nature ecology & evolution

Article

https://doi.org/10.1038/s41559-025-02701-y

Dominant species predict plant richness and biomass in global grasslands

Received: 12 January 2025

Accepted: 4 April 2025

Published online: 13 May 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

The bidirectional relationship between plant species richness and community biomass is often variable and poorly resolved in natural grassland ecosystems, impeding progress in predicting impacts of environmental changes. Most biological communities have long-tailed species abundance distributions (for example, biomass, cover, number of individuals), a general property that may provide predictive power for species richness and community biomass. Here we show mathematical relationships between community characteristics and the abundance of dominant species arising from long-tailed distributions and test these predictions using observational and experimental data from 76 grassland sites across 6 continents. We find that community biomass provides little predictive ability for community richness, consistent with previous findings. By contrast, the relative abundance of dominant species quantitatively predicts species richness, whereas their absolute abundance quantitatively predicts community biomass under both ambient and altered environmental conditions, as expected mathematically. These results are robust to the type of abundance measure used. Three types of simulated data further show the generality of these results. Our integrative framework, arising from a few dominant species and mathematical properties of species abundance distributions, fills a persistent gap in our ability to predict community richness and biomass under ambient and anthropogenically altered conditions.

Species richness and community biomass are two of the most fundamental characteristics of plant communities that drive ecosystem functions and services such as resilience to disturbances, carbon storage and nutrient cycling[1], yet prediction of these community characteristics, and their relationship, has remained elusive[2,3]. At biogeographic spatial scales, species richness and community biomass have shared environmental drivers leading to covariation[4]. This relationship is often strong in experiments in which plant composition or richness is maintained as a treatment, such as in grassland and forest biodiversity experiments or intercropping systems, through seed addition and weed removal[5–9]. However, the biomass–richness relationship is often variable and poorly resolved when communities naturally assemble[10–16]. The inability to predict observed plant richness and biomass in natural systems has hindered progress in predicting the impacts of anthropogenic perturbations and changing environmental conditions on plant community characteristics.

In contrast to the highly variable biomass–richness relationships observed in naturally assembled communities, species abundance distributions (SADs) are consistently long tailed and strongly skewed (for example, log-normal distribution, Fisher's log series distribution, (zero-sum) multinomial distribution)[17–19], with a few highly abundant species dominating the community and many relatively rare species[19–25]. This long-tailed abundance distribution has biological importance, because the most dominant species often exerts a disproportionate impact on ecosystem processes, functions and community dynamics[23,26–29]. Mathematically, long-tailed distributions provide a

e-mail: zhangpengfei@lzu.edu.cn

https://doi.org/10.1038/s41559-025-02701-y

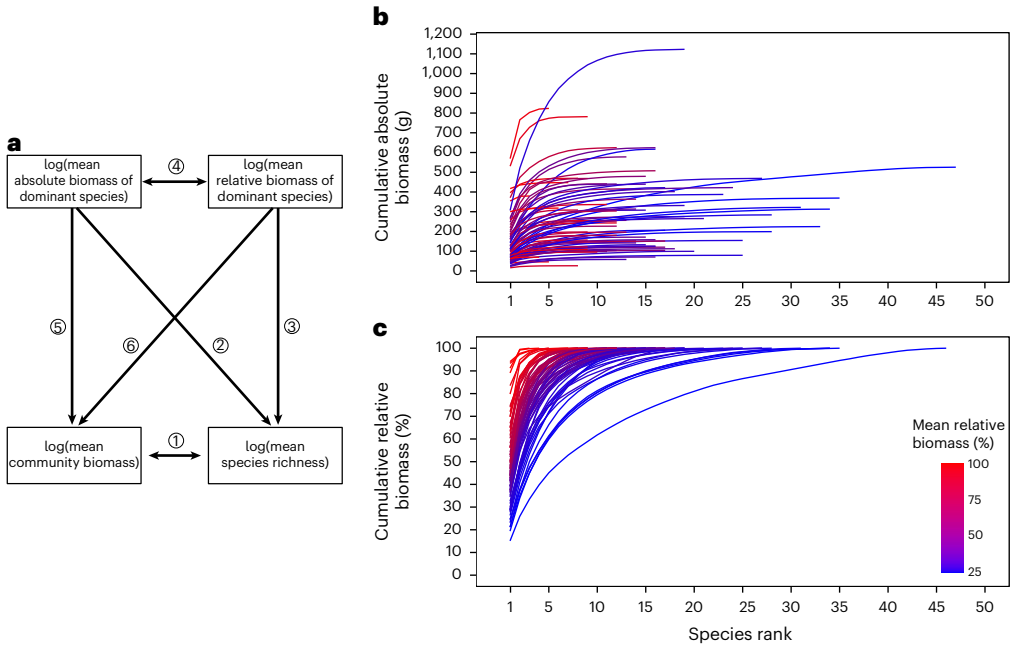


Fig. 1 | Structural relationships and species abundance patterns across global grassland sites. a–**c**, Structural equation meta-model (**a**) and site-level patterns from 76 grassland sites throughout the world under ambient conditions of cumulative absolute abundance curve (**b**) and cumulative relative abundance curve (**c**) based on species-level biomass estimates. The numbers in circles in **a** represent bivariate relationships. The *x*-axis in **b** and **c** shows the rank of species from most to least abundant. Line colours from blue to red in **b** and **c** represent the site-level mean relative biomass of the 2 most dominant species at each of the 76 sites under ambient conditions from low to high (from 25.9% to 99.5%; see Supplementary Table 1 for additional details).

strong description of species abundances (whether assessed through species-level biomass, absolute cover or number of individuals) in biological communities[18,30–33], suggesting the possibility that dominant species and properties of the long-tailed distribution may provide an integrated framework for predicting total community biomass and richness.

Dominant species abundances within SADs can be quantified using either absolute abundance estimates (for example, counts, biomass or absolute cover) or relativized abundance estimates (for example, proportions, or percentages), with each offering distinct insights into community structure[34]. For instance, while both types of estimates provide information on species diversity, dominance and species-level abundance trends through time and across space[35], only absolute abundance includes information on total population size[34] and, by extension, community biomass[28,36–39]. For long-tailed distributions, it can be mathematically shown that the absolute abundance of the most dominant species provides a strong prediction for the total abundance of all species in a community[40]. Although currently lacking a mathematical demonstration, measures of dominance based on relativized abundance of dominant species are constrained by the percentage occupied by the abundance of all other species in the community. Thus, we hypothesize that because of the long-tailed distribution of abundance, the relative abundance of dominant species should be related to species richness of the community[24,41–44].

Here we test whether the absolute and relative abundance of dominant species can provide predictive insights into both community biomass and richness across relevant gradients of species richness and environmental conditions in global grasslands. We begin by demonstrating the mathematical relationship between the relative abundance of dominant species and species richness. We follow this by testing these relationships with empirical data using observational (that is, ambient conditions) and experimental (that is, altered environmental conditions) data collected at a 1-m^2 scale at 76 grassland sites on 6 continents contributing to the Nutrient Network (NutNet) distributed experiment (Extended Data Fig. 1a,b)[45]. We use bivariate regression (paths 1–6 in Fig. 1a and Extended Data Fig. 1c) and structural equation modelling (meta-model in Fig. 1a and Extended Data Fig. 1c) to test whether the absolute and relative abundance of dominant species effectively predict community biomass and species richness across sites under a wide range of edaphic and climatic conditions. We then test whether the relationships persist under experimentally imposed environmental changes that reorganize the relative abundance of species[46]. For the analyses we present in the main text, we estimate species-level biomass from directly measured species-level cover using a power law relationship[47]. This approach is validated by strongly concordant results in an analysis of data from seven sites with direct measures of both species-level cover and species-level biomass (Extended Data Fig. 2a–g)[47]. We show that the generality of these results is underlain by the mathematical characteristics of long-tailed species abundance distributions: our results do not depend on the type of abundance measure used and produce qualitatively similar results using species-level biomass, species areal cover or species areal cover scaled by plot biomass (Extended Data Fig. 2a–g and Supplementary Figs. 1–5). Finally, we reveal the generality and robustness of these results by analysing simulated data from three different long-tailed distributions (log-normal distribution, Fisher's log series distribution and multinomial distribution) to bridge the mathematical relationships and empirical data.

Results and discussion

Mathematical relationship of dominant species and richness

We begin with the empirical observation that the abundance (that is, biomass, cover or number of individuals) of species within a single relatively homogeneous plot follows a long-tailed distribution[22,32], such as the log-normal or Fisher's log series distribution. Our mathematical modelling shows that the relative abundance of the most dominant species predicts the total community species richness in a plot. We let the abundance of the most dominant species in the plot be defined as

$$A_1(S) \equiv \max\{Y_1, Y_2, \ldots Y_s\} \quad (1)$$

https://doi.org/10.1038/s41559-025-02701-y

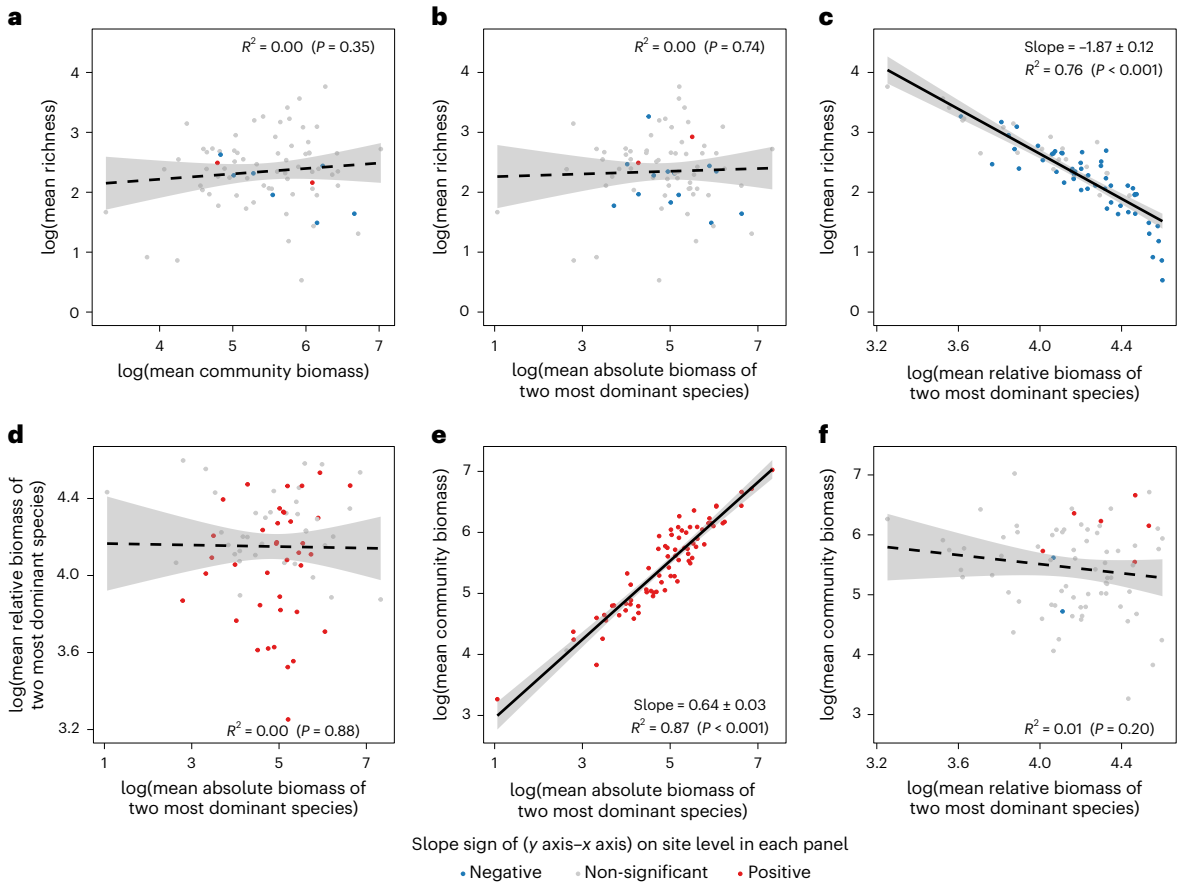


Fig. 2 | Grassland richness and biomass relationships under ambient conditions. a–f, The relationship between mean richness and mean community biomass (**a**), mean absolute biomass of the two most dominant species (**b**) and mean relative biomass of the two most dominant species (**c**); between the mean relative biomass and mean absolute biomass of the two most dominant species (**d**); and between mean community-level biomass and mean absolute biomass of the two most dominant species (**e**) and mean relative biomass of the two most dominant species (**f**), at 76 sites under ambient conditions (each site ≈3 blocks; each block ≈10 plots). All data were natural log transformed. The correlation between the *y*-axis and *x*-axis variables of each panel on the site level is indicated as significantly positive (red), uncorrelated (grey) and significantly negative (blue) (Extended Data Fig. 3a–f). The dashed and solid lines indicate that the overall relationship is not significant ($P > 0.05$) and significant ($P < 0.05$), respectively, with shaded areas indicating 95% confidence intervals. Significant slopes in **c** and **e** are reported as mean ± s.e.m. All statistical tests are conducted as two sided.

where Y_i is the abundance of the ith species and S is the species richness. The random variables Y_i are assumed to be independent and identically distributed, and their tail probability function is denoted by $\bar{F}(y)$. Then, the relative abundance of the most dominant species is

$$\hat{A}_1(S) = \frac{A_1(S)}{\sum_{i=1}^{s} Y_i} \sim \frac{A_1(S)}{S \times E[Y_1]} \tag{2}$$

where $E[Y_1]$ represents the expected value (mean) of the random variables Y_1, calculated as the average outcome over many realizations of Y_1. For a large S, we have the asymptotic result[48]

$$\frac{A_1(S)}{b_S} \sim 1 \text{ where } b_S \equiv \bar{F}^{-1}(1/S). \tag{3}$$

To be more precise, b_S is defined in terms of the cumulative distribution function (CDF) of the abundance distribution. Its specific values are then calculated for the two distributions (log-normal and Fisher's log series) analysed in detail in Methods.

Applying a logarithm to both sides yields

$$\log\left(\hat{A}_1(S)\right) \sim \log(b_S) - \log(S) - \log(E[Y_1]), \tag{4}$$

This relationship reveals that if the following condition on the tail behaviour of the distribution of Y_i holds:

$$\frac{\log(b_S)}{\log(S)} \to 0 \text{ as } S \to \infty, \tag{5}$$

then the logarithm of the relative abundance of the most dominant species, $\hat{A}_1$, is nearly linearly dependent on the logarithm of the species richness S. These relationships can also be shown to hold for the top two dominant species (Methods). For both the log-normal and Fisher's log series distributions, it can be shown that equation (5) holds (Methods), thus suggesting that empirically, the dominant species' relative abundance may be predictive of species richness.

Predicting biomass and richness under ambient conditions

The cumulative absolute and relative abundance (that is, biomass estimates here) curves under ambient conditions from 76 NutNet sites spanning a wide range of grassland floras (Fig. 1b,c) and abiotic environmental conditions (Extended Data Fig. 1a,b and Supplementary Table 1) show that greater mean species richness does not necessarily correspond to greater mean community biomass (as indicated by the mixed order of colours in Fig. 1b). However, sites with greater mean species richness have a lower proportion of relative abundance attributed to one or a few dominant species (as shown by the colour gradient transitioning from red to blue in Fig. 1c).

As predicted by the mathematical relationship in equation (5) and in Methods, the mean relative biomass estimates of the two most

https://doi.org/10.1038/s41559-025-02701-y

Table 1 | The R^2 of various relationships among four natural log-transformed variables (top) and the R^2 of relationships between natural log-transformed gamma diversity and natural log-transformed mean relative biomass of dominant species (bottom) in the context of selecting different numbers of dominant species (DS_n; n is from 1 to 5) from NutNet data under both ambient and altered conditions

		DS_n				
		n=1	n=2	n=3	n=4	n=5
Variables						
The R^2 in ambient conditions	log(mean richness)–log(mean absolute biomass of DS_n)	0.00^{NS}	0.00^{NS}	0.00^{NS}	0.00^{NS}	0.01^{NS}
	log(mean richness)–log(mean relative biomass of DS_n)	0.75***	0.76***	0.73***	0.70***	0.67***
	log(mean relative biomass of DS_n)–log(mean absolute biomass of DS_n)	0.00^{NS}	0.00^{NS}	0.00^{NS}	0.00^{NS}	0.01^{NS}
	log(mean community biomass)–log(mean absolute biomass of DS_n)	0.84***	0.87***	0.88***	0.88***	0.88***
	log(mean community biomass)–log(mean relative biomass of DS_n)	0.00^{NS}	0.01^{NS}	0.01^{NS}	0.01^{NS}	0.01^{NS}
The R^2 in altered environmental conditions	log(mean richness)–log(mean absolute biomass of DS_n)	0.01^{NS}	0.00^{NS}	0.00^{NS}	0.00^{NS}	0.00^{NS}
	log(mean richness)–log(mean relative biomass of DS_n)	0.79***	0.84***	0.84***	0.82***	0.79***
	log(mean relative biomass of DS_n)–log(mean absolute biomass of DS_n)	0.04*	0.00^{NS}	0.00^{NS}	0.00^{NS}	0.00^{NS}
	log(mean community biomass)–log(mean absolute biomass of DS_n)	0.90***	0.91***	0.90***	0.89***	0.88***
	log(mean community biomass)–log(mean relative biomass of DS_n)	0.00^{NS}	0.00^{NS}	0.00^{NS}	0.00^{NS}	0.00^{NS}
Gamma diversity						
	The R^2 of log(gamma diversity)–log(mean relative biomass of DS_n) in ambient conditions	0.45***	0.47***	0.43***	0.37***	0.31***
	The R^2 of log(gamma diversity)–log(mean relative biomass of DS_n) in altered conditions	0.35***	0.39***	0.37***	0.33***	0.28***

The values in the table represent R^2 from ordinary least squares (OLS) linear regression analysis, with symbols in the upper right indicating the corresponding P value ranges: *P<0.05, ***P<0.001 and 'NS' denotes non-significant results (P>0.05). All statistical tests were conducted as two sided.

dominant species accurately predicted mean community richness (slope = −1.87 ± 0.12, R^2 = 0.76, P < 0.001; Fig. 2c). Similarly, as predicted mathematically[40], the mean absolute biomass estimates of the two most dominant species accurately predicted mean total community biomass (slope = 0.64 ± 0.03, R^2 = 0.87, P < 0.001; Fig. 2e) under ambient conditions. By contrast, there were no significant relationships among the other pairwise combinations of these four variables (P > 0.05; Fig. 2a,b,d,f). The empirical results were robust to the number of dominant species included (one to five dominant species; Table 1). Here we present results of the two most dominant species (referred to as dominant species) because these had a slightly higher correlation with community biomass, species richness and gamma diversity (that is, site-level richness) (Table 1, top and bottom). The site-level results of these relationships under ambient conditions are presented in Extended Data Fig. 3a–f.

Predicting biomass and richness under altered conditions

Using community responses to identical nutrient supply and herbivore density treatments across 76 sites, we investigated whether these relationships remained consistent under altered environmental conditions and found nearly identical results (Fig. 3), even though the treatments at most sites altered the identity of the dominant species (Supplementary Table 3a). Importantly, the slope of the relationship between dominant species abundance and community richness and biomass did not differ under ambient and altered environmental conditions (see mean ± s.e. of slopes in purple and black regression lines in Fig. 3c,e). This slope arises from the parameters of the underlying long-tailed distribution of Y_i, underscoring the generality and robust predictive ability of dominant species even under environmental change. Furthermore, the slope did not differ between two different methods for calculating absolute and relative biomass of dominant species across seven sites (see mean ± s.e. of slopes in purple and black regression lines in Extended Data Fig. 2c–g), clarifying that the predictive ability of dominant species is robust to methodology for describing the species-level abundance data (for example, biomass, cover or number of individuals).

Our results provided quantitative predictions of total community biomass and richness using information on only the two most dominant species. In both the observational and experimental communities, a 10% increase in relative biomass of the two most dominant species corresponded to reduction of approximately 20% in species richness (slope = −1.87 ± 0.12, R^2 = 0.76, P < 0.001 under ambient conditions, and slope = −2.08 ± 0.11, R^2 = 0.83, P < 0.001 under experimentally altered conditions; Fig. 3c). In addition, a 10% increase in absolute biomass of the two most dominant species was associated with an increase of approximately 6.6% in total community biomass (slope = 0.64 ± 0.03, R^2 = 0.87, P < 0.001 under ambient conditions, and slope = 0.68 ± 0.02, R^2 = 0.92, P < 0.001 under experimentally altered conditions; Fig. 3e). The site-level results of these relationships under altered conditions are presented in Extended Data Fig. 3g–l.

When we analysed each year (from year 1 to 15 in Supplementary Table 2a) and each treatment (control, nitrogen (N), phosphorus (P), potassium and micronutrients (Kμ), NP, NKμ, PKμ, NPKμ, fencing and fencing plus NPKμ in Supplementary Table 2b) separately under altered environmental conditions, the results were consistent with the data averaged across all years (last row in Supplementary Table 2a) or all treatments (last row in Supplementary Table 2b). This strong and consistent predictive ability of dominant species was further shown by the ability of the model fit to the observational data to quantitatively predict richness and biomass of communities under experimentally altered conditions (Fig. 4c,d).

We then used structural equation modelling (SEM) to test whether the distinct correlations between relative and absolute biomass of dominant species, community biomass and species richness persist after accounting for the correlations among the variables[3]. The SEMs using datasets from both ambient and altered environmental conditions (Fig. 5a,b) showed results consistent with the bivariate analyses (Figs. 2 and 3). Furthermore, we obtained qualitatively similar results using species-level absolute and relative cover (Supplementary Figs. 1–5 and Table 4) as well as gamma diversity (Extended Data Fig. 4). This consistency shows the robustness of these

https://doi.org/10.1038/s41559-025-02701-y

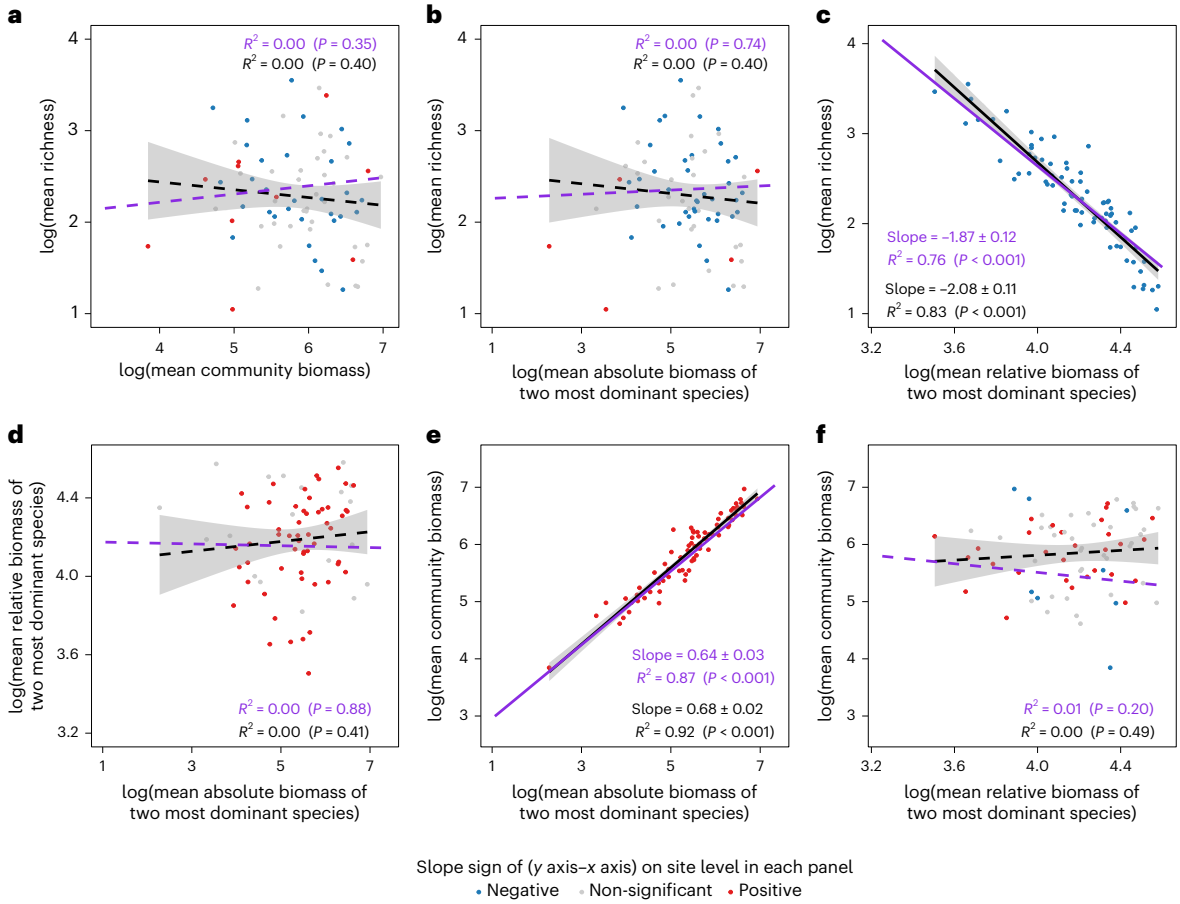


Fig. 3 | Grassland richness and biomass relationships under altered environmental conditions. a–f, The relationship between mean richness and mean community biomass (**a**), mean absolute biomass of the two most dominant species (**b**) and mean relative biomass of the two most dominant species (**c**); between the mean relative biomass and mean absolute biomass of the two most dominant species (**d**); and between the mean community-level biomass and the mean absolute biomass of the two most dominant species (**e**) and mean relative biomass of the two most dominant species (**f**), at 76 sites under altered environmental conditions (1–15 years; each site ≈3 blocks; each block ≈10 plots). All data were natural log transformed. The correlation between the *y*-axis and *x*-axis variables of each panel on the site level is indicated as significantly positive (red), uncorrelated (grey) and significantly negative (blue) (Extended Data Fig. 3g–l). The purple lines are regression curves for the ambient conditions in Fig. 2. The purple font indicates R^2 and *P* values for the ambient conditions in Fig. 2. The black lines are regression curves for the altered environmental conditions. The black font indicates R^2 and *P* values for the altered environmental conditions. The dashed and solid lines indicate that the overall relationship is not significant ($P > 0.05$) and significant ($P < 0.05$), respectively, with shaded areas indicating 95% confidence intervals. Significant slopes in **c** and **e** are reported as mean ± s.e.m. All statistical tests were conducted as two sided.

mathematical relationships across different abundance measures and diversity metrics.

Taken together, the mathematical demonstration and empirical results highlight that the absolute and relative abundance of dominant species provide unique and highly repeatable predictions of the total biomass and richness of biological communities[34,38,49–54]. Simultaneously considering both the absolute and relative abundance of dominant species provides valuable insights into the relationship underlying community richness and biomass, enhancing the ability to separately predict these factors in the face of global changes[34,51,55–60].

Simulated data highlight generality and a role for biology

We used three different types of randomly generated long-tailed-distributed data, drawn from log-normal, Fisher's log series and multinomial data distributions, to further test the generality of these empirical relationships and to better understand how biological systems may differ from purely statistical data. The results of the simulated data for all three long-tailed distributions mirror the relationships predicted mathematically (here and ref. 40) and observed empirically (Fig. 3 versus Extended Data Figs. 5–7 for bivariate relationships; Fig. 5a,b versus Fig. 5c–h for SEMs), reinforcing the generality of these relationships between dominant species and both community biomass and richness. While similar in many ways, dominant species in the biological data were more likely to remain dominant over time compared with simulated data (40–53% of the most dominant biological species remain dominant in NutNet data (Supplementary Table 3a), compared with 15–38% of the most dominant simulated species in log-normally distributed data (Supplementary Table 3b), compared with 17–45% of the most dominant simulated species in log series-distributed data (Supplementary Table 3c), and compared with 14–37% of the most dominant simulated species in multinomial distributed data (Supplementary Table 3d)). This difference has important biological implications. Species that are always dominant maintain community functions and attributes[23,26,29,61], whereas species that have recently become dominant (owing to, for example, invasion, environmental change) often introduce new attributes into the community, altering community functioning[29,62,63]. The persistence of always-dominant species under altered environmental conditions probably stabilizes community attributes arising from the identity of dominant species[23,28,34,35,38,64]. This suggests that the characteristics of biological communities are more highly conserved than is predicted solely from the properties

https://doi.org/10.1038/s41559-025-02701-y

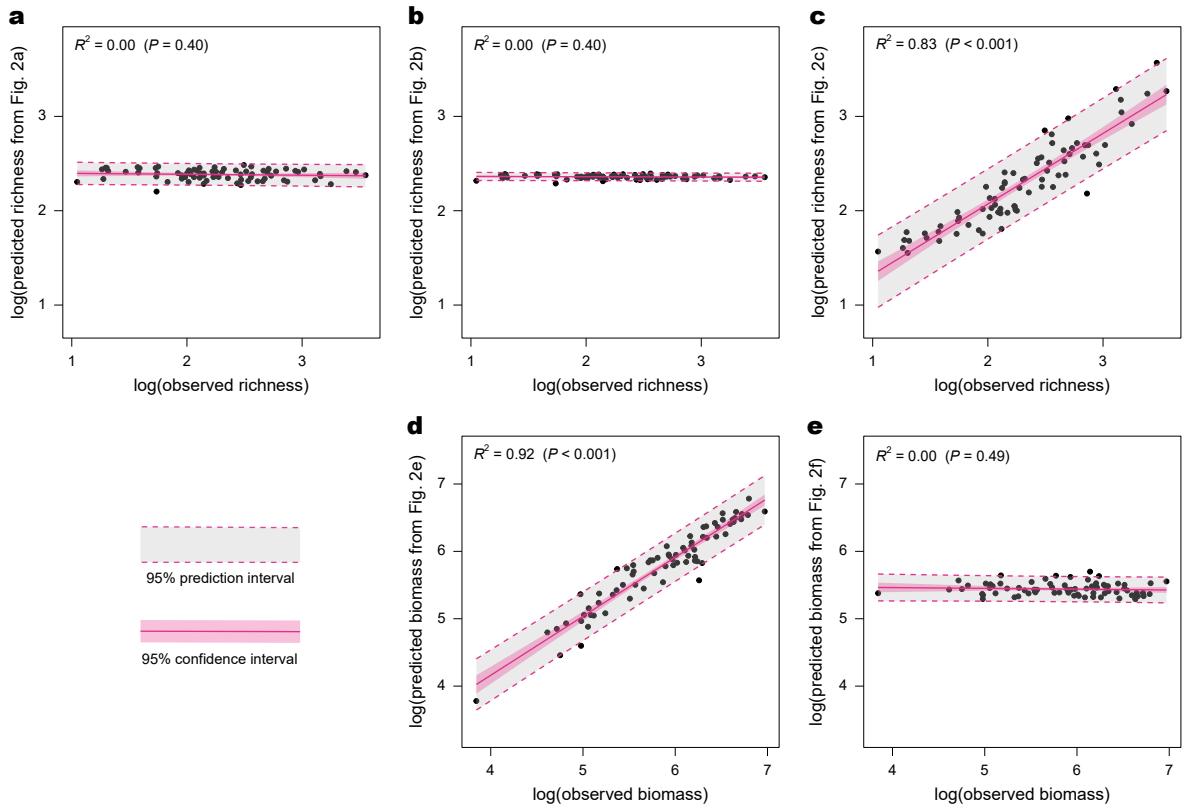


Fig. 4 | The regression model for ambient conditions predicts the outcome for global grasslands under altered environmental conditions. a–**e**, The relationship between the predicted species richness according to the regression model of the ambient conditions and the actual mean species richness of altered environmental conditions of each site (**a**–**c**) and between the predicted community biomass according to the regression model of the ambient conditions and the actual mean community biomass of the altered environmental conditions of each site (**d**,**e**). The regression models used to predict the vertical axis variables in **a**, **b**, **c**, **d** and **e** are, respectively, from the model in **a**, **b**, **c**, **e** and **f** of Fig. 2. In these five panels, the grey-shaded area is the 95% prediction interval, and the pink-shaded area represents the 95% confidence interval, around the regression line. All statistical tests are conducted as two sided.

of the log-normal distribution, Fisher's log series distribution and multinomial distribution.

In conclusion, our mathematical derivation, analysis of ambient and experimentally altered grassland communities and simulations using three different long-tailed distributions show that the relative abundance (whether assessed through biomass, absolute cover or number of individuals) of dominant species effectively and consistently predicts total community species richness, whereas the absolute abundance of dominant species consistently predicts total community abundance (for example, biomass). These relationships remain robust after reorganization of communities under environmental changes and do not depend on the type of abundance measure used. Because of the mathematical underpinnings of long-tailed distributions, our results show that the most dominant species in biological communities can be used to predict both community biomass and richness, even under anthropogenic perturbations and changing environmental conditions. Although our empirical insights are derived from grassland ecosystems, the underlying principles may extend beyond the realm of biology, as they encapsulate fundamental properties of long-tailed distributions that are observed in systems from geochemistry to neurology and epidemiology. Further validation across different systems will be critical to assess the broader applicability of these patterns.

Methods

Site description and experimental design

In our analyses, we use data collected at 76 sites, which are part of the NutNet, a collaborative global research network conducting a globally replicated experiment[45]. All the sites are located in areas dominated by low-statured, primarily herbaceous vegetation including old fields and pastures; tall-, mixed- and short-grass prairies; alpine tundra; montane meadow; savannah and shrub-steppe; desert-, mesic- and semi-arid grasslands; and annual grasslands, which we refer to collectively as grasslands[65]. These sites encompass a wide range of environmental conditions including elevation (0.5–4,241 m above sea level), mean annual precipitation (203–2,114 mm per year; Extended Data Fig. 1a), mean annual temperature (−3.3 to 27.3 °C; Extended Data Fig. 1a) and latitude (69° N–52° S; Extended Data Fig. 1b). These sites also include a wide range of site means of community biomass (26.2–1,122.9 g m^{-2}), local mean (alpha) plant richness (1.7–43.1 species per 1 m^2 quadrat) and site-level (gamma) plant richness (14–128 species per site). Detailed information about the 76 NutNet study sites is presented in Supplementary Table 1.

At each site, local researchers established an identical experiment that manipulates the supplies of various biologically limiting elements (for example, nitrogen, phosphorus, potassium and various micronutrients) and the density of mammalian grazers. The experiment is composed of 10 treatments applied at the scale of 5 m × 5 m plots and replicated within each site in a completely randomized block design, with most sites having three complete blocks (range 1–6 blocks per site). Each sampling area was separated by more than 1.5 m from neighbouring plots (1-m walkway and 0.5-m within-plot buffer) to minimize spillover effects of treatments on adjacent plots.

The treatments included a factorial combination of three nutrient treatments (N, P and Kμ), each at two levels (control, and nutrient added) for a total of eight treatments. The last two treatments use

https://doi.org/10.1038/s41559-025-02701-y

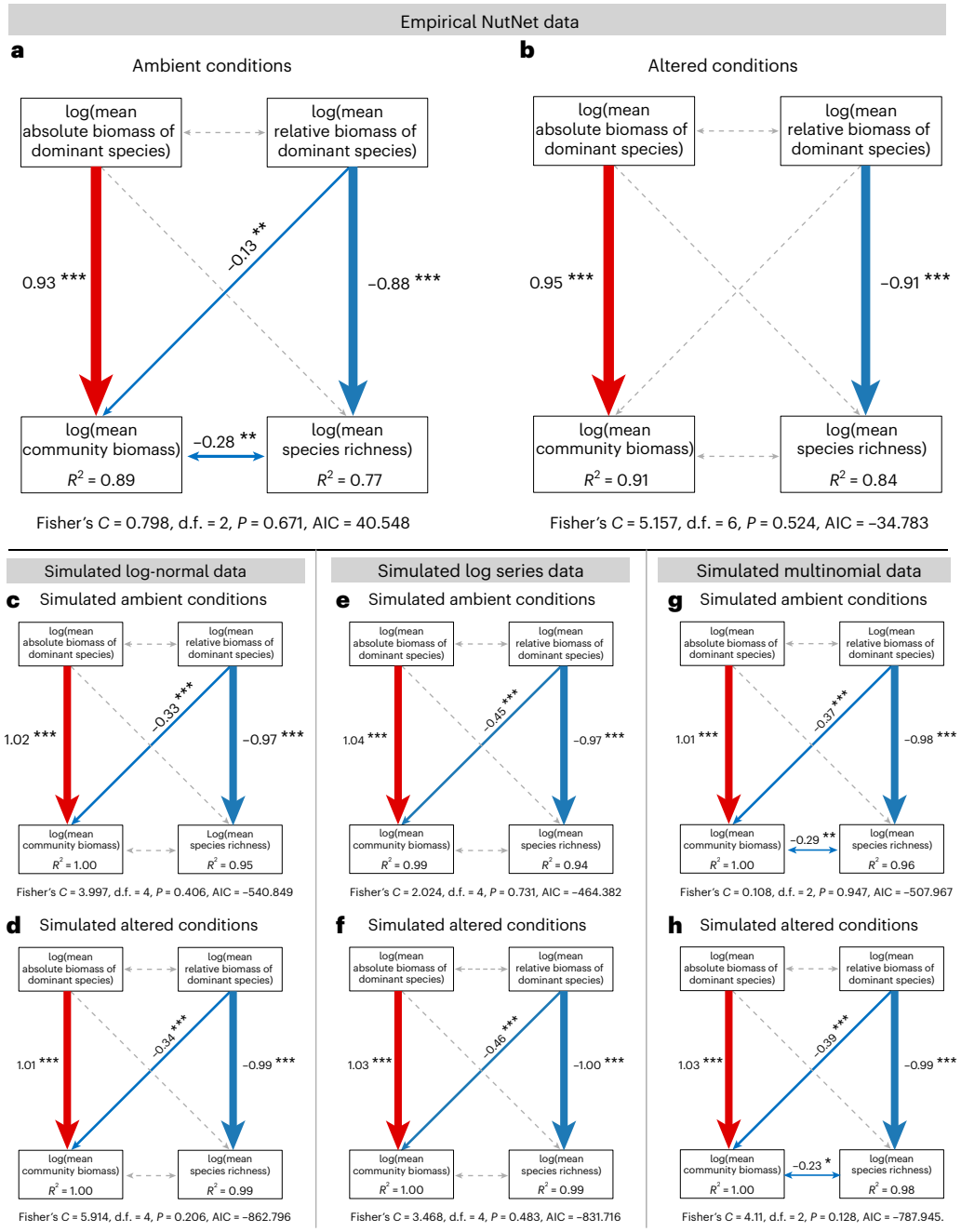


Fig. 5 | Structural relationships across empirical and simulated grassland data under ambient and altered conditions. a–h, SEMs under ambient (**a**) and altered (**b**) environmental conditions from empirical NutNet grassland data and under simulated ambient (**c**,**e**,**g**) and simulated altered (**d**,**f**,**h**) conditions from simulated log-normal data (**c**,**d**), simulated log series data (**e**,**f**) and simulated multinomial data (**g**,**h**). All data were natural log transformed. The red and blue lines mean significantly positive and negative relationships, respectively. The conditional R^2 for each component model is given in the box of response variables. The dashed and solid lines indicate that the relationship is not significant ($P > 0.05$) and significant ($P < 0.05$), respectively. The value next to each solid line represents the standardized regression coefficient for that path, while the symbol in the upper right indicates the significance level of the regression coefficient (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). All statistical tests were conducted as two sided. The goodness-of-fit metrics for each model in **a**–**h**, including Fisher's *C*, degrees of freedom, *P* value and AIC, are presented beneath each model. The detailed bivariate relationships among the four variables of the simulated data are analysed in Extended Data Figs. 5–7 (simulated log-normal data, Extended Data Fig. 5; simulated log series data, Extended Data Fig. 6; and simulated multinomial data, Extended Data Fig. 7).

fencing to exclude herbivores: fencing without nutrient added (fencing) and fencing with all nutrients added (fencing plus NPKμ). In total, the ten different treatments are as follows: control, N, P, Kμ, NP, NKμ, PKμ, NPKμ, fencing and fencing plus NPKμ.

The specific nutrient treatments were as follows: 10 g N m^{-2} yr^{-1} as timed-release urea ($(NH_2)_2CO$), 10 g P m^{-2} yr^{-1} as triple superphosphate ($Ca(H_2PO_4)_2$), 10 g K m^{-2} yr^{-1} as potassium sulphate (K_2SO_4) and 100 g m^{-2} yr^{-1} of a micronutrient mix containing Fe (15%), S (14%), Mg (1.5%), Mn (2.5%), Cu (1%), Zn (1%), B (0.2%) and Mo (0.05%). Micronutrients were added only in the first year to prevent build-up to toxic levels. Fences were 230 cm tall with the lower 90 cm surrounded by a 1-cm woven wire mesh. An additional 30-cm outward-facing flange was stapled to the ground to exclude digging animals, though not fully subterranean animals.

The local scientists collect identical data using standardized protocols starting in the year before the application of experimental treatments (ambient conditions) and in the years following the treatments (experimentally altered conditions). The first set of sites were established in 2007 and include 1 observational year and 15 post-treatment years, and the newest sites have 1 observational year and 1 post-treatment year. Taken together, our data include 21,233 unique site–block–treatment–year combinations, including 18,705 plots representing altered conditions and 2,528 plots under ambient conditions, collected over a span of up to 16 years (range of 2–16 years) across 76 sites (Extended Data Fig. 1b and Supplementary Table 1).

Sampling protocols

Within each 5 m × 5 m plot, a randomly designated 1 m × 1 m quadrat was permanently marked and sampled annually at peak plant biomass. In each quadrat, absolute cover was visually estimated to the nearest 1% for every species rooted within the quadrat. The cover of each species per quadrat was estimated independently, such that the total summed absolute cover of all species (that is, community-level absolute cover) can exceed 100% in multilayer canopies. Species richness per quadrat was determined by the number of species in the cover data. The relative cover of each species was calculated as the ratio of its absolute cover to the community-level absolute cover, ensuring that the sum of the relative cover of all species within each quadrat equalled 100% (Extended Data Fig. 1e).

Adjacent to the permanent 1 m × 1 m cover quadrat, aboveground live biomass was estimated by clipping all aboveground biomass at ground level within two 1 m × 0.1 m strips (totalling 0.2 m^2), with the location of these strips being moved each year. All biomass samples were dried at 60 °C to constant mass and then weighed to the nearest 0.01 g. The weights were multiplied by five to estimate grams per square metre, representing the community biomass of the 1 m^2 quadrat (Extended Data Fig. 1d).

Species absolute and relative biomass

At 7 of the 76 sites, the aboveground live biomass samples were sorted to the species level. At these sites, we identified robust power law relationships between species-level absolute biomass and absolute cover[47], despite the data being derived independently from biomass strips and cover quadrats (Extended Data Fig. 2a). These robust correlations allowed us to estimate species absolute biomass from community biomass and species cover[47], even though direct species-level biomass data were collected only at these seven sites. The absolute biomass of each species in a quadrat was calculated by multiplying the community biomass by the proportion of absolute cover of the corresponding species in that quadrat. Our results were similar whether we used absolute or relative cover (results not shown). The calculation for the relative biomass of each species followed the same principle as that for relative cover. It was determined by the ratio of the absolute biomass of each species to the total community biomass such that the sum of the relative biomass of all species within each quadrat was equal to 100%.

Simulated long-tailed-distributed data

We generated three different types of long-tailed-distributed data (for example, log-normal distribution, Fisher's log series distribution and multinomial distribution) that matched the mean and standard deviation of species richness and community biomass in the 76 NutNet sites. To achieve this, we initially planned to generate 30,000 simulated community-level data points from 100 simulated sites, each with 300 simulated plots (3 blocks × 10 treatments × 10 years) per site, over a simulated period of 10 years (that is, year 0 of simulated ambient conditions and year 1–9 of simulated altered environmental conditions). To ensure that the simulated data closely resembled real-world grassland community characteristics, the mean and standard deviation were set to 12 and 7 for species richness, and 360 g and 240 g for community biomass, respectively, reflecting the ranges of values observed at the 76 NutNet sites.

The simulated data conformed to a log-normal distribution at the community level. From 200,000 iterations (100 sites × 2,000 iterations), we filtered 27,222 eligible data points by limiting community species richness to between 1 and 50, and community biomass to between 1 g and 8,000 g. The community biomass in each of these 27,222 eligible data points was then allocated to the number of species indicated by the corresponding species richness, following the log-normal distribution, the Fisher's log series distribution and the multinomial distribution, respectively. The zero-sum multinomial distribution requires that species abundances are distributed based on a competitive process in which gains by one species must result in losses for others. However, the total abundance of our observational (that is, ambient conditions) and experimental (that is, altered environmental conditions) data across 76 grassland sites on 6 continents is usually not constant, so this precondition cannot be satisfied. Therefore, we simulated the multinomial distribution without forcing the total abundance to be constant. Ultimately, each of the 100 simulated sites had approximately 270 plot-level values for community biomass and species richness. These 270 or so plots in each simulated site were randomly assigned to 1 of 10 different simulated treatments, 3 simulated blocks and 10 simulated years from year 0 to year 9, analogous to those in NutNet data. In total, the 27,222 plots contained 341,452 species-level values for absolute biomass and relative biomass.

Statistical analysis

To investigate the relationships among community biomass, species richness, absolute abundance of dominant species and relative abundance of dominant species under both ambient and altered environmental conditions using (1) a mathematical formula derivation and (2) empirical NutNet data and simulated data from long-tailed distributions, we conducted the following statistical analyses.

Mathematical modelling of the relationship between the relative abundance of dominant species and species richness. In this section, we provide further details on the derivation of the relationship between (1) the relative abundance of the single most dominant species and species richness in a site based on (i) log-normally distributed and (ii) Fisher's log series-distributed data, and also showed that this relationship (2) extends to the sum of the two most dominant species. This calculation elucidated how the connections between the relative abundance of dominant species and species richness are predicted consequences of the log-normal distribution of species abundance. While our mathematical modelling has focused on the log-normal distribution and Fisher's log series distribution, the shared common characteristic of a long-tailed distribution ensures the robustness of our approach, with potential variation only in the coefficient.

For modelling of the single most dominant species based on (i) log-normally distributed and (ii) Fisher's log series-distributed data, recall that the abundance of the most dominant species in the plot is $A_1(S) \equiv \max\{Y_1, Y_2, \ldots Y_S\}$ where the Y_i are independent and identically distributed random variables representing the abundance of each species. Recall that the tail probability function, which depends on the distribution of Y_i, is denoted by $\bar{F}(y) \equiv P(Y_i > y)$. In this section, we show that both the log-normal and log series distributions satisfy the condition provided in equation (5):

$$\frac{\log(b_S)}{\log(S)} \to 0$$

for

$$b_S \equiv \bar{F}^{-1}(1/S).$$

For log-normally distributed species abundance, suppose the Y_i are log-normally distributed with parameters μ and σ (mean and standard deviation of the logarithm, respectively). Note that as $\bar{F}$ is a strictly decreasing function, its inverse is also strictly decreasing. Therefore, if an α exists such that $\bar{F}(a) \le 1/S$, then $b_S = \bar{F}^{-1}(1/S) \le a$. For positive x, define the function

$$\ell(x) = \frac{\log(x) - \mu}{\sigma},$$

we can then use Mill's inequality[66] (and the cumulative distribution function of log-normal) to obtain the upper bound

$$\bar{F}(x) \le \frac{1}{\sqrt{2\pi}\ell(x)} \exp\left[-\frac{\ell(x)^2}{2}\right]. \quad (6)$$

Define

$$a_S = e^{\mu} \exp\left[\sigma\left(2\log\left(\frac{S}{\sqrt{2\pi}}\right)\right)^{1/2}\right].$$

We will now use equation (6) to verify that $\bar{F}(a_S) \le 1/S$. First, check that

$$\ell(a_S) = \left(2\log\left(\frac{S}{\sqrt{2\pi}}\right)\right)^{1/2},$$

$$\frac{\ell(a_S)^2}{2} = \log\left(\frac{S}{\sqrt{2\pi}}\right),$$

and

$$\exp\left[-\frac{\ell(a_S)^2}{2}\right] = \frac{\sqrt{2\pi}}{S}.$$

Therefore,

$$\bar{F}(a_S) \le \frac{1}{\sqrt{2\pi}\ell(a_S)} \exp\left[-\frac{\ell(a_S)^2}{2}\right] = \frac{1}{S}\frac{1}{\sqrt{2\log\left(S/\sqrt{2\pi}\right)}} \le \frac{1}{S},$$

and we conclude that

$$b_S \le a_S = e^{\mu} \exp\left[\sigma\left(2\log\left(\frac{S}{\sqrt{2\pi}}\right)\right)^{1/2}\right],$$

or

$$\log b_S \le \log a_S = \mu + \sqrt{2\sigma^2}\left(\log\left(\frac{S}{\sqrt{2\pi}}\right)\right)^{1/2}.$$

Thus, equation (5) is verified.

For log series-distributed species abundance, suppose the Y_i are random variables with log series distribution with parameter $0 < p < 1$. Let Z be a geometrically distributed random variable with parameter p and tail distribution $\bar{F}_Z(x)$. Note that for $k \in \mathbb{N}$, the tail distribution of Y_i is

$$\bar{F}(k) = \sum_{j=k+1}^{\infty} -\frac{1}{\log(1-p)}\frac{p^j}{j}$$

and the tail distribution of Z is

$$\bar{F}_Z(k) = \sum_{j=k+1}^{\infty} (1-p)p^j.$$

Therefore, there exists a constant, C_1, independent of k, such that $\bar{F}_Z(k) \ge C_1\bar{F}(k)$, so if

$$\frac{\log \bar{F}_Z^{-1}(1/S)}{\log(S)} \to 0$$

holds, this implies that equation (5) holds for the log series tail probability $\bar{F}$ as well. Define $\hat{k}$ as follows:

$$\bar{F}_Z(\hat{k}) = \sum_{j=\hat{k}+1}^{\infty} (1-p)p^j = \frac{1}{S}.$$

Solving for $\hat{k}$ yields

$$p^{\hat{k}+1} = 1/S$$

$$(\hat{k}+1)\log(p) = -\log(S)$$

$$\hat{k} = C_2\log(S),$$

so

$$\frac{\log\left(\bar{F}_Z^{-1}\right)(1/S)}{\log(S)} \sim \frac{\log(\log(S))}{\log(S)} \to 0.$$

For extension to the top two dominant species, we show that our modelling results extend to the sum of the top two dominant relative abundances. To this end, denote the index of the maximally abundant species in the site $i_{\max}$. Then, the abundance of the second most dominant species is defined as $A_2(S) \equiv \max\{Y_i \text{ for } i \in \{1,\ldots,S\}/i_{\max}\}$. Then, consider the sum of the top two dominant relative abundances:

$$T \equiv \frac{A_1(S) + A_2(S)}{\sum_{i=1}^{S} Y_i} \sim \frac{A_1(S) + A_2(S)}{S \times E[Y_i]}.$$

Taking the logarithm of both sides yields

$$\log(T) \sim \log(A_1(S) + A_2(S)) - \log(S) - \log(E[Y_i]).$$

Thus, a nearly linear relationship between $\log(T)$ and $\log(S)$ is again ensured as long as the following condition holds

$$\frac{\log(A_1(S)+A_2(S))}{\log(S)} \to 0. \quad (7)$$

Note that due to the trivial upper bound

$$\log(A_1(S) + A_2(S)) \le \log(2A_1(S)),$$

the second condition (Using Empirical NutNet Data and Simulated Data from Long-Tailed Distributions) then holds for all distributions satisfying equation (5) in the main text.

Using empirical NutNet data and simulated data from long-tailed distributions. We ranked all species in each plot from highest to lowest in terms of absolute biomass using the 'BiodiversityR' package[67]. This ranking enabled us to construct cumulative absolute and relative biomass distribution curves for each plot and even for each site (Fig. 1b,c and Extended Data Fig. 1d,e). It is important to note that ranking species within each plot can result in different species occupying the same rank across different plots[43,44]. Our analysis thus focuses on abundance distributions rather than specific species, making it applicable to sites with diverse grassland floras and varying abiotic environmental conditions.

Species in each plot were divided into two groups: dominant and non-dominant species, based on their biomass ranking. The absolute and relative biomass of dominant species in each plot could then be calculated. For instance, the two species with the highest biomass in each plot were categorized as dominant, while all others were classified as non-dominant (Extended Data Fig. 1d,e). To ensure the robustness of our analysis, we repeated the analyses using data ranging from one to five dominant species (Table 1). All results were qualitatively consistent, and we focus on results from the two most dominant species, because this level had a slightly higher correlation with community biomass, richness and gamma diversity (see R^2 of selecting different numbers of dominant species in Table 1 (top and bottom)). We natural log-transformed values of community biomass, species richness, and absolute and relative biomass of dominant species at the plot level before analysis.

We used mixed-effects models with the 'nlme' package[68] to determine the pairwise bivariate relationships among our four focal variables for each site (community biomass, species richness, absolute biomass of dominant species and relative biomass of dominant species) (Extended Data Fig. 3). Sites and blocks nested within sites were incorporated as random effects, allowing for variability in both intercepts and slopes of the regression among sites, contingent on model selection outcomes. To address interactions between fixed and random effects, we used a model-selection approach grounded in the minimization of the Bayesian information criterion as per Pinheiro and Bates' methodology[68]. This approach involved comparing models with and without each random effect to ascertain the necessary level of variation to be included in the model. Consistently, model selection favoured the inclusion of variability among sites while excluding variability attributable to blocks. Based on these analyses, we categorized sites as having significantly positive, non-significant or significantly negative relationships (Extended Data Fig. 3).

We also examined the relationships among the site-level means of these four variables under ambient conditions (year = 0), altered environmental conditions (year > 0), and each year (from year 1 to year 15 of NutNet data; Supplementary Table 2a) and each treatment (control, N, P, Kμ, NP, NKμ, PKμ, NPKμ, fencing and fencing plus NPKμ of NutNet data; Supplementary Table 2b) of altered environmental conditions. We developed linear models of the pairwise bivariate relationships among these four variables across 76 sites under both ambient and altered environmental conditions (Figs. 2 and 3) and across 100 simulated sites under both simulated ambient and simulated altered environmental conditions (Extended Data Figs. 5–7). In addition, we calculated gamma diversity in NutNet data under both ambient and altered environmental conditions based on species count at each site and analysed its relationship with the mean of the relative biomass of dominant species (Extended Data Fig. 4).

By comparing the predicted values from regression models of variables under ambient conditions with the actual measured values under altered environmental conditions (Fig. 4), we estimated the impact of environmental changes on these relationships in NutNet data.

Moreover, by comparing the identities of the dominant species after treatment (years > 0) with those in the year before treatment (year = 0) within the same quadrat, we quantified the count and proportion of dominant species occurrences that remained consistent before and after treatment. For instance, when comparing the two (or one, three, four or five) most dominant species, if the pre-treatment dominant species in a quadrat are sp1 and sp2, and post-treatment they change to sp1 and sp3 or sp2 and sp3, the count of consistency is 1, resulting in a proportion of 50%. If the species remain as sp1 and sp2 after treatment, the count is 2 with a proportion of 100%. Conversely, if they shift to sp3 and sp4, the count is 0 with a proportion of 0%. This analysis enables us to assess the proportion of dominant species that remain unchanged after treatments in both NutNet data and the simulated data (Supplementary Table 3a–d).

Finally, we constructed a piecewise SEM (Fig. 1a and Extended Data Fig. 1c) with the 'piecewiseSEM' package[69] to assess relationships among these four variables across 76 sites (Fig. 5a,b) and across 100 simulated sites of three different types of long-tail-distributed simulated data (Fig. 5c–h) after accounting for correlations among the variables. In light of the conceptual framework depicted in Fig. 1a and Extended Data Fig. 1c, we assumed that both absolute and relative biomass of dominant species could affect community biomass and species richness, and that there is a covariance relationship between the absolute and relative biomass of dominant species and between community biomass and species richness. The piecewise SEM was evaluated using Fisher's C statistic, P value (where $P > 0.05$ indicated a good fit) and the Akaike Information Criterion (AIC)[69]. To verify that our results do not depend on the type of species-level abundance measure used, we repeated the identical analysis based on species-level absolute and relative cover data from this experiment (Supplementary Figs. 1–5 and Table 4).

All analyses were conducted in R 4.3.1 (ref. 70).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All the data that support the findings of this article are freely available via the Environmental Data Initiative (EDI) Data Portal (https://doi.org/10.6073/pasta/442895326274ea09942bd04e6ea92df2)[71].

Code availability

The R code used to perform the analyses is freely available via the EDI Data Portal (https://doi.org/10.6073/pasta/442895326274ea09942bd04e6ea92df2)[71].

References

1. Willig, M. R. Biodiversity and productivity. *Science* **333**, 1709–1710 (2011).
2. Mittelbach, G. G. et al. What is the observed relationship between species richness and productivity? *Ecology* **82**, 2381–2396 (2001).
3. Grace, J. B. et al. Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529**, 390–393 (2016).
4. Lavers, C. & Field, R. A resource-based conceptual model of plant diversity that reassesses causality in the productivity–diversity relationship. *Glob. Ecol. Biogeogr.* **15**, 213–224 (2006).
5. Reich, P. B. et al. Impacts of biodiversity loss escalate through time as redundancy fades. *Science* **336**, 589–592 (2012).
6. Tilman, D., Reich, P. B. & Isbell, F. Biodiversity impacts ecosystem productivity as much as resources, disturbance, or herbivory. *Proc. Natl Acad. Sci. USA* **109**, 10394–10397 (2012).
7. Huang, Y. et al. Impacts of species richness on productivity in a large-scale subtropical forest experiment. *Science* **362**, 80–83 (2018).
8. Bongers, F. J. et al. Functional diversity effects on productivity increase with age in a forest biodiversity experiment. *Nat. Ecol. Evol.* **5**, 1594–1603 (2021).
9. Li, C. et al. The productive performance of intercropping. *Proc. Natl Acad. Sci. USA* **120**, e2201886120 (2023).
10. Grace, J. B. et al. Does species diversity limit productivity in natural grassland communities? *Ecol. Lett.* **10**, 680–689 (2007).
11. Adler, P. B. et al. Productivity is a poor predictor of plant species richness. *Science* **333**, 1750–1753 (2011).
12. Fraser, L. H. et al. Worldwide evidence of a unimodal relationship between productivity and plant species richness. *Science* **349**, 302–305 (2015).

13. Tredennick, A. T. et al. Comment on "Worldwide evidence of a unimodal relationship between productivity and plant species richness". *Science* **351**, 457 (2016).
14. Liang, J. et al. Positive biodiversity–productivity relationship predominant in global forests. *Science* **354**, aaf8957 (2016).
15. Duffy, J. E., Godwin, C. M. & Cardinale, B. J. Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature* **549**, 261–264 (2017).
16. Dee, L. E. et al. Clarifying the effect of biodiversity on productivity in natural ecosystems with longitudinal data and methods for causal inference. *Nat. Commun.* **14**, 2607 (2023).
17. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
18. Magurran, A. E. & Henderson, P. A. Explaining the excess of rare species in natural species abundance distributions. *Nature* **422**, 714–716 (2003).
19. McGill, B. J. et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* **10**, 995–1015 (2007).
20. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58 (1943).
21. MacArthur, R. H. On the relative abundance of bird species. *Proc. Natl Acad. Sci. USA* **43**, 293–295 (1957).
22. Whittaker, R. H. Dominance and diversity in land plant communities: numerical relations of species express the importance of competition in community function and evolution. *Science* **147**, 250–260 (1965).
23. Smith, M. D. & Knapp, A. K. Dominant species maintain ecosystem function with non-random species loss. *Ecol. Lett.* **6**, 509–517 (2003).
24. Roswell, M., Dushoff, J. & Winfree, R. A conceptual guide to measuring species diversity. *Oikos* **130**, 321–338 (2021).
25. Cooper, D. L. M. et al. Consistent patterns of common species across tropical tree communities. *Nature* **625**, 728–734 (2024).
26. Grime, J. P. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* **86**, 902–910 (1998).
27. Dangles, O. & Malmqvist, B. Species richness–decomposition relationships depend on species dominance. *Ecol. Lett.* **7**, 395–402 (2004).
28. Winfree, R., Fox, J. W., Williams, N. M., Reilly, J. R. & Cariveau, D. P. Abundance of common species, not species richness, drives delivery of a real-world ecosystem service. *Ecol. Lett.* **18**, 626–635 (2015).
29. Avolio, M. L. et al. Demystifying dominant species. *New Phytol.* **223**, 1106–1126 (2019).
30. Preston, F. W. The commonness, and rarity, of species. *Ecology* **29**, 254–283 (1948).
31. Limpert, E., Stahel, W. A. & Abbt, M. Log-normal distributions across the sciences: keys and clues. *BioScience* **51**, 341–352 (2001).
32. Ulrich, W., Ollik, M. & Ugland, K. I. A meta-analysis of species–abundance distributions. *Oikos* **119**, 1149–1155 (2010).
33. Callaghan, C. T., Borda-de-Água, L., Van Klink, R., Rozzi, R. & Pereira, H. M. Unveiling global species abundance distributions. *Nat. Ecol. Evol.* **7**, 1600–1609 (2023).
34. Callaghan, C. T., Santini, L., Spake, R. & Bowler, D. E. Population abundance estimates in conservation and biodiversity research. *Trends Ecol. Evol.* **39**, 515–523 (2024).
35. McNaughton, S. J. & Volf, L. L. Dominance and the niche in ecological systems: dominance is an expression of ecological inequalities arising out of different exploitation strategies. *Science* **167**, 131–139 (1970).
36. Leibold, M. A., Chase, J. M. & Ernest, S. K. M. Community assembly and the functioning of ecosystems: how metacommunity processes alter ecosystems attributes. *Ecology* **98**, 909–919 (2017).
37. Bannar-Martin, K. H. et al. Integrating community assembly and biodiversity to better understand ecosystem function: the Community Assembly and the Functioning of Ecosystems (CAFE) approach. *Ecol. Lett.* **21**, 167–180 (2018).
38. Harte, J., Brush, M., Newman, E. A. & Umemura, K. An equation of state unifies diversity, productivity, abundance and biomass. *Commun. Biol.* **5**, 874 (2022).
39. Ladouceur, E. et al. Linking changes in species composition and biomass in a globally distributed grassland experiment. *Ecol. Lett.* **25**, 2699–2712 (2022).
40. Nair, J., Wierman, A. & Zwart, B. *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation* Vol. 53 (Cambridge Univ. Press, 2022).
41. Bock, C. E., Jones, Z. F. & Bock, J. H. Relationships between species richness, evenness, and abundance in a southwestern savanna. *Ecology* **88**, 1322–1327 (2007).
42. Locey, K. J. & White, E. P. How species richness and total abundance constrain the distribution of abundance. *Ecol. Lett.* **16**, 1177–1185 (2013).
43. Zhang, P. et al. SRU_D: a simple non-destructive method for accurate quantification of plant diversity dynamics. *J. Ecol.* **107**, 2155–2166 (2019).
44. Zhang, P. et al. Space resource utilization of dominant species integrates abundance- and functional-based processes for better predictions of plant diversity dynamics. *Oikos* **4**, e09519 (2023).
45. Borer, E. T. et al. Finding generality in ecology: a model for globally distributed experiments. *Methods Ecol. Evol.* **5**, 65–73 (2014).
46. Wilfahrt, P. A. et al. Nothing lasts forever: dominant species decline under rapid environmental change in global grasslands. *J. Ecol.* **111**, 2472–2482 (2023).
47. Pan, X. et al. The convex relationship between plant cover and biomass: implications for assessing species and community properties. *J. Veg. Sci.* **35**, e13288 (2024).
48. De Haan, L. Sample extremes: an elementary introduction. *Stat. Neerl.* **30**, 161–172 (1976).
49. Goldberg, D. E. & Barton, A. M. Patterns and consequences of interspecific competition in natural communities: a review of field experiments with plants. *Am. Nat.* **139**, 771–801 (1992).
50. Mac Nally, R. Use of the abundance spectrum and relative–abundance distributions to analyze assemblage change in massively altered landscapes. *Am. Nat.* **170**, 319–330 (2007).
51. Chao, A., Chiu, C. H. & Jost, L. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annu. Rev. Ecol. Evol. Syst.* **45**, 297–324 (2014).
52. Kissling, W. D. et al. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biol. Rev.* **93**, 600–625 (2018).
53. Pearse, I. S., Sofaer, H. R., Zaya, D. N. & Spyreas, G. Non-native plants have greater impacts because of differing per-capita effects and nonlinear abundance–impact curves. *Ecol. Lett.* **22**, 1214–1220 (2019).
54. Gotelli, N. J. et al. Estimating species relative abundances from museum records. *Methods Ecol. Evol.* **14**, 431–443 (2023).
55. Dawson, G. The usefulness of absolute ("census") and relative ("sampling" or "index") measures of abundance. *Stud. Avian Biol.* **6**, 554–558 (1981).
56. Theodose, T. A. & Bowman, W. D. Nutrient availability, plant abundance, and species diversity in two alpine tundra communities. *Ecology* **78**, 1861–1872 (1997).
57. Royle, J. A. & Nichols, J. D. Estimating abundance from repeated presence–absence data or point counts. *Ecology* **84**, 777–790 (2003).

58. Chao, A., Hsieh, T. C., Chazdon, R. L., Colwell, R. K. & Gotelli, N. J. Unveiling the species-rank abundance distribution by generalizing the Good–Turing sample coverage theory. *Ecology* **96**, 1189–1201 (2015).
59. Burton, A. C. et al. Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *J. Appl. Ecol.* **52**, 675–685 (2015).
60. Cerini, F., Childs, D. Z. & Clements, C. F. A predictive timeline of wildlife population collapse. *Nat. Ecol. Evol.* **7**, 320–331 (2023).
61. Tilman, D. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* **80**, 1455–1474 (1999).
62. Allan, E. et al. More diverse plant communities have higher functioning over time due to turnover in complementary dominant species. *Proc. Natl Acad. Sci. USA* **108**, 17034–17039 (2011).
63. Mallmin, E., Traulsen, A. & De Monte, S. Chaotic turnover of rare and abundant species in a strongly interacting model community. *Proc. Natl Acad. Sci. USA* **121**, e2312822121 (2024).
64. Gilbert, B., Turkington, R. & Srivastava, D. S. Dominant species and diversity: linking relative abundance to controls of species establishment. *Am. Nat.* **174**, 850–862 (2009).
65. Fay, P. A. et al. Grassland productivity limited by multiple nutrients. *Nat. Plants* **1**, 15080 (2015).
66. Karatzas, I. & Shreve, S. E. *Brownian Motion and Stochastic Calculus* (Springer Science & Business Media, 1998).
67. Kindt, R. & Coe, R. *Tree Diversity Analysis: A Manual and Software for Common Statistical Methods for Ecological and Biodiversity Studies* (World Agroforestry Centre, 2005).
68. Pinheiro, J. & Bates, D. *Mixed-Effects Models in S and S-PLUS* (Springer Science & Business Media, 2006).
69. Lefcheck, J. S. PIECEWISESEM: piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).
70. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2023).
71. Borer, E. T. et al. Species cover, community biomass, and richness in global grasslands from NutNet (2007–2023): dominant species predict plant richness and biomass in global grasslands ver 4. *Environmental Data Initiative* https://doi.org/10.6073/pasta/442895326274ea09942bd04e6ea92df2 (2025).

Acknowledgements

We thank each of the researchers who have contributed data and ideas to NutNet (http://www.nutnet.org). Grants to P.Z. came from the National Natural Science Foundation of China (grant number 32101267) and the Start-Up Funds of Introduced Talent in Lanzhou University (grant number 561120205). Azi.cn and azitwo.cn were conducted at the Gansu Gannan Grassland Ecosystem National Observation and Research Station. Coordination and data management in NutNet have been supported by funding to E.T.B. and E.W.S. from the National Science Foundation Research Coordination Network (NSF-DEB-1042132) and Long-Term Ecological Research (NSF-DEB-1234162 to Cedar Creek LTER) programmes, and the Institute on the Environment (DG-0001-13). We also thank the Minnesota Supercomputer Institute for hosting project data and the Institute on the Environment for hosting network meetings. N.E. was funded by the German Research Foundation (DFG, FZT 118, 202548816; Ei 862/29-1). Y.L. was funded by MPG Ranch; N.G.S. was supported by the US National Science Foundation (DEB-2045968) and Texas Tech University. S.M.P. was supported by the Australian government through the Great Western Woodlands TERN SuperSite. G.M.W. and C.R.D. were supported by the Australian Government through the desert ecology TERN sites funded by NCRIS and the ARC. P.T. was funded by UBACyT20020190100212BA, PICT-2019-2019-02324 and Familia Bordeu and Pepo. A.J. was funded by Federal Ministry of Education and Research (BMBF) (grant numbers FKZ 031B0516C and 031B1067C). J.A.C. was funded by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant number 101002987).

Author contributions

Conceptualization: P.Z., E.T.B. and E.W.S. conceived and developed the study with substantial input from J.F., A.S.M., W.S.H., P.B.A., Y.H., N.E. and M.S. Formal analysis: P.Z. performed the analysis, J.F. developed the mathematical formula derivation and J.D.B. contributed to the analysis. Writing—original draft: P.Z. and E.T.B. wrote the paper with substantial input from E.W.S., J.F., A.S.M., W.S.H., P.B.A., N.E. and M.S. Writing—review and editing: all co-authors contributed data and reviewed, approved and had the opportunity to comment on the paper. An author contribution matrix is provided in Supplementary Table 5.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-025-02701-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-025-02701-y.

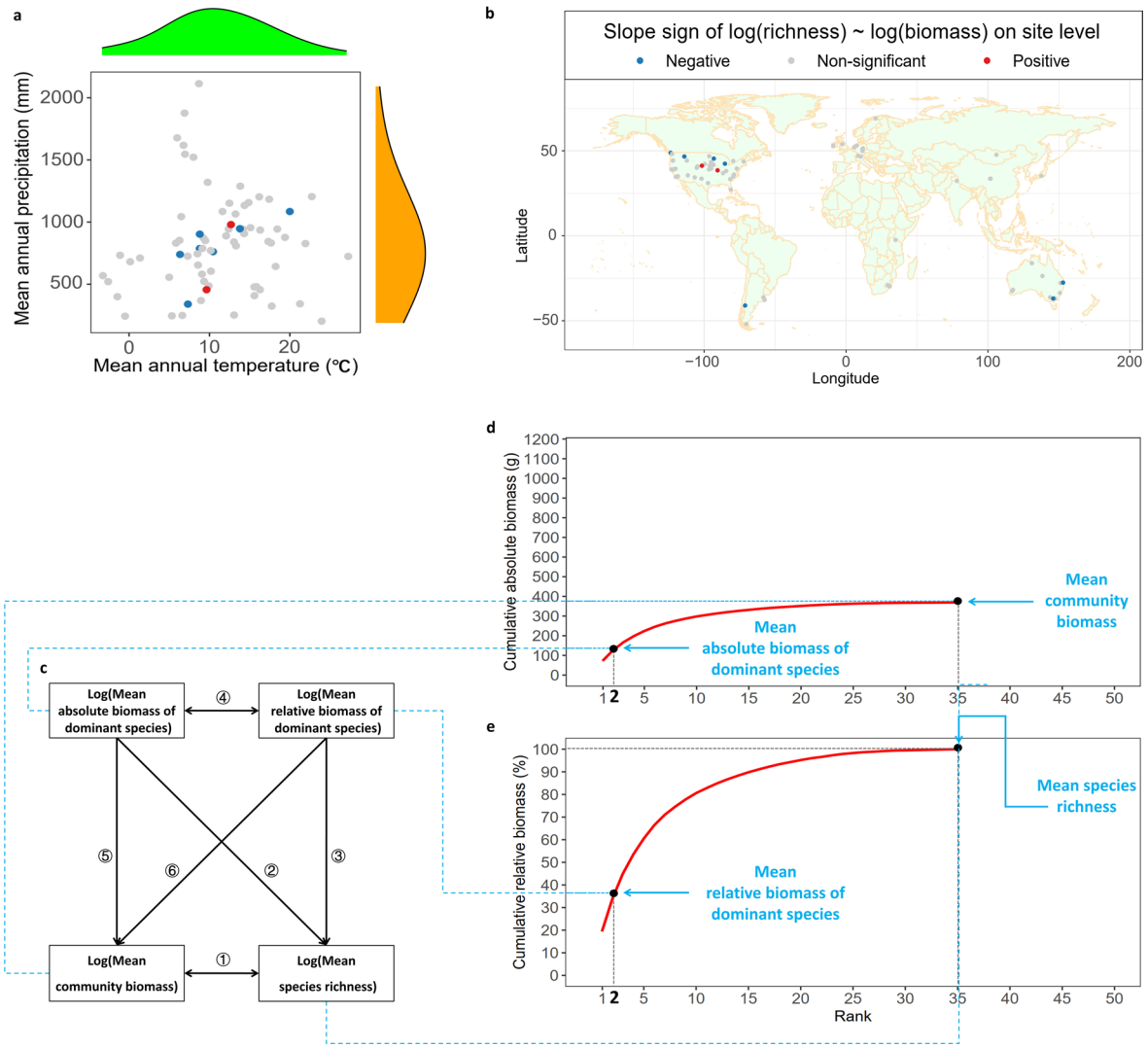
Correspondence and requests for materials should be addressed to Pengfei Zhang.

Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Pengfei Zhang[1,2] ✉, Eric W. Seabloom[2], Jasmine Foo[3], Andrew S. MacDougall[4], W. Stanley Harpole[5,6,7], Peter B. Adler[8], Yann Hautier[9], Nico Eisenhauer[6,10], Marie Spohn[11], Jonathan D. Bakker[12], Ylva Lekberg[13,14], Alyssa L. Young[15], Clinton Carbutt[16,17], Anita C. Risch[18], Pablo L. Peri[19], Nicholas G. Smith[20], Carly J. Stevens[21], Suzanne M. Prober[22], Johannes M. H. Knops[23], Glenda M. Wardle[24], Christopher R. Dickman[24], Anne Ebeling[25], Christiane Roscher[5,6], Holly M. Martinson[26], Jason P. Martina[27], Sally A. Power[28], Yujie Niu[29,30], Zhengwei Ren[1], Guozhen Du[1], Risto Virtanen[31], Pedro Tognetti[32], Michelle J. Tedder[16], Anke Jentsch[29], Jane A. Catford[33] & Elizabeth T. Borer[2]

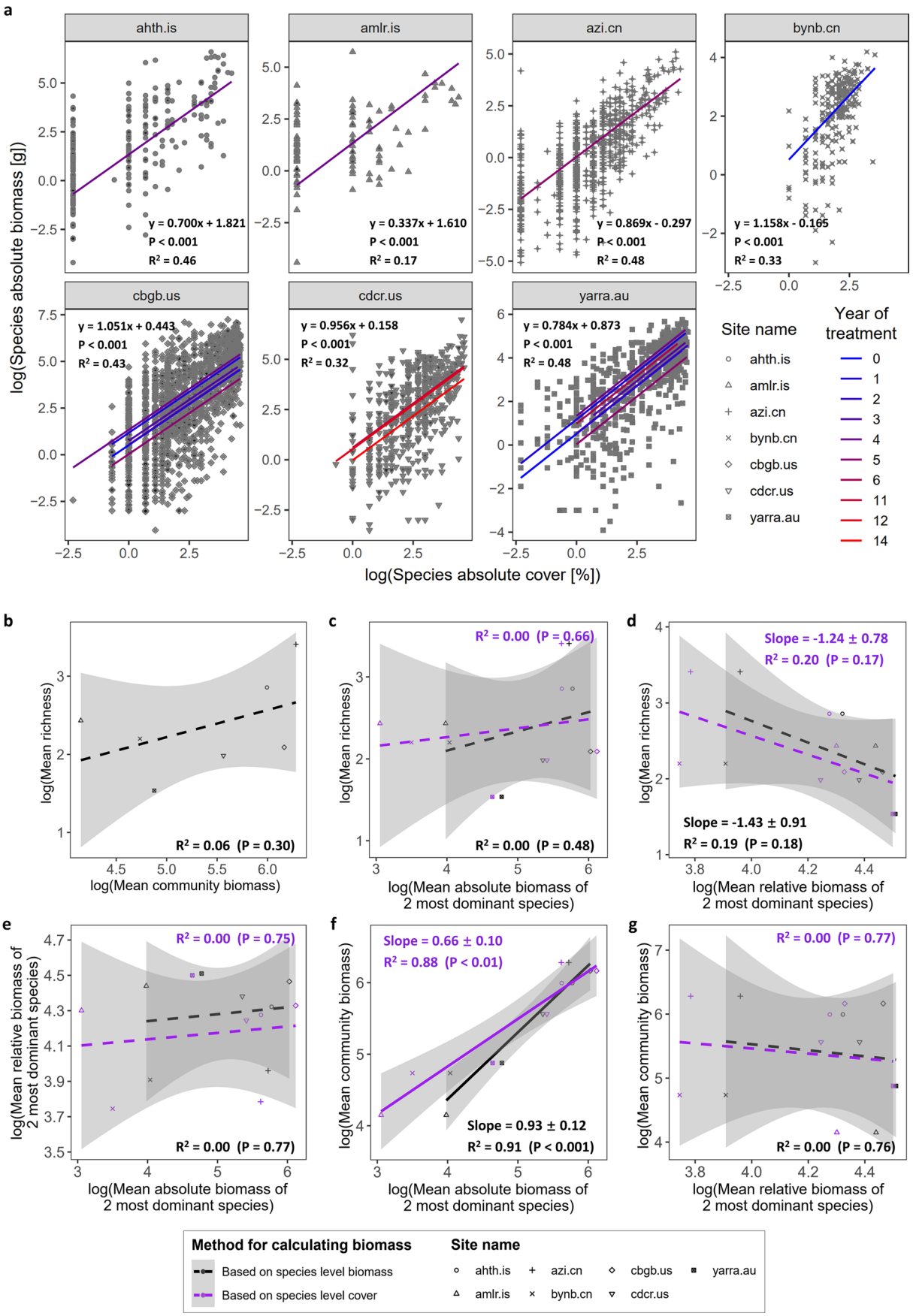
[1]State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, College of Ecology, Lanzhou University, Lanzhou, People's Republic of China. [2]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN, USA. [3]School of Mathematics, University of Minnesota, Minneapolis, MN, USA. [4]Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada. [5]Department of Physiological Diversity, Helmholtz Center for Environmental Research—UFZ, Leipzig, Germany. [6]German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany. [7]Martin Luther University Halle-Wittenberg, Halle, Germany. [8]Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT, USA. [9]Ecology and Biodiversity Group, Department of Biology, Utrecht University, Utrecht, The Netherlands. [10]Institute of Biology, Leipzig University, Leipzig, Germany. [11]Department of Soil and Environment, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden. [12]School of Environmental and Forest Sciences, University of Washington, Seattle, WA, USA. [13]MPG Ranch Missoula, Florence, MT, USA. [14]Department of Ecosystem and Conservation Sciences, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT, USA. [15]Department of Biology, University of North Carolina Greensboro, Greensboro, NC, USA. [16]School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa. [17]Scientific Services, Ezemvelo KZN Wildlife, Pietermaritzburg, South Africa. [18]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland. [19]Instituto Nacional de Tecnología Agropecuaria (INTA)–Universidad Nacional de la Patagonia Austral (UNAP)–CONICET, Río Gallegos, Argentina. [20]Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA. [21]Lancaster Environment Centre, Lancaster University, Lancaster, UK. [22]CSIRO Environment, Canberra, Australian Capital Territory, Australia. [23]Health and Environmental Sciences Department, Xi'an Jiaotong-Liverpool University, Suzhou, People's Republic of China. [24]Desert Ecology Research Group, School of Life and Environmental Sciences and ARC Training Centre in Data Analytics for Resources and Environments (DARE), The University of Sydney, Sydney, New South Wales, Australia. [25]Institute of Ecology and Evolution, University Jena, Jena, Germany. [26]Department of Biology, McDaniel College, Westminster, MD, USA. [27]Department of Biology, Texas State University, San Marcos, TX, USA. [28]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales, Australia. [29]Department of Disturbance Ecology and Vegetation Dynamics, Bayreuth Center of Ecology and Environmental Research (Bayceer), University of Bayreuth, Bayreuth, Germany. [30]College of Grassland Science, Key Laboratory of Grassland Ecosystem of the Ministry of Education, Gansu Agricultural University, Lanzhou, People's Republic of China. [31]Ecology & Genetics, University of Oulu, Oulu, Finland. [32]IFEVA, Facultad de Agronomía, Universidad de Buenos Aires-CONICET, Buenos Aires, Argentina. [33]Department of Geography, King's College London, London, UK. ✉e-mail: zhangpengfei@lzu.edu.cn

https://doi.org/10.1038/s41559-025-02701-y



Extended Data Fig. 1 | Climatic (a) and geographic (b) distribution of 76 experimental sites from NutNet data; and (c) structural equation meta-model to characterize the possible relationships among log(mean absolute biomass of dominant species), log(mean relative biomass of dominant species), log(mean community biomass), and log(mean species richness); and the link between these four variables in the structural equation model and one site-level patterns of (d) cumulative absolute biomass curve and (e) cumulative relative biomass curve taking azi.cn site under ambient conditions as an example. Here, species-level abundance is estimated based on species-level biomass. The red, gray, and blue dots in Fig. a and b mean that the relationship between log(species richness) and log(community biomass) on the site-level under ambient conditions (year = 0) is significantly positive, non-significant correlated, and significantly negative correlated, respectively (see Extended Data Fig. 3a).

https://doi.org/10.1038/s41559-025-02701-y

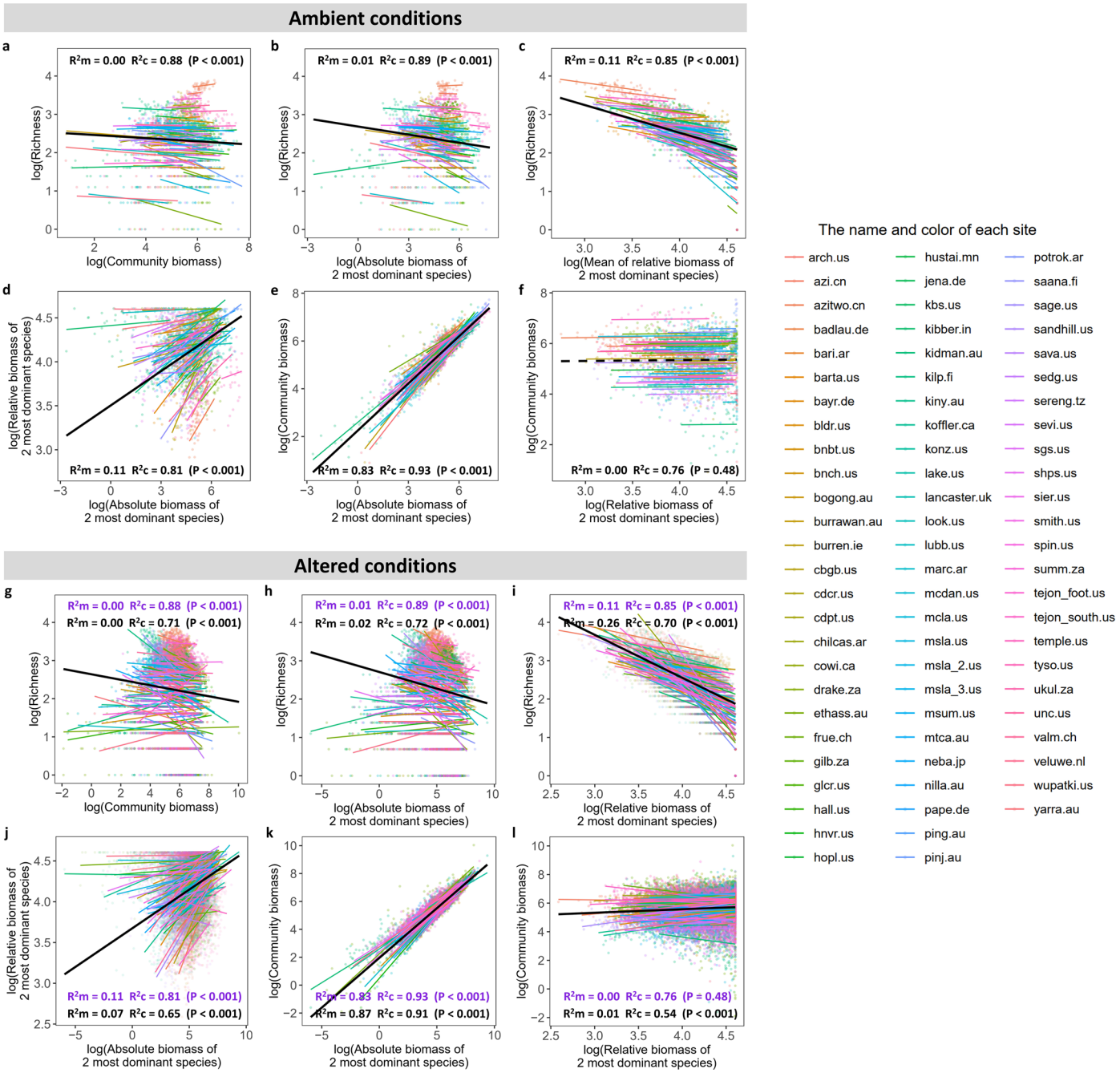


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41559-025-02701-y

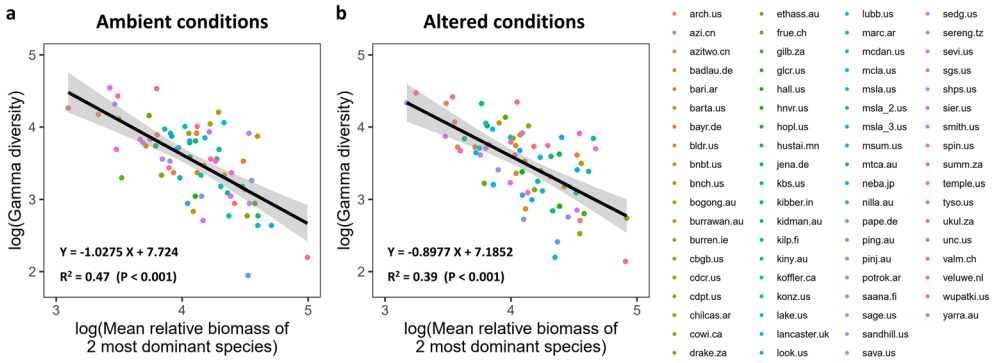
Extended Data Fig. 2 | (a) The relationship between directly measured species-level absolute biomass and species-level absolute cover for the 7 sites that simultaneously collected these data; and (b-g) patterns of these 7 sites based on species level biomass data (black) and species level cover data (purple). All data were natural log-transformed. Each regression curve in Fig. a represents one year (in different colors). Relationships between mean richness and (**b**) mean community biomass, and (**c**) mean absolute biomass of the two most dominant species, and (**d**) mean relative biomass of the two most dominant species; and (**e**) between mean relative biomass and mean absolute biomass of the two most dominant species; and between mean community biomass and (**f**) mean absolute biomass of the two most dominant species, and (**g**) mean relative biomass of the two most dominant species, of these 7 sites. The black dots and lines in Fig. b to g are the results of each pattern by directly calculating the absolute and relative biomass of the two most dominant species based on the species level biomass data. The purple dots and lines in Fig. b to g are the results of each pattern by indirect calculations of the absolute and relative biomass of the two dominant species based on species level cover data. The dashed and solid lines indicate that the overall relationship is not significant ($P > 0.05$) and significant ($P < 0.05$), respectively, with shaded areas indicating 95% confidence intervals. Different sites are represented by points of different shapes. Slopes in Fig. d and f are reported as the mean ± SEM. All statistical tests are conducted as two-sided.

https://doi.org/10.1038/s41559-025-02701-y



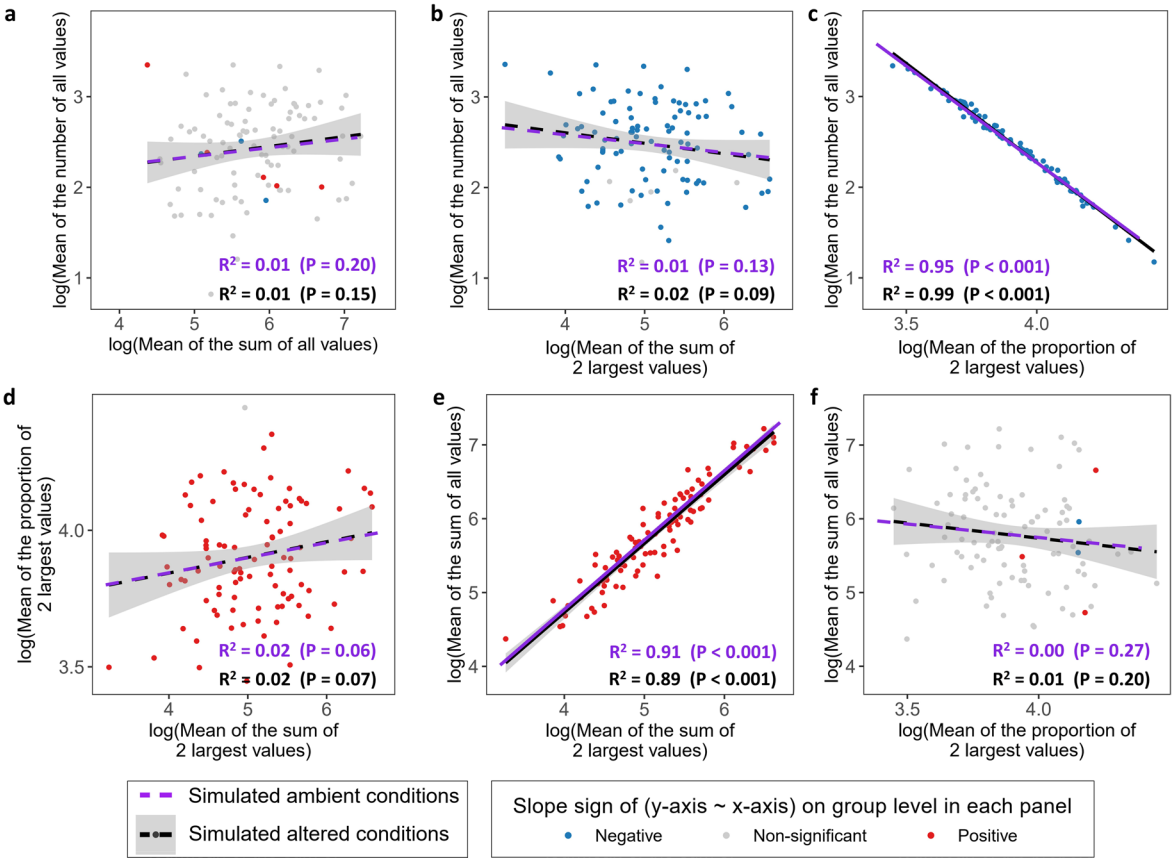
Extended Data Fig. 3 | Site-level relationships in NutNet under (a-f) ambient conditions and (g-l) altered conditions. The relationship between richness and (**a**, **g**) community biomass, and (**b**, **h**) absolute biomass of the two most dominant species, and (**c**, **i**) relative biomass of the two most dominant species; and (**d**, **j**) between relative biomass and absolute biomass of the two most dominant species; and between community level biomass and (**e**, **k**) absolute biomass of the two most dominant species, and (**f**, **l**) relative biomass of the two most dominant species of each site under ambient and altered conditions (76 sites; each site ≈ 3 blocks; each block ≈ 10 plots). All data were natural log-transformed to improve normality. The black dashed and solid lines indicate that the worldwide relationship is not significant (P > 0.05) and significant (P < 0.05), respectively. Both marginal R^2 (R^2m) and conditional R^2 (R^2c) are presented in the figures. All statistical tests are conducted as two-sided.

https://doi.org/10.1038/s41559-025-02701-y



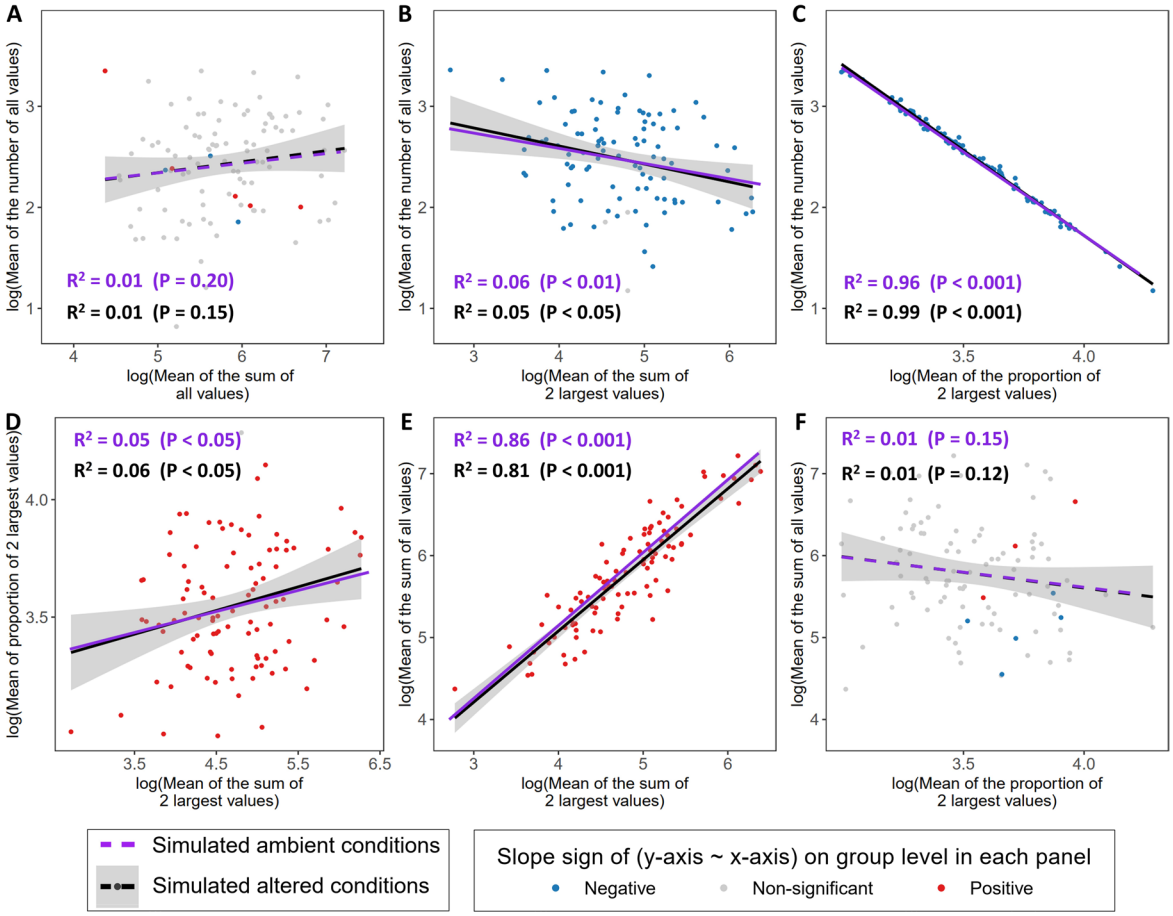
Extended Data Fig. 4 | Worldwide relationships between log(gamma diversity) and log(mean relative biomass of the two most dominant species) across 76 NutNet sites under (a) ambient and (b) altered conditions. All data were natural log-transformed to improve normality. The solid lines indicate that the overall relationship is significant ($P < 0.05$), with shaded areas indicating 95% confidence intervals. All statistical tests are conducted as two-sided.

https://doi.org/10.1038/s41559-025-02701-y



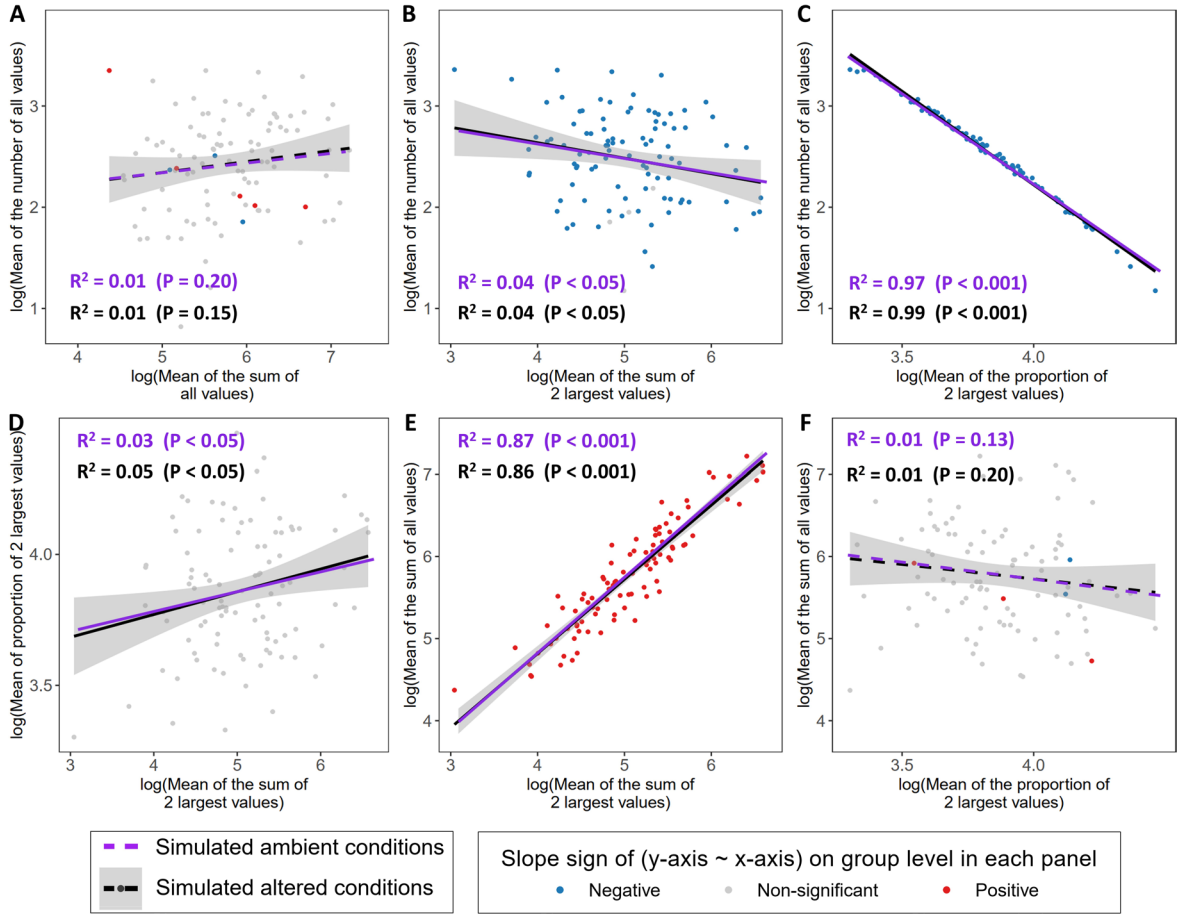
Extended Data Fig. 5 | Overall relationships generated from simulated lognormal distribution data under simulated ambient (purple lines) and altered environmental (dots and black lines) conditions. The relationship between mean of the number of all values and (**a**) mean of the sum of all values, and (**b**) mean of the sum of the two largest values, and (**c**) mean of the proportion of the two largest values; and (**d**) mean of the proportion of the two largest values and mean of the sum of the two largest values; and between mean of the sum of all values and (**e**) mean of the sum of the two largest values, and (**f**) mean of the proportion of the two largest values, of 100 simulated sites under simulated altered environmental conditions (100 simulated sites; simulated year > 0; each simulated site includes 3 simulated blocks; each simulated block includes 10 simulated plots). All data were natural log-transformed to improve normality. The red, gray, and blue dots mean that the relationship between the y-axis and x-axis variables of each panel on the simulated site-level is significantly positive, non-significant correlated, and significantly negative correlated under simulated altered conditions, respectively. The purple lines are regression curves for the ambient conditions. The purple fonts are R2 and P values for the ambient conditions. The black lines are regression curves for the altered environmental conditions. The black fonts are R2 and P values for the altered environmental conditions. The dashed and solid lines indicate that the overall relationship is not significant ($P > 0.05$) and significant ($P < 0.05$), respectively, with shaded areas indicating 95% confidence intervals. All statistical tests are conducted as two-sided.

https://doi.org/10.1038/s41559-025-02701-y



Extended Data Fig. 6 | Overall relationships generated from simulated Fisher's log series distribution data under simulated ambient (purple lines) and altered environmental (dots and black lines) conditions. The relationship between mean of the number of all values and (**a**) mean of the sum of all values, and (**b**) mean of the sum of the two largest values, and (**c**) mean of the proportion of the two largest values; and (**d**) mean of the proportion of the two largest values and mean of the sum of the two largest values; and between mean of the sum of all values and (**e**) mean of the sum of the two largest values, and (**f**) mean of the proportion of the two largest values, of 100 simulated sites under simulated ambient and altered environmental conditions (100 simulated sites; simulated year > 0; each simulated site includes 3 simulated blocks; each simulated block includes 10 simulated plots). All data were natural log-transformed to improve normality. The red, gray, and blue dots mean that the relationship between the y-axis and x-axis variables of each panel on the simulated site-level is significantly positive, non-significant correlated, and significantly negative correlated under simulated altered conditions, respectively. The purple lines are regression curves for the ambient conditions. The purple fonts are R2 and P values for the ambient conditions. The black lines are regression curves for the altered environmental conditions. The black fonts are R2 and P values for the altered environmental conditions. The dashed and solid lines indicate that the overall relationship is not significant ($P > 0.05$) and significant ($P < 0.05$), respectively, with shaded areas indicating 95% confidence intervals. All statistical tests are conducted as two-sided.

https://doi.org/10.1038/s41559-025-02701-y



Extended Data Fig. 7 | Overall relationships generated from simulated multinomial distribution data under simulated ambient (purple lines) and altered environmental (dots and black lines) conditions. The relationship between mean of the number of all values and (**a**) mean of the sum of all values, and (**b**) mean of the sum of the two largest values, and (**c**) mean of the proportion of the two largest values; and (**d**) mean of the proportion of the two largest values and mean of the sum of the two largest values; and between mean of the sum of all values and (**e**) mean of the sum of the two largest values, and (**f**) mean of the proportion of the two largest values, of 100 simulated sites under simulated ambient and altered environmental conditions (100 simulated sites; simulated year > 0; each simulated site includes 3 simulated blocks; each simulated block includes 10 simulated plots). All data were natural log-transformed to improve normality. The red, gray, and blue dots mean that the relationship between the y-axis and x-axis variables of each panel on the simulated site-level is significantly positive, non-significant correlated, and significantly negative correlated under simulated altered conditions, respectively. The purple lines are regression curves for the ambient conditions. The purple fonts are R2 and P values for the ambient conditions. The black lines are regression curves for the altered environmental conditions. The black fonts are R2 and P values for the altered environmental conditions. The dashed and solid lines indicate that the overall relationship is not significant ($P > 0.05$) and significant ($P < 0.05$), respectively, with shaded areas indicating 95% confidence intervals. All statistical tests are conducted as two-sided.

nature portfolio

Corresponding author(s): Pengfei Zhang

Last updated by author(s): Mar 19, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We did not use software to collect our data for this analysis.
Data analysis	We used standard packages in R (v 4.3.1) to analyze our data and generate figures.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All the data that support the findings of this study are available from https://doi.org/10.6073/pasta/89bbbeee6dc3990274fd079bdda92f1e.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

All the sites are located in areas dominated by low statured, primarily herbaceous vegetation including old fields and pastures, tallgrass, mixed and shortgrass prairies, alpine tundra, montane meadow, savannah and shrub-steppe, desert and semi-arid grassland, and annual grasslands, which we refer to collectively as grasslands. These sites encompass a wide range of environmental conditions including elevation (0.5 to 4241 m above sea level), mean annual precipitation (MAP; 203 to 2114 mm per year, Extended Data Fig. 1a), mean annual temperature (MAT; -3.3 to 27.3 ℃, Extended Data Fig. 1a), and latitude (69° N to 52° S, Extended Data Fig. 1b). These sites also include a wide range of site means of community biomass (26.2 to 1122.9 g per m2), local mean (alpha) plant richness (1.7 to 43.1 species per 1 m2 quadrat), and site level (gamma) plant richness (14 to 128 species per site).
At each site, local researchers established an identical experiment which manipulates the supplies of various biologically limiting elements (e.g., nitrogen, phosphorus, potassium, and various micronutrients) and the density of mammalian grazers. The experiment is composed of 10 treatments applied at the scale of 5 m × 5 m plots and replicated within each site in a completely randomized block design, with most sites having three complete blocks (range 1 to 6 blocks per site). Each sampling area was separated by more than 1.5 m from neighbouring plots (1 m walkway and 0.5 m within-plot buffer) to minimize spillover effects of treatments on adjacent plots.
The treatments included a factorial combination of three nutrient treatments (1. Nitrogen (N), 2. Phosphorus (P), and 3. Potassium and Micronutrients (K□μ)), each at two levels (Control, and Nutrient Added) for a total of eight treatments. The last two treatments use fencing to exclude herbivores: Fencing without nutrient added (Fencing) and fencing with all nutrients added (Fencing+NPKμ). In total, the ten different treatments are: Control, N, P, Kμ, NP, NKμ, PKμ, NPKμ, Fencing, and Fencing+NPKμ.
The specific nutrient treatments were as follows: 10 g N m-2 yr-1 as timed-release urea ((NH2)2CO), 10 g P m-2 yr-1 as triple superphosphate (Ca(H2PO4)2), 10 g K m-2 yr-1 as potassium sulphate (K2SO4), and 100 g m-2 yr-1 of a micronutrient mix containing Fe (15%), S (14%), Mg (1.5%), Mn (2.5%), Cu (1%), Zn (1%), B (0.2%) and Mo (0.05%). Micronutrients were only added in the first year to prevent the build up to toxic levels. Fences were 230 cm tall with the lower 90 cm surrounded by 1-cm woven wire mesh. An additional 30-cm outward-facing flange was stapled to the ground to exclude digging animals, though not fully subterranean animals.
The local scientists collect identical data using standardized protocols starting in the year prior to application of experimental treatments (ambient conditions) and in the years following the treatments (experimentally altered conditions). The first set of sites were established in 2007 and include 1 observational year and 15 post-treatment years and the newest sites have 1 observational year and 1 post-treatment year. Taken together, our data includes 21,233 unique site-block-treatment-year combinations, including 18,705 plots representing altered environmental conditions and 2,528 plots under ambient conditions, collected over a span of up to 16 years across 76 sites (Extended Data Fig. 1b).

Research sample

To measure absolute cover of each species, within each 5 m × 5 m plot, a randomly designated 1 m × 1 m quadrat was permanently marked and sampled annually at peak plant biomass. In each quadrat, absolute cover was visually estimated to the nearest 1% for every species overhanging the quadrat. The cover of each species per quadrat was estimated independently, such that the total summed absolute cover of all species (i.e., community-level absolute cover) can exceed 100% in multilayer canopies.
The relative cover of each species was calculated as the ratio of its absolute cover to the community-level absolute cover, ensuring that the sum of the relative cover of all species within each quadrat equalled 100% (Extended Data Fig. 1e).
Species richness per quadrat was determined by the number of species in the cover data.
To measure community biomass, adjacent to the permanent 1 m × 1 m cover quadrat, aboveground live biomass was estimated by clipping all aboveground biomass at ground level within two 1 m × 0.1 m strips (totalling 0.2 m2), with the location of these strips being moved each year. All biomass samples were dried at 60 ℃ to constant mass and then weighed to the nearest 0.01 g. The

	weights were multiplied by five to estimate grams per square meter, representing the community biomass of the 1 m2 quadrat (Extended Data Fig. 1d).
Sampling strategy	Within each 5 m × 5 m plot, a randomly designated 1 m × 1 m quadrat was permanently marked and sampled annually at peak plant biomass. In each quadrat, absolute cover was visually estimated to the nearest 1% for every species overhanging the quadrat. The cover of each species per quadrat was estimated independently, such that the total summed absolute cover of all species (i.e., community-level absolute cover) can exceed 100% in multilayer canopies. Species richness per quadrat was determined by the number of species in the cover data. The relative cover of each species was calculated as the ratio of its absolute cover to the community-level absolute cover, ensuring that the sum of the relative cover of all species within each quadrat equalled 100% (Extended Data Fig. 1e). Adjacent to the permanent 1 m × 1 m cover quadrat, aboveground live biomass was estimated by clipping all aboveground biomass at ground level within two 1 m × 0.1 m strips (totalling 0.2 m2), with the location of these strips being moved each year. All biomass samples were dried at 60 ℃ to constant mass and then weighed to the nearest 0.01 g. The weights were multiplied by five to estimate grams per square meter, representing the community biomass of the 1 m2 quadrat (Extended Data Fig. 1d).
Data collection	At the time of peak biomass each year at each of the 76 sites in this study, total aboveground biomass of all plants rooted within two 0.1 m2 (10 x 100 cm) strips in each experimental plot was clipped using hand shears or clippers and measured on a balance. Clipped vegetation was separated by hand into live and dead components, dried in a drying oven at 60oC ℃for 48 hrs, and weighed to the nearest 0.01 g. For plots with shrubs and subshrubs, all leaves and current year's woody growth were collected, dried and weighed. We also quantified species richness and composition by visually estimating the percent cover of each species to the nearest 1% in a randomly designated, but permanently marked, 1 x 1 m subplot within each 25 m2 plot. All data were recorded each year in a standardized spreadsheet. These data were collected over the course of this 16-year experiment by site PIs, postdocs, graduate students, and other lab personnel associated with each of the 76 sites contributing to this experiment.
Timing and spatial scale	Because we sought to characterize the relationships among community biomass, species richness, absolute abundance of dominant species, and relative abundance of dominant species in global grassland under both ambient and altered environmental conditions, aboveground biomass was collected at the seasonal point of peak growing season biomass at each site in each year (sampling dates vary by site location, including continent and hemisphere) to provide data at a biologically comparable point in the season for sites around the world. At each of the 76 sites in this study, total aboveground biomass of all plants rooted within two 0.1 m2 (10 x 100 cm) strips was collected from each of the 5 x 5 m experimental plots. Start dates of the experimental treatments varied among sites, but the final year of sampling for all sites included in this study was 2023. When we analyzed each year or each treatment separately under altered environmental conditions, the results were consistent with the data averaged across all years or all treatments (Extended Data Table 1a, b).
Data exclusions	NA
Reproducibility	This study, replicated at 76 sites around the world, is designed specifically to examine reproducibility -- the responses that are general across sites and those that are contingent on site-level characteristics. So, yes, the reproducibility of results across site conditions is one of the central goals of this study.
Randomization	This study is a randomized block design within each site. All experimental treatments were randomly assigned to plots at the outset of the study.
Blinding	Plots are numbered, thus investigators do not track the treatments being applied to each plot at the time of data collection, but rather associate data with each treatment via plot number after data collection. So, yes, investigators are blind to the treatments at the time of data collection.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	This was a 16-year-long study at 76 sites spanning 6 continents. We have provided a figure (Extended Data Figure 1a) that includes the mean annual precipitation and mean annual temperature of every site.
Location	We have provided a figure (Extended Data Figure 1b) that includes the latitudes and longitudes of every site.
Access & import/export	Biomass was measured using standard methods in each investigator lab. Absolute cover was estimated to the nearest 1% for every species overhanging the each quadrat.
Disturbance	The experiment included building and maintaining fences to reduce access by large mammals and addition of elemental nutrients. Investigators accessed sites, often by walking on trails. Therefore, the disturbance beyond the treatments, themselves, was minimal.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☒	☐ Demonstrate how to render a vaccine ineffective
☒	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐ Increase transmissibility of a pathogen
☒	☐ Alter the host range of a pathogen
☒	☐ Enable evasion of diagnostic/detection modalities
☒	☐ Enable the weaponization of a biological agent or toxin
☒	☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA