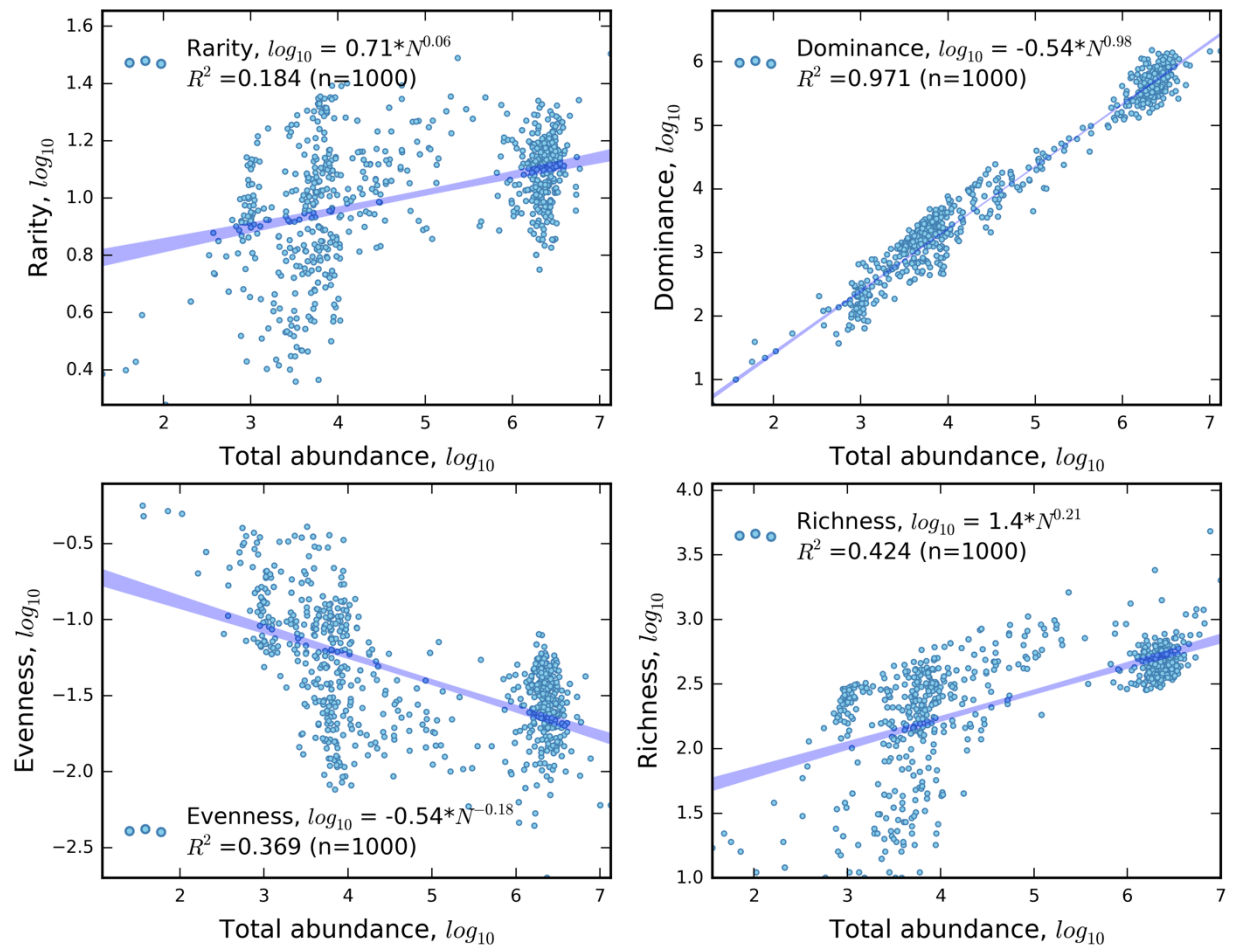
Sub-figure E. North American Breeding Bird Survey (BBS)

BBS



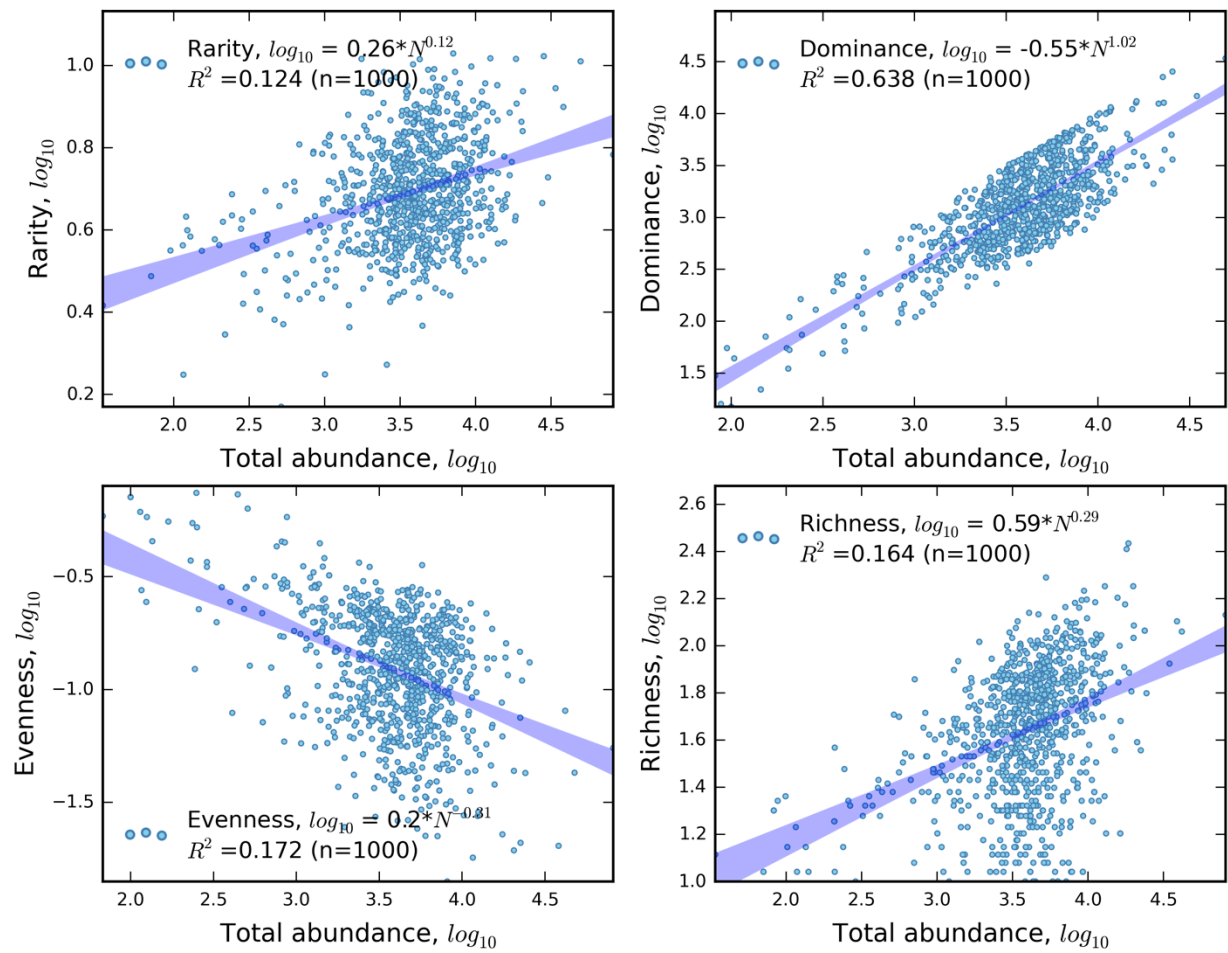
Sub-figure F. Data obtained from projects uploaded to the National Argonne Laboratory's MG-RAST metagenomic server (MGRAST).

MGRAST



Sub-figure G. Human Microbiome Project (HMP)

HMP



Sub-figure H. Earth Microbiome Project, closed reference OTU data (EMPclosed)

EMPclosed

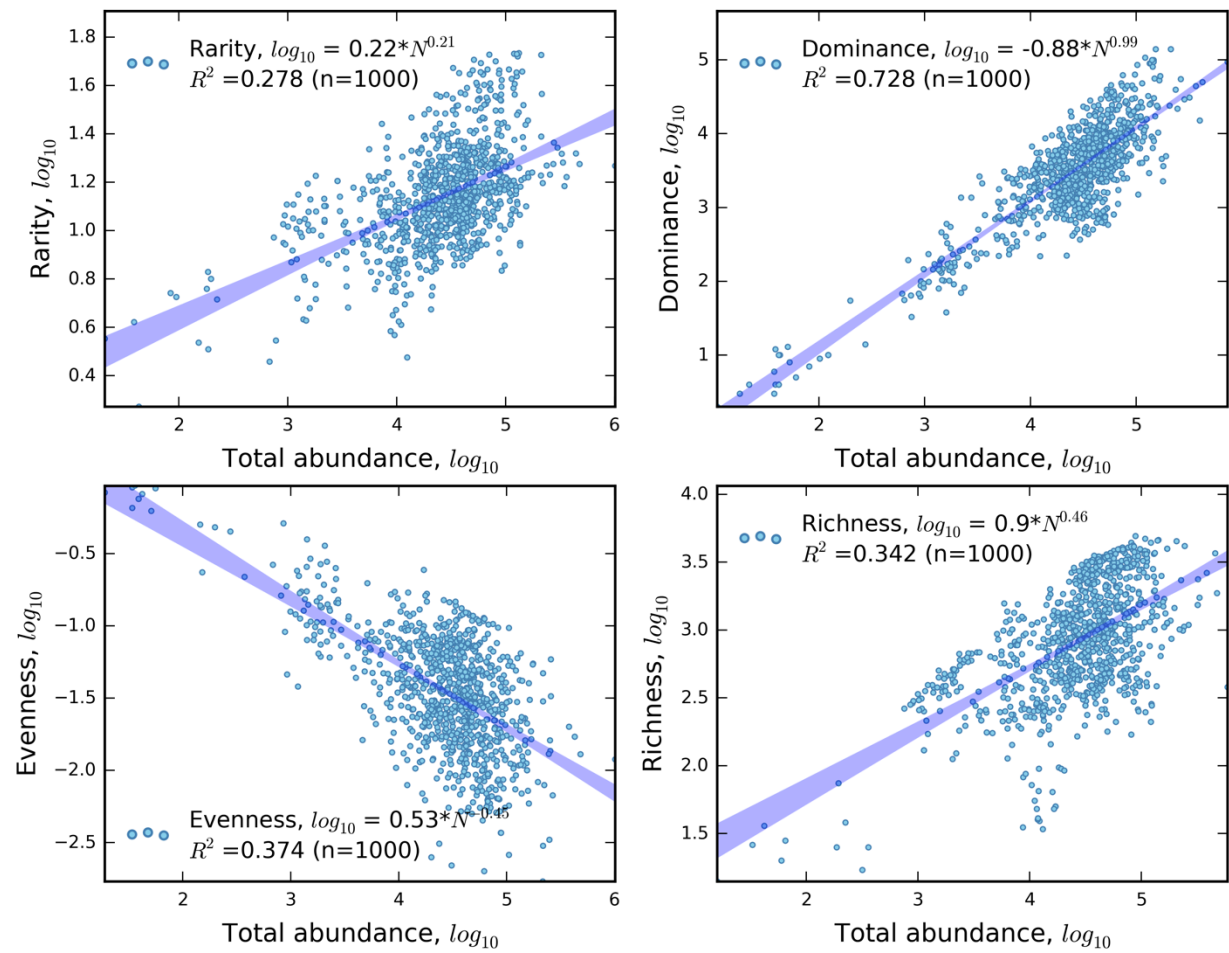


Figure S8. Binning taxa according to 95, 97, and 99 percent sequence similarity among 16S rRNA genes did not effect our results. Here, we use a subset our data from MG-RAST to show that relationships of diversity do not differ when using 95, 97, or 99% similarity. The metrics are the same as those used in Fig. 1 in the main body, that is rarity (log-modulo skewness), dominance (N_{max}), Simpson's evenness metric, and observed richness (S).

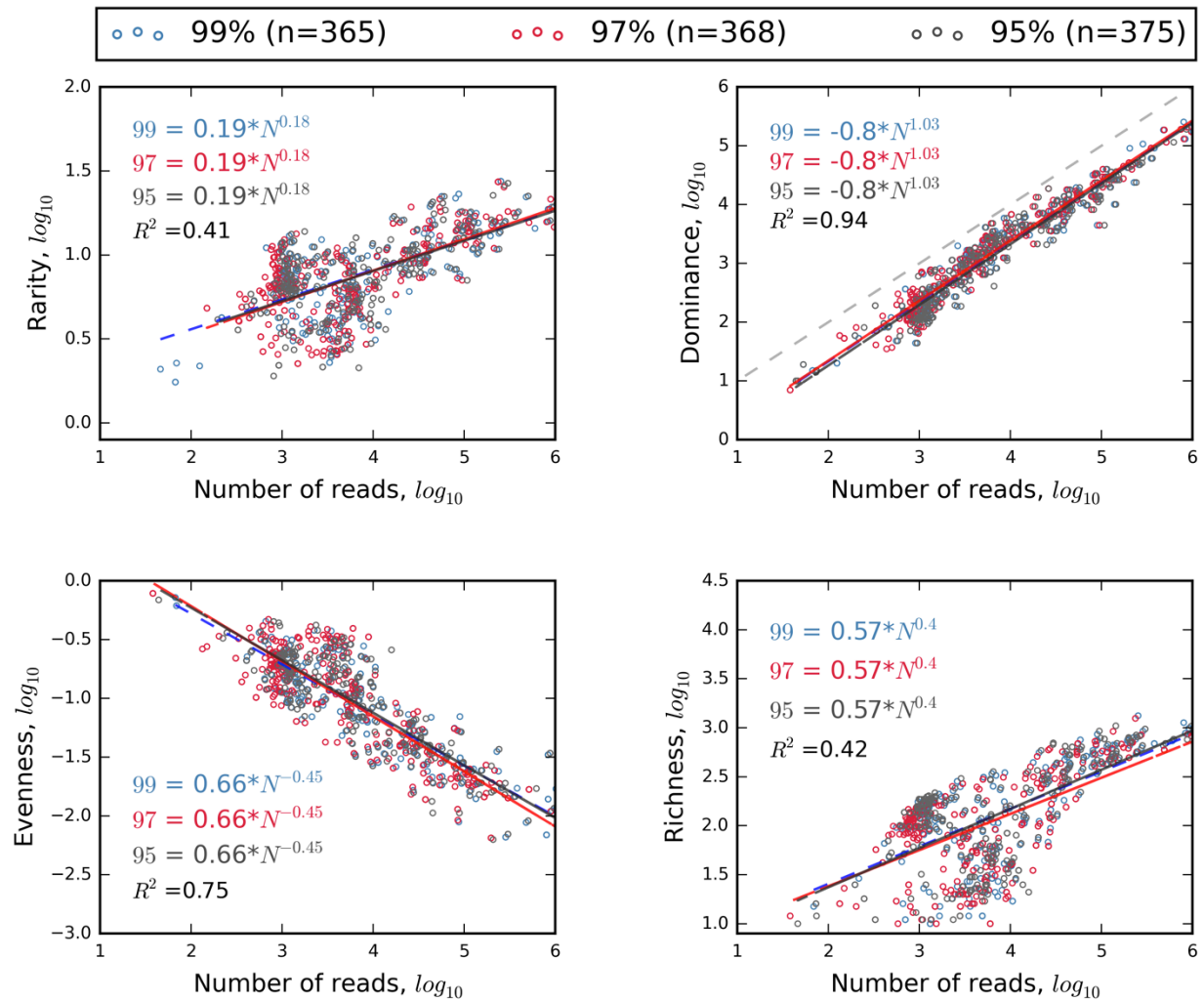


Figure S9. Including and excluding singleton taxa among microbes did not effect our results. Taxonomic units resulting from molecular surveys and characterized by a single 16S rRNA read are often considered dubious in microbial community studies. However, we found no consistent and noticeable differences when either including or excluding microbial singletons. The metrics are the same as those used in Fig. 1 in the main body, that is rarity (log-modulo skewness), dominance (N_{max}), Simpson's evenness metric, and observed richness (S).

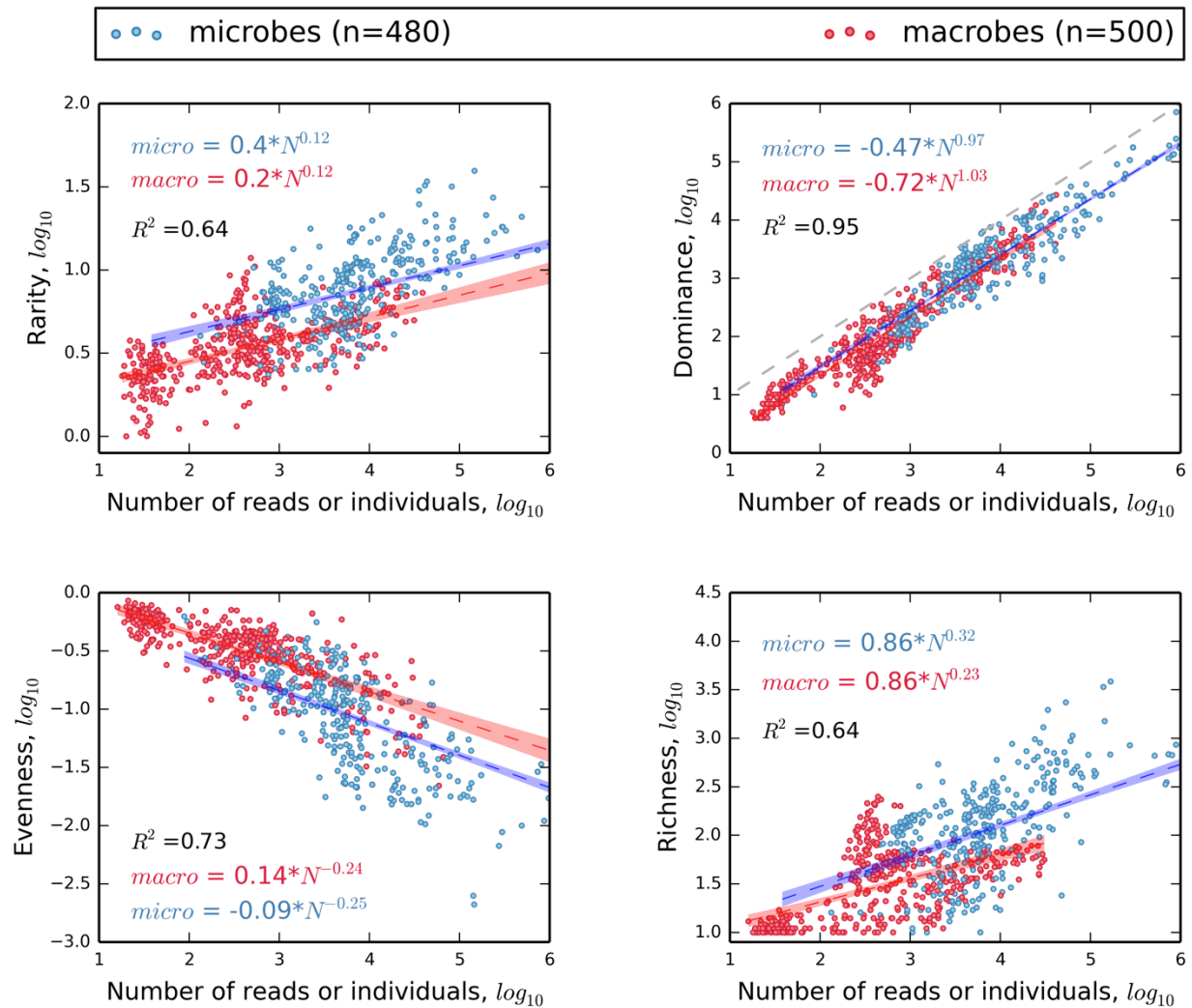


Figure S9. Flow of using observed N to obtain predicted N_{max} , and then use those values to parameterize the lognormal model.

Obtaining bootstrapped predictions of S for a microbiome or microbial community where values of total abundance (N) have been reported. Below, N_{max} is the predicted abundance of the most abundant species.

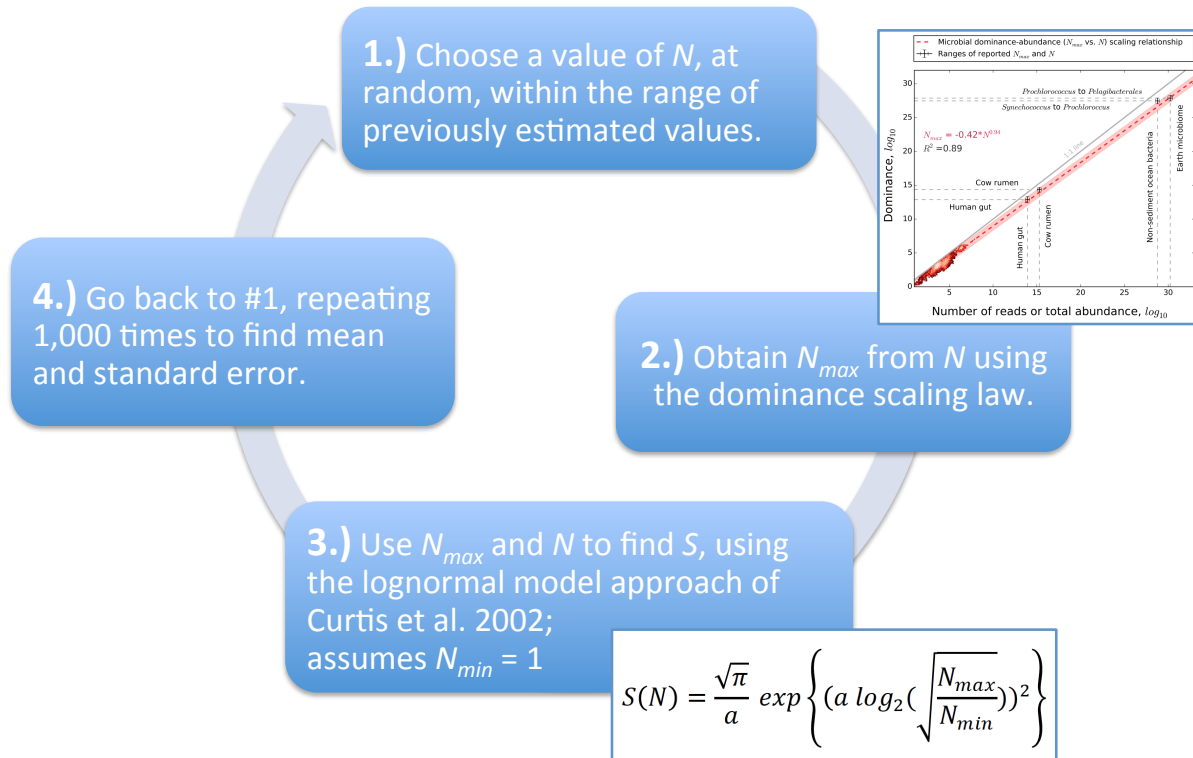


Table S1. Comparing fits of Power-law, Semi-log, Exponential and linear models. The power-law model provides the best fit to the data in regards to rarity, dominance, richness, and evenness. In regards to evenness, the power-law model explains a degree of variation similar to that of the exponential model (73% vs. 75%), but its AIC and BIC scores were nearly ten times less than those of the exponential; making it the preferred model.

Rarity	R-squared	AIC	BIC
power-law	0.625	-292.092	-274.504
Semi-log	0.457	3530.146	3547.733
exponential	0.508	-129.035	-111.447
linear	0.345	3642.024	3659.612
Dominance	R-squared	AIC	BIC
power-law	0.938	304.839	322.427
Semi-log	0.548	15516.828	15534.416
exponential	0.679	1286.795	1304.383
linear	0.832	14901.395	14918.983
Evenness	R-squared	AIC	BIC
power-law	0.734	133.599	151.187
Semi-log	0.75	-1068.973	-1051.385
exponential	0.582	405.159	422.747
linear	0.531	-690.923	-673.335
Richness	R-squared	AIC	BIC
power-law	0.602	690.997	708.585
Semi-log	0.254	9232.117	9249.705
exponential	0.442	893.132	910.72
linear	0.163	9301.414	9319.002