nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-020-1280-9

Check for updates

The results of biodiversity–ecosystem functioning experiments are realistic

Malte Jochum[1,2,3]✉, Markus Fischer[1], Forest Isbell[4], Christiane Roscher[2,5], Fons van der Plas[6], Steffen Boch[1,7], Gerhard Boenisch[8], Nina Buchmann[9], Jane A. Catford[10], Jeannine Cavender-Bares[4], Anne Ebeling[11], Nico Eisenhauer[2,3], Gerd Gleixner[8], Norbert Hölzel[12], Jens Kattge[2,8], Valentin H. Klaus[9], Till Kleinebecker[13], Markus Lange[8], Gaëtane Le Provost[14], Sebastian T. Meyer[15], Rafael Molina-Venegas[1,16], Liesje Mommer[17], Yvonne Oelmann[18], Caterina Penone[1], Daniel Prati[1], Peter B. Reich[19,20], Abiel Rindisbacher[1], Deborah Schäfer[1], Stefan Scheu[21,22], Bernhard Schmid[23,24], David Tilman[4,25], Teja Tscharntke[26], Anja Vogel[2,3,11], Cameron Wagg[27], Alexandra Weigelt[2,6], Wolfgang W. Weisser[15], Wolfgang Wilcke[28] and Peter Manning[14]

A large body of research shows that biodiversity loss can reduce ecosystem functioning. However, much of the evidence for this relationship is drawn from biodiversity–ecosystem functioning experiments in which biodiversity loss is simulated by randomly assembling communities of varying species diversity, and ecosystem functions are measured. This random assembly has led some ecologists to question the relevance of biodiversity experiments to real-world ecosystems, where community assembly or disassembly may be non-random and influenced by external drivers, such as climate, soil conditions or land use. Here, we compare data from real-world grassland plant communities with data from two of the largest and longest-running grassland biodiversity experiments (the Jena Experiment in Germany and BioDIV in the United States) in terms of their taxonomic, functional and phylogenetic diversity and functional-trait composition. We found that plant communities of biodiversity experiments cover almost all of the multivariate variation of the real-world communities, while also containing community types that are not currently observed in the real world. Moreover, they have greater variance in their compositional features than their real-world counterparts. We then re-analysed a subset of experimental data that included only ecologically realistic communities (that is, those comparable to real-world communities). For 10 out of 12 biodiversity–ecosystem functioning relationships, biodiversity effects did not differ significantly between the full dataset of biodiversity experiments and the ecologically realistic subset of experimental communities. Although we do not provide direct evidence for strong or consistent biodiversity–ecosystem functioning relationships in real-world communities, our results demonstrate that the results of biodiversity experiments are largely insensitive to the exclusion of unrealistic communities and that the conclusions drawn from biodiversity experiments are generally robust.

Concerns over the consequences of biodiversity loss for human well-being triggered the growth of biodiversity–ecosystem functioning (BEF) research, an important field of ecology over the past 25 years[1–8]. Some of the most influential studies in this field are based on BEF experiments (hereafter, biodiversity experiments), in which communities of varying diversities

[1]Institute of Plant Sciences, University of Bern, Bern, Switzerland. [2]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [3]Institute of Biology, Leipzig University, Leipzig, Germany. [4]Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, MN, USA. [5]Department of Physiological Diversity, UFZ, Helmholtz Centre for Environmental Research, Leipzig, Germany. [6]Systematic Botany and Functional Biodiversity, Institute of Biology, Leipzig University, Leipzig, Germany. [7]WSL Swiss Federal Research Institute, Birmensdorf, Switzerland. [8]Max Planck Institute for Biogeochemistry, Jena, Germany. [9]Institute of Agricultural Sciences, ETH Zürich, Zürich, Switzerland. [10]Department of Geography, King's College London, London, UK. [11]Institute of Ecology and Evolution, University of Jena, Jena, Germany. [12]Institute of Landscape Ecology, University of Münster, Münster, Germany. [13]Institute of Landscape Ecology and Resources Management, Giessen University, Giessen, Germany. [14]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt, Germany. [15]Terrestrial Ecology Research Group, School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany. [16]GloCEE–Global Change Ecology & Evolution group, Department of Life Sciences, University of Alcalá, Madrid, Spain. [17]Plant Ecology and Nature Conservation Group, Wageningen University and Research, Wageningen, The Netherlands. [18]Geoecology, Geoscience Department, University of Tübingen, Tübingen, Germany. [19]Department of Forest Resources, University of Minnesota, Saint Paul, MN, USA. [20]Hawkesbury Institute for the Environment, Western Sydney University, Penrith South DC, New South Wales, Australia. [21]JF Blumenbach Institute of Zoology and Anthropology, University of Göttingen, Göttingen, Germany. [22]Centre for Biodiversity and Sustainable Land Use, University of Göttingen, Göttingen, Germany. [23]Department of Geography, University of Zürich, Zürich, Switzerland. [24]Institute of Ecology, College of Urban and Environmental Sciences, Peking University, Beijing, China. [25]Bren School of Environmental Science and Management, University of California, Santa Barbara, CA, USA. [26]Agroecology, Department of Crop Sciences, University of Göttingen, Göttingen, Germany. [27]Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland. [28]Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. ✉e-mail: malte.jochum@idiv.de

are randomly assembled and the responses of ecosystem processes are measured[9,10]. These experiments, often conducted using grassland communities[10–14], aim to isolate the effects of species richness from those of other factors known to affect ecosystem processes, such as climate, nutrient availability and the presence of particular plant functional types. By doing so, they have provided strong evidence that biodiversity can affect the functioning of ecosystems—most commonly with a positive but saturating relationship between diversity and plant productivity[1,2,7,11,12,15,16]. However, the relevance of biodiversity experiments to real-world ecosystems (that is, those where community assembly is influenced by external drivers, such as climate, soil conditions or land use) has been repeatedly questioned[17–26]. Criticisms highlight several common features of experimental designs—namely, random assembly (as opposed to non-random assembly or disassembly of real-world ecosystems[20]), initial sowing of even species abundances (but see refs. [27–30]) and the repeated removal of non-target species (but see refs. [31,32]). These factors may alter community assembly processes, leading to unrealistic communities that possess functional properties that are rare or absent in the real world. Although numerous researchers have argued for the relevance of biodiversity experiments[22,23,33,34] and provided evidence to counter these criticisms[28,35,36], we do not know how closely plant communities in biodiversity experiments resemble those of related real-world ecosystems (but see ref. [37] for a local-scale comparison), or whether the presence of unrealistic communities affects the conclusions drawn from these experiments.

To close these knowledge gaps, we take a two-step approach. First, we perform a comprehensive, quantitative assessment of the differences and similarities between plant communities from biodiversity experiments and related real-world ecosystems. Second, we test the robustness of the conclusions drawn from biodiversity experiments to the removal of unrealistic communities—those least comparable to real-world communities. In the first step, we quantitatively compared the plant communities of two of the world's largest and longest-running grassland biodiversity experiments with those of nearby real-world communities where diversity gradients are created by natural environmental variation and global-change drivers. These experiments are the Jena Experiment, established in 2002 in Jena, Germany (hereafter, the Jena Experiment)[10,32], and the BioDIV experiment, established in 1994 at the Cedar Creek Ecosystem Science Reserve, Minnesota, United States (hereafter, BioDIV)[11,38–40] (Fig. 1). We compared experimental communities from the Jena Experiment with those of agricultural grasslands in three regions of Germany, spanning a broad range of site conditions and land-use intensities (the Biodiversity Exploratories[41,42]) and grasslands close to the Jena Experiment (hereafter, Jena real world). BioDIV's experimental communities were compared with nearby, naturally assembled prairie–grassland communities at Cedar Creek, including fertilized grasslands[35,43,44] and those undergoing successional change[45] (Methods and Extended Data Fig. 1). We combined species-specific cover data from annual vegetation surveys (3,329 and 9,954 plot-year combinations in the German and the US datasets, respectively) with phylogenetic information and plant functional-trait data to characterize and quantitatively compare plant communities on the basis of a range of properties known to represent important dimensions of biodiversity and to independently influence ecosystem functioning[46–49]. These properties included measures of taxonomic diversity and evenness, phylogenetic diversity, functional diversity and community abundance-weighted means (CWMs) of selected functional traits of vascular plants (hereafter referred to as community properties) and were examined in a principal component analysis (PCA) (see Methods for definitions of all community properties; Fig. 1). On the basis of this multidimensional, multivariate comparison of plant-community properties, we identified plots from biodiversity experiments whose communities fell outside the multidimensional community-property space occupied by real-world plant communities (hereafter, unrealistic communities). This was achieved by calculating the intersection of three-dimensional convex hull volumes defined by experimental and real-world communities (Fig. 1 and Methods). In the second step of our analysis, we fit linear models to test how plant species richness affected eight selected ecosystem functions from both the above- and belowground subsystems. This was done for both the full datasets and the subsets of realistic plots.

Results and discussion

Plant communities in biodiversity experiments and related real-world systems. The results of our multidimensional, multivariate comparison showed that experimental plant communities occupy a larger area of multivariate community-property space than real-world communities, despite the latter covering a wide range of climatic, edaphic and management conditions, particularly in the German dataset[41,50] (Fig. 1a,e). This finding was robust to the inclusion or exclusion of particular community properties and the choice of overlap calculation methodology (Supplementary Information on Sensitivity Analyses 1, Supplementary Fig. 1 and Supplementary Tables 1–3) and was supported by additional data collected at Jena. These data showed that experimental communities migrated towards the narrow space occupied by real-world communities when not weeded (that is, Jena invasion; see Extended Data Fig. 2), thus also indicating that the differences between real-world and biodiversity-experiment communities in multivariate community-property space were due to experimental maintenance rather than differences in plot conditions, species pools or initially random versus natural community assembly.

Next, for each community property in each region (Germany and the United States), we determined the proportion of biodiversity-experiment plots that fell within the community-property range of the related real-world plots (the violin plots[51] in Extended Data Figs. 3 and 4, and Supplementary Tables 4 and 5). Specifically, in Germany, Simpson's evenness (SEve), species richness, Faith's phylogenetic diversity (PD), functional richness (FRic) and mean nearest taxon distance (MNTD) showed the lowest proportion of biodiversity-experiment plots in the real-world range of these properties. Experimental communities at Jena showed higher values of SEve and MNTD and lower species richness, PD and FRic than their real-world counterparts. In contrast, in the US dataset, CWM value of leaf dry matter content (LDMC), functional evenness (FEve), CWM value of specific leaf area (SLA), CWM value of leaf nitrogen (leaf N) and FRic showed the lowest proportion of experimental plots in the real-world range of community properties, and all these community properties showed lower values in the experimental than in the real-world communities.

Overall, three conclusions can be drawn from this comparative analysis. First, biodiversity experiments successfully create plant communities that vary greatly in functionally important community properties. Second, real-world communities are confined to narrower regions of multivariate community-property space than those of experiments. Third, while the properties of many experimental communities are not observed in related real-world communities, a subset of randomly assembled experimental communities are functionally comparable to real-world communities (Fig. 1 and Supplementary Tables 4 and 5), even though their taxonomic community composition may differ (see Supplementary Information on Sensitivity Analyses 1, section E, and Supplementary Fig. 2).

The comparative analysis was used to define which plant communities from biodiversity experiments could be deemed comparable to real-world systems (that is, realistic). This analysis revealed that, when using 12 community properties selected using variance inflation factors (vif) to reduce redundant information (Methods), 28% and 77% of the experimental plots were deemed realistic in Jena and BioDIV, respectively (Supplementary Tables 2 and 3).

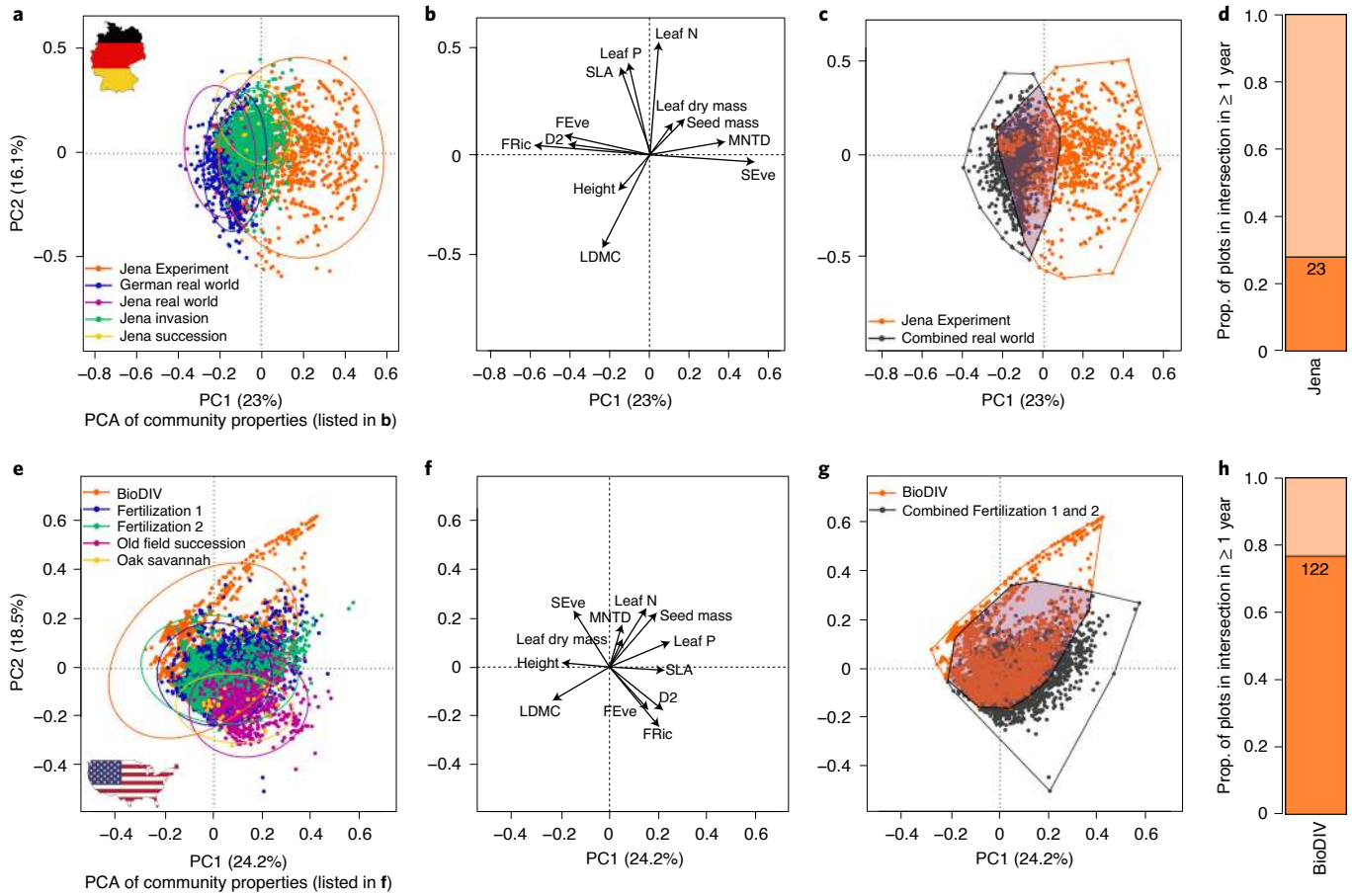


Fig. 1 | Experimental versus real-world communities. a–d, German comparison ($n = 3{,}329$ plot-year combinations). **e–h**, US comparison ($n = 9{,}954$ plot-year combinations). Panels **a–c** and **e–g** show the first two axes of a PCA on 12 plant-community properties (see **b** and **f**, variance-inflation factor-selected CWM traits, functional diversity, phylogenetic diversity and taxonomic diversity metrics). Each dot represents a single plot in a single year. In **a**,**e**, the distribution of the experimental (orange) and various real-world plots with 95% confidence ellipses (variables scaled for PCA) are shown for each subset. In **b**,**f**, the PCA factor loadings for the community properties are shown (arrows have been proportionally increased in length to improve visibility—"const = 25" in R vegan biplot function[87]; see Extended Data Fig. 7 for the PCA factor loadings and ref. [101] for the full dataset). In **c**,**g**, two-dimensional representations of the three-dimensional convex hull volumes for experimental (orange) and combined real-world communities (German real-world and Jena real-world plots for the German comparison, Fertilization 1 and 2 plots for the US comparison, grey) and their intersections (shaded areas) are shown. In **d**,**h**, the number and proportion of biodiversity experiment plots in the intersection (that is, realistic plots; strong colour) are shown. Each plot with at least one annual community in the intersection is defined as realistic. The number of years of vegetation data for each project is as follows: Jena Experiment (13), German real world (8), Jena real world (1), Jena invasion (13), Jena succession (7), BioDIV (19), Fertilization 1 (23), Fertilization 2 (10), old field succession chronosequence (7) and oak savannah (1). Taxonomic diversity indices: inverse Simpson's diversity index (D2) and Simpson's evenness (SEve); phylogenetic diversity indices: mean nearest taxon distance (MNTD); functional diversity indices: functional richness (FRic), and functional evenness (FEve); CWM values of leaf nitrogen (Leaf_N) and phosphorus (Leaf_P), specific leaf area (SLA), leaf dry mass, leaf dry matter content (LDMC), seed mass and plant height. For the definitions of these properties, please see Methods. Country outlines in symbols were created with R package mapdata[108].

The plant communities of these realistic biodiversity-experiment plots had significantly higher sown diversity (in Jena, average = 21.7 realistic versus 3.5 unrealistic; in BioDIV, 7.8 versus 1.7) and more sown functional groups (in Jena, 2.8 versus 1.9; in BioDIV, 3.5 versus 1.5), but lower SEve (in Jena, 0.5 versus 0.7; in BioDIV, 0.6 versus 0.9; Fig. 1) than the unrealistic experimental plots (Fig. 1, Extended Data Figs. 3 and 4, and Supplementary Tables 6 and 7). Although the constraining was not based on species richness, the diversity gradient in Jena was truncated in the realistic subset of plots. In Jena, the average minimum species richness across years was 1 in the unconstrained dataset (all plots) and 3.7 in the constrained dataset (realistic plots only). In contrast, BioDIV covered a relatively narrow range of species richness, and the equivalent real-world communities were also relatively species poor, so here the gradient was not truncated (Fig. 2 and Supplementary Table 9). As such, the low-diversity plots in the Jena Experiment, although necessary for an experimental design that can identify diversity effects and their underlying mechanisms[52], are generally found to be unrealistic when compared with current German real-world communities. Note that study-specific differences in vegetation survey areas could not be avoided, although their impact on the results was minimized (see Methods and Supplementary Figs. 7 and 8 for more detail).

The selection of realistic experimental plots was largely insensitive to most methodological choices, such as the exclusion of certain community properties and the overlap calculation method used (see Supplementary Information on Sensitivity Analyses 1 for the details). For example, using all 21 instead of only the 12 vif-selected community properties resulted in slight changes in the number and identity of plots selected as realistic (91–96% of the main analysis plots included for Jena, 85–95% for BioDIV; Supplementary Tables 3

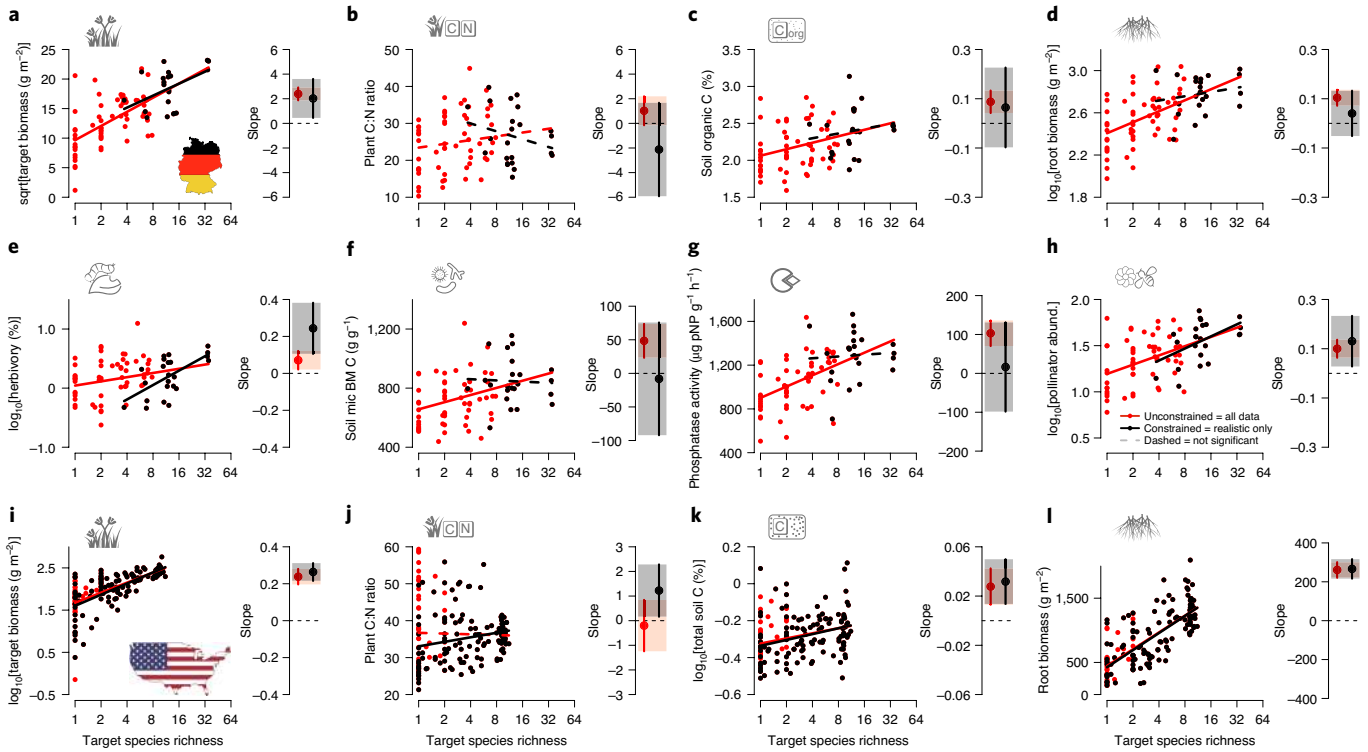


Fig. 2 | BEF relationships. a–l, Relationships between realized target plant species richness (averaged per plot between 2006 and 2015; all *x* axes are on a $\log_2$ scale) and various ecosystem functions in the German (**a–h**, Jena Experiment) and US (**i–l**, BioDIV) biodiversity experiments containing all plots (unconstrained, all dots and red lines) and only realistic plots (constrained, black dots and lines). The insets show slope estimates with 95% confidence intervals (error bars and shaded areas) for all plots (unconstrained, red) and only realistic plots (constrained, black). The sample sizes are $n = 82$ for **a–c** (red), $n = 80$ for **d–g** (red), $n = 79$ for **h** (red), $n = 23$ for **a–h** (black), $n = 159$ for **i** (red), $n = 158$ for **j**,**k** (red), $n = 150$ for **l** (red), $n = 122$ for **i–k** (black) and $n = 117$ for **l** (black). For details on the model parameters (such as sample sizes, slope estimates, confidence intervals, *P* values and adjusted R^2 values), see Extended Data Fig. 5. The dashed regression lines show non-significant relationships ($P > 0.05$). Note that **a–d** and **i–l** show the same ecosystem functions for both experiments (in BioDIV, total soil C represents soil organic C in **k**). BM, biomass; pNP, p-nitrophenylphosphate. Where indicated in the *y*-axis label, the data were transformed to meet the model assumptions. The response variables were averaged over all available years. The icons depict ecosystem functions (as specified in the *y*-axis labels) and were modified from originals by Hamish (flowers (**a**,**b**,**i**,**j**)), Saeful Muslim (bee (**h**), caterpillar (**e**), leaf (**e**)), Alice Noir (microbe (**f**)), Lluis Pareras (coral (**f**)), Creative Stall (bacteria (**f**)), Atif Arshad (pacman (**g**)), Made (flower (**h**)) and amantaka (root (**d**,**l**)), from thenounproject.com. Country outlines in symbols were created with R package mapdata[108].

and 4). However, the selection of realistic plots was sensitive to some methodological choices. Within our sensitivity analyses, the results were relatively sensitive to the following: changing the number of PCA axes used to compute multidimensional overlap, altering the criterion for defining inclusion/exclusion in the overlap, basing our comparison on species abundances rather than community properties and reducing the real-world data to include only those plots with comparable land use to the experiment (for details, see sections B, D, E and F of Supplementary Information on Sensitivity Analyses 1, Supplementary Tables 2 and 3, and Supplementary Figs. 1 and 3). For example, when using species-abundance-based non-metric multidimensional scaling (NMDS) to define realistic communities (Supplementary Fig. 2), in the German dataset, very few experimental plots (2%) fell within the real-world NMDS realm and were selected as realistic. In the US dataset, 33% of plots were selected as realistic. For BEF relationships based on these alternative analyses, see below. As such, as long as the overall analysis framework of using plant-community properties in PCAs to determine multidimensional overlap is used, as opposed to species-abundance-based NMDS, our conclusions are robust to the methodological decisions taken.

BEF relationships in unconstrained versus constrained experimental data subsets. Our comparison of BEF relationships in full datasets of biodiversity experiments (unconstrained, all plots) versus realistic subsets of plots (constrained, realistic plots only) was conducted for the following ecosystem functions: plant aboveground and belowground (root) biomass, plant aboveground carbon-to-nitrogen (C:N) ratio, soil organic carbon content, invertebrate leaf herbivory, soil microbial biomass C, phosphatase activity in the soil and pollinator abundance (Fig. 2). This comparison showed that, in both experiments and across the different ecosystem functions, the slopes of the experimental BEF relationships were relatively insensitive to the removal of unrealistic communities (but see the discussion of significance changes below). A paired *t*-test on pairs of unconstrained and constrained slopes for the 12 BEF relationships shown in Fig. 2 showed no significant change in the slope estimates ($t = 1.40$, d.f. $= 11$, $P = 0.19$, $n = 12$), and the confidence intervals for the slope estimates overlapped each other's mean for all but two model pairs. The two exceptions were both initially weak BEF relationships: Jena Experiment herbivory, where the positive slope increased when constrained to realistic plots, and BioDIV plant C:N, where a non-significant, slightly negative slope turned into a positive significant one (Extended Data Fig. 5). The finding that the slope of the BEF relationship was largely unaffected by the exclusion of unrealistic communities was robust to changing the set of community properties in the PCA and the method used to identify realistic communities (Supplementary Fig. 3). The goodness of fit (adjusted R^2 values) was also only partly affected by constrain-

ing the dataset (mean R^2, 0.24 versus 0.15 for unconstrained and constrained models, respectively; Extended Data Fig. 5), and the average percentage change in maximum functioning was ±10.3% (s.e.m., 4%; Supplementary Table 8). When using the realistic plots defined based on all 21 instead of the 12 vif-selected community properties in the PCA, the BEF-slope changes from unconstrained to constrained data subsets were largely unaltered (Supplementary Information on Sensitivity Analyses 1 and Supplementary Fig. 3). For BioDIV, when using species-abundance-based NMDS to define the overlap, the constrained BEF relationships were comparable to or more strongly positive than the unconstrained relationships (Supplementary Fig. 3). Together, these results show that the form, strength and magnitude of the relationship between biodiversity and functioning that has been identified in biodiversity experiments weakens somewhat, but is generally robust to the removal of unrealistic communities.

In 4 out of 12 cases, constraining the data led to a change from a significant to a non-significant relationship (Jena soil organic C content, root biomass, soil microbial biomass C and phosphatase activity; Fig. 2). To check whether this change in significance was driven by the smaller sample size of the constrained dataset, we assessed the sensitivity of the results to reduced replication. This was done by performing a sensitivity analysis in which we randomly reduced the size of the unconstrained dataset of the Jena Experiment. This analysis showed that the slope of the BEF relationship in the realistic subset for these four relationships was shallower than most slopes estimated from randomly selected data (Supplementary Information on Sensitivity Analyses 2 and Supplementary Fig. 4). This suggests that for certain ecosystem functions (particularly soil processes in the Jena Experiment), the strength of the BEF relationship might be overestimated in biodiversity experiments.

The truncated species-richness gradient of the realistic plots at Jena was associated with a 31% reduction in the range of functioning covered across the truncated reduced biodiversity gradient (Supplementary Table 9). Therefore, to investigate whether the shallower slope and loss of significance in realistic data subsets at Jena was driven by the truncation of the species-richness gradient, we performed an additional sensitivity analysis for the four Jena soil functions in question (Supplementary Information on Sensitivity Analyses 2 and Supplementary Fig. 5). When we restricted the random choice of Jena Experiment plots to the shorter gradient of species richness covered by the realistic plots in the main analysis, the vast majority of BEF relationships in the sensitivity analysis turned non-significant (between 84 and 100 of 100 repetitions; see Supplementary Fig. 5). This indicates that it is primarily the shortened species-richness gradient, rather than reduced sample size, that drives the weakening of some BEF relationships when constrained (Supplementary Fig. 5). These results show that removing the lower end of the species-richness gradient leaves only the saturating, right-hand side of the commonly observed BEF relationship[1] in some constrained experimental datasets, for which the slope is shallower. These shallower slopes do not demonstrate that experiments falsely predict a stronger BEF relationship at low richness, but they do indicate that some real-world systems do not vary over the full richness gradient found in experiments, thus potentially explaining the relatively weaker BEF relationships observed across real-world diversity gradients, compared with experiments[4].

Finally, several observational real-world studies have shown that other aspects of biodiversity (such as functional composition) are stronger predictors of ecosystem functioning than species richness[4,53,54], while experiments show a dominant effect of species richness and related variables[47,55]. We therefore investigated whether the identity of the community properties that best explain function was affected by our constraining to realistic experimental plots. This analysis demonstrated that the relative importance of plant-community properties in explaining experimental ecosystem functioning changed slightly due to the constraining to realistic experimental plots: for Jena aboveground biomass, PD and CWM SLA gained importance and FEve and MNTD lost importance; and for BioDIV soil organic C, SEve slightly gained importance and CWM seed mass lost importance (see Supplementary Table 10 for the details). However, there was no large, systematic shift in the identity of the plant-community properties that best explain ecosystem functioning.

Our results show that the BEF relationships observed in biodiversity experiments are not an experimental artefact caused by the presence of unrealistic communities. The question remains, however, as to how important biodiversity is as a driver of ecosystem functioning in the real world, relative to factors such as land use or climate[12,21,56]. Although strong and positive BEF relationships have been reported in real-world studies[4,9,36,57–59], other studies describe weak or negative relationships[4,60,61]. This inconsistency (and the discrepancy between experimental and real-world patterns) is commonly attributed to the presence of covarying environmental or biological factors that also drive ecosystem functioning[62] and that obscure, confound or negate the effects of biodiversity (such as nutrient availability, climate and the dominant functional traits of the community[53,57,63–65]). These factors are likely to be closely coupled in real-world ecosystems but decoupled in experiments. Indeed, across our datasets, the average correlation strength of the eight measures of dominant functional traits (CWMs) with SEve and functional and phylogenetic diversity properties was slightly higher in real-world than in experimental data subsets (the mean absolute correlation coefficients were 0.18 and 0.22 in German and US real-world plots, compared with 0.08 and 0.16 in their respective experiments; Supplementary Tables 11 and 12).

While it would be desirable to directly compare the experimental BEF relationships described in this study with those observed in real-world systems, both theoretical and empirical studies show that simple, bivariate relationships between species richness and functioning will not necessarily be positive, even if there are strong underlying effects of biodiversity on ecosystem functioning[57,63]. Previous investigations have shown neutral or negative relationships between plant species richness and biomass for the German real-world dataset included in our study[66]. Furthermore, the relationship between species richness and a production-only ecosystem-service scenario (heavily based on plant shoot biomass) was negative, even when accounting for land-use intensity in a structural equation modelling framework[53]. This negative relationship may be driven by extremely strong covariation between species richness and functional composition (the Pearson correlation between species richness and CWM SLA is as strong as $r = -0.9$ in one region), making it virtually impossible to distinguish between the effects of diversity and those of functional composition using conventional methods. For the fertilization studies at Cedar Creek, negative relationships between diversity and productivity across space were observed because fertilized plots possess high productivity and low diversity, but when fertilization reduced plant species richness, this also reduced productivity over time[35]. Consequently, adequately investigating real-world BEF relationships requires specific, in-depth knowledge of the identity and interplay of additional drivers of both species richness and ecosystem functions[57,63] and analysis frameworks capable of disentangling covariation in and simultaneous reciprocal effects between these interrelated drivers.

While the biodiversity experiments used in our analysis cover a wide range of plant-community properties, only a fraction of this multidimensional space is occupied by related real-world communities. The remainder of the space covered by the experimental communities is currently not observed in the real-world communities that we considered; however, this unrealized plant-community property space may be useful in predicting ecosystem functioning

in the future, when novel combinations of species and environmental conditions may emerge[33,67].

Conclusions

We show that, although biodiversity experiments deliberately include plant communities that may not currently occur under real-world conditions, the BEF relationship is generally robust to the exclusion of these communities. Sensitivity analyses suggest that, where BEF relationships did become weaker and non-significant, this change was primarily driven by the truncated species-richness gradient in the realistic subset of experimental plots. This indicates that experiments do not overestimate possible BEF relationships, but rather that some real-world biodiversity gradients may not currently span the gradient in which biodiversity loss has its strongest impact. Nevertheless, it is conceivable that future changes to biodiversity may occur over this low to very low range.

Although we do not provide direct evidence for strong BEF relationships in real-world communities, our results complement previous reports of significant BEF relationships in the real world[4,36,42,58,59,64] by showing that constraining experimental datasets to contain only real-world-comparable plant communities does not change the core conclusions of BEF research. However, to advance this field, we must acknowledge both the strengths and the limitations of biodiversity experiments. Specifically, our improved understanding should be used to develop a new generation of experiments—for example, experiments that focus on more realistic patterns of community change[68]. At the same time, we must maintain and further examine the valuable resource of long-term biodiversity experiments, such as by re-analysing existing experimental data to simulate a range of possible biodiversity-change scenarios. By moving beyond critiques of experimental design and placing experimental BEF research in the context of natural communities, we advance the current debate from verbal arguments to a quantitative investigation, thus increasing the robustness and applicability of BEF research.

Methods

Overview and data origin. We chose two of the largest and longest-running grassland biodiversity experiments in the world for our comparison. The Jena Experiment[10,32] was chosen as a Central European example of a long-term, intensively studied biodiversity experiment[32,69]. In the Jena main experiment, combinations of 1, 2, 4, 8, 16 and 60 species from a pool of 60 Arrhenatherion grassland species[70] were sown in 82 originally 20 m × 20 m plots on a former agricultural field in 2002. This species richness gradient was crossed with a gradient of functional group richness (one to four functional groups; small herbs, tall herbs, grasses and legumes), where species were randomly chosen from the respective functional groups[10]. Jena Experiment plots are maintained by weeding (two or three times per year). All plots are mown twice per year, and mown biomass is removed (a common management practice of meadows in the region) and do not receive any fertilizers. The Jena Experiment includes two invasion subexperiments, which are nested within the main experiment plots as subplots. One set of these Jena invasion plots was not weeded after initial sowing and was studied regularly until 2009; another set was weeded initially, but weeding halted in 2010 (ref. [32]). Here, we use the former for 2003–2009 and the latter for 2010–2015. Jena mown succession plots were not initially sown and are excluded from all management except for the mowing. These plots represent intermediate successional stages between the biodiversity experiment and the real-world systems, so they were included in the multivariate analysis of community-property overlap (Fig. 1). However, given that they are influenced by initial sowing, and that vegetation surveys were performed using different methodology (see below), they were not considered real-world counterparts when constraining the Jena Experiment to realistic plots (see below).

As a real-world counterpart to the Jena Experiment, we chose the grassland plots of the Biodiversity Exploratories project (German real world). This large-scale, long-term research project was established in 2006 to assess the effects of land-use intensity on biodiversity and ecosystem functioning in three regions of Germany[41]. The 150 grassland plots measure 50 m × 50 m and were selected to cover a wide and representative range of land-use intensities, here composed of varying levels of mowing frequency, grazing intensity and fertilization[71]. Species richness in the Exploratories grasslands ranges from 9 to 70 species within a 4 m × 4 m subplot, across all years used in our study (see Supplementary Fig. 6 for details on land-use intensity in the Biodiversity Exploratories plots and its impact on the comparability of experimental and real-world communities). The Exploratories data were augmented by the inclusion of data from 14 grasslands in the Saale river valley near the Jena Experiment (unpublished data; Jena real world). These grasslands are usually mown twice per year; most are unfertilized, and some are moderately fertilized.

The Cedar Creek biodiversity experiment e120 (BioDIV; refs. [11,16,38,72]) was selected as a North American example of a long-term biodiversity experiment, while a suite of other naturally assembled grasslands at Cedar Creek served as nearby real-world communities. BioDIV was established in 1994, when 1, 2, 4, 8 or 16 species were randomly drawn from an 18-species pool and sown across 168 13 m × 13 m plots at the Cedar Creek Ecosystem Science Reserve in Minnesota, United States.

Several datasets of local experiments and observation plots served as local real-world comparisons for BioDIV. Experiments e001 (Fertilization 1) and e002 (Fertilization 2) were set up in 1982 to study the long-term effects of fertilization with nitrogen and other nutrients, ranging from low rates of nutrient inputs that are similar to atmospheric N deposition rates to high rates of fertilization similar to those used in agriculture. These experiments consist of 324 plots located across three successional grassland fields (324 plots = 2 fertilization experiments × 3 old fields × 9 fertilization treatments × 6 replicates) that differ in their age since abandonment from agriculture and 45 plots in one never-ploughed oak savannah in Fertilization 1 (45 plots = 9 nutrient treatments × 5 replicates)[43]. The plot sizes were 4 m × 4 m in the younger fields and 2 m × 4 m in the oak savannah. In contrast to Fertilization 1, the Fertilization 2 plots were agriculturally disked before receiving nutrient addition treatments. The plot-level species richness in the two fertilization studies ranged from 1 to 28 species across all years used in our study. Established in 1983 and 1989, the Cedar Creek project e014 (old field succession chronosequence) offers vegetation data from four to six observational transects in each of 23 different fields repeated seven times between 1983 and 2011 to study succession after agricultural abandonment[45]. Cedar Creek project e093 (oak savannah), established in 1991, offers data from 30 2 m × 2 m prairie opening plots of natural vegetation[73,74]. This combination of Cedar Creek datasets was chosen to represent a variety of real-world plant communities that were comparable to the BioDIV experiment. Note that while Central European grasslands depend on anthropogenic management (mowing and grazing) to prevent succession to forest, the US prairies are naturally fire disturbed—hence the selection of agricultural plots as the German real-world grassland. Please note that while all above-described datasets were used to illustrate multivariate overlap in plant-community properties (Fig. 1a,b,e,f), only a subset was used to constrain the biodiversity experiment data to realistic plots, as different vegetation-survey techniques in the old field succession chronosequence and the oak savannah datasets (transects and subplots) made these data relatively incomparable (Fig. 1c,d,g,h; see below). For an overview of the datasets used in this study and online resources to obtain the original data, see Extended Data Fig. 1.

Plant-community properties. *Vascular plant cover and biomass.* In the Jena Experiment, vegetation surveys were performed annually in the second half of May on a 3 m × 3 m subplot of each plot, and species-specific cover data were collected. Note that, in the Jena main plots, only target species (vascular plants originally sown in the respective plots) were recorded. Vegetation surveys of the invasion and succession plots were performed annually in 2 m × 2.25 m subplots (2003–2009) or 3 m × 3 m subplots (2010–2015), assessing all present species. We used Jena vegetation data from 2003–2015 (succession data only from 2003–2009). In the Biodiversity Exploratories (German real-world plots), species-specific vascular plant cover was estimated annually on a 4 m × 4 m subplot of each plot between mid-May and mid-June. Here, we used all data from 2008–2015. Data from the 3 m × 3 m vegetation surveys of the Jena real-world plots were available for May 2011.

To test whether the different vegetation survey areas in Jena and the Biodiversity Exploratories might bias the relative abundances of vascular plant species and thus the calculation of abundance-weighted community properties, a separate survey of 27 Biodiversity Exploratories plots (which covered a strong land-use intensity gradient) was performed by sampling species-specific cover in a series of nested 4 m × 4 m (16 m^2, comparable to the Exploratories vegetation survey area), 3 m × 3 m (9 m^2, comparable to the Jena Experiment and Jena real world plots) and 2 m × 2 m (4 m^2, similar to the Jena invasion and succession plots) subplots. As the cover estimates did not show any sign of systematic variation (Supplementary Figs. 7 and 8), we concluded that the different survey areas were unlikely to bias our analysis for the relative-abundance-weighted community properties. We also compared species richness for the 27 16 m^2 and 9 m^2 subplots using a paired *t*-test. This showed a significantly lower species richness in the smaller subplots ($t = 7.30$, d.f. = 26, $P < 0.001$, $n = 27$). On average, the 9 m^2 subplots had only 89% of the species richness of the 16 m^2 subplots. Downscaling species-richness-related community properties on the basis of such a coarse relationship, established for only a subset of plots in only one year, seemed inappropriate. However, the data show that our results should be robust to the differing vegetation survey areas of the datasets included in our study, as species richness and most other taxonomic-diversity community properties (except for D2 and SEve) were removed from the multidimensional comparison (PCA approach) on the basis of the assessment of vif (see below).

For BioDIV, a combination of species-specific cover data (1996–2000) and species-specific aboveground peak biomass (2001–2015) data was used to calculate plant community relative abundance. Previous analyses have shown that this difference in methodology does not affect the conclusions of analyses investigating species-richness effects on biomass[75]. Cover estimates for BioDIV were obtained by averaging the estimates from four permanently marked subplots (each 0.5 m × 1 m) within each plot. Species-specific biomass in BioDIV was obtained by annually clipping 0.1 m × 6 m strips on each plot, drying the clippings and sorting the resulting biomass to species.

For Fertilization 1 and Fertilization 2, species-specific plant aboveground biomass data were collected annually at peak biomass by clipping a 0.1 m × 3 m strip of vegetation per plot, sorting it and drying it. The years 1982–2004 were used for Fertilization 1 and 1982–1991 for Fertilization 2, as these years maintained the original, balanced treatment design, which was later changed to add further treatments. For the old field succession chronosequence plots, species-specific cover values were used for seven years between 1983 and 2011. Each of the 23 fields had four transects (except for two fields with six transects) of 25 subplots each. For comparability to the other datasets, the 25 transect subplots of 0.5 m × 1 m in each transect were treated as one plot by averaging species-specific cover values across the subplots within transects, resulting in four (or six) plots for each of the 23 fields (96 plots = 21 fields × 4 plots + 2 fields × 6 plots). For the oak savannah dataset, only the plant species cover from 1991 was used; the later years were excluded because they were affected by a seed addition treatment. Species-specific cover was averaged across the 16 0.5 m × 0.5 m subplots per plot.

For the comparative analyses, different years were chosen for these different datasets due to the varying availability of measurements and to ensure a consistently balanced design of the experimental treatments in cases where additional treatments were added at a later stage. The transects in the old field succession chronosequence are likely to inflate certain community properties because their subplots span further across the respective sites than a square plot of the same area would. Similarly, the averaging across subplots in the oak savannah dataset might influence the direct comparability to the biodiversity experiment data. As such, data from the old field succession chronosequence and the oak savannah dataset are shown in Fig. 1e to put the BioDIV data into perspective by adding different kinds of real-world data. However, when it came to constraining the biodiversity experiment data with the real-world data (Fig. 1g), we took a conservative approach and included only those real-world datasets that were most comparable in terms of survey methodology (Fertilization 1 and Fertilization 2; combined US real world). Similarly, for the Jena Experiment real-world counterparts, we considered only the German real-world and Jena real-world plots as purely non-biodiversity experiment plots in Fig. 1c (combined German real world).

To enable direct comparisons of plant communities, the species-specific cover and biomass values for all projects were transformed to relative abundance, in which the single abundance values in each community sum to 100. To do this, all Jena Experiment cover values (originally estimated on a decimal scale[76]) were first transformed to percent cover values[77]. Where vegetation covered more or less than 100% of the vegetation survey area (29% of all communities in the German dataset had total cover values below 100%), it was scaled to 100% for the calculation of relative abundance and, subsequently, community properties. Some communities had a low overall cover, indicating bare ground. Specifically, although communities with a high percentage of bare ground were present in both experiments and the real world, they were more common in the German biodiversity experiment than in its real-world counterparts. An equivalent assessment in the US datasets was not possible, as relative abundance was here based on biomass rather than cover data for most communities (see above). Removing high-bare-ground communities, where possible, might have led to an arbitrary, artificial convergence of plant-community properties from biodiversity experiments and real-world communities that would have weakened the direct comparison between those plant communities, a central aim of this study. Consequently, all communities were retained in the analysis.

Species synonyms and phylogeny. As we used plant species cover, biomass and trait data from multiple sources based on research across decades and different geographic regions, there was considerable variation in the classification and nomenclature of species. Additionally, since the TRY database[78] was queried for plant traits and we also used a phylogenetic backbone tree (see below), the various datasets contained species names that might not all currently have the status of accepted names, challenging the linkage of the different datasets. This issue was dealt with by creating "code" data frames that linked all original spellings, including outdated and synonym names, that appeared in the original data files to the respective accepted species names obtained using The Plant List via the function TPL in R package Taxonstand[79].

To calculate phylogenetic diversity metrics and to use phylogenetic relatedness to assist the imputation of missing trait data, a phylogenetic tree of all plant species was created and included in our study. We adopted the nomenclatural criteria in The Plant List v.1.1 (ref. [80]) for the species in our dataset and pruned the updated vascular plant megaphylogeny by Qian and Jin[81] to include only the species in our study ($n = 664$). We used the software SUNPLIN (ref. [82]) to add the species lacking from the megaphylogeny ($n = 132$, or 19.9% of all species in our study) at random within the crown nodes of the corresponding monophyletic genera. In a few cases where the genera of the missing species were polyphyletic (*Potentilla*, *Medicago*, *Solidago* and *Galium*) or paraphyletic (*Calamagrostis* and *Vicia*), we inserted the species at random within the nodes representing the most recent common ancestors that unequivocally contain them[83]. We repeated this procedure iteratively to obtain 50 phylogenetic trees (see Supplementary Fig. 9 for one example tree and the distribution of randomly inserted species). When using the phylogenetic trees in the subsequent data analysis (the calculation of phylogenetic diversity metrics and plant trait imputation), all 50 trees were used, and the results were averaged.

Functional trait data. To calculate CWM trait values for all plant communities, functional trait data from the TRY database (see Supplementary Table 13) were complemented with in situ collected trait data from Cedar Creek that were not published in TRY. Plant species-specific functional trait values were calculated separately for the German and US species subsets.

Trait data for leaf area (mm^2), leaf dry mass (mg), LDMC ($g\,g^{-1}$), leaf N ($mg\,g^{-1}$), leaf P ($mg\,g^{-1}$), plant height (m), SLA ($mm^2\,mg^{-1}$) and seed mass (dry mass in mg) were assembled[84]. These traits were selected because they are important for ecosystem functioning[46,47] and data for them were available. For the details of processing TRY and other trait data to generate species-level values, see Supplementary Methods.

To fill gaps in the trait data, trait values from same-genus species with available trait information were inferred. The phylopars function in the R package Rphylopars[85] was then employed to impute missing data on the basis of available information on other traits and the phylogenetic tree[86]. Before imputation, all trait data were natural-log transformed. To account for phylogenetic uncertainty (see above), trait data for all 50 phylogenetic trees were imputed and averaged. The plant species and their trait values were then visualized in a PCA for each region (Supplementary Fig. 10) to check for strong outliers and check the outlier species' ability to score extreme values. For details on the importance of species without original trait data (before genus inference and imputation) and for the number of species with identical trait information after inference and imputation, see Supplementary Table 14.

Calculation of plant-community properties. Before the plant-community properties were calculated, tree species, occurring as seedlings, were removed from all datasets. This was because of their strong impact on the calculated CWMs and functional metrics, due to strong differences in trait expression between sapling (observed in the grasslands) and adult trees (studied for functional traits), and the fact that most grasslands in these climates (including the experiments) are grazed, mown or burned regularly, thus preventing tree invasion. Plant-community properties were calculated for each plot–year combination so that the temporal development (succession) of the plots was accounted for in our analysis. As taxonomic diversity indices, we calculated species richness (S), Shannon's diversity index (H), Simpson's diversity index (D1) and D2 (calculated as $D1 = 1 - D$ and $D2 = 1/D$, where D is the sum over all p_i^2 and p_i are the relative abundances of all species i) with the functions specnumber and diversity in R package vegan[87] and Simpson's evenness SEve (by dividing D2 by species richness)[88–91]. As phylogenetic diversity indices, we used Faith's phylogenetic diversity (PD), mean pairwise distance (MPD) and mean nearest taxon distance (MNTD)[92] with the functions pd, mpd and mntd in R package picante[93], where MPD and MNTD were calculated with abundance weighting. All three phylogenetic diversity properties were calculated for each of the 50 phylogenetic trees and averaged to account for phylogenetic uncertainty (see above). For the calculation of the functional diversity indices FRic, FEve, functional divergence (FDiv), functional dispersion (FDis), Rao's quadratic entropy (RaoQ)[94–96] and CWM traits, the function dbFD in the R package FD (refs. [95,97]) was used with correction method cailliez. As the function dbFD relies on the computation of a Gower dissimilarity matrix where zero-dissimilarity values between two species (identical trait values) are not allowed, we slightly altered the trait values of a small number of species by deliberately increasing all trait values by 0.001 to 0.002% for the function to run. For each of the respective species pairs, only the species with the lower overall cover (throughout the regional dataset) received this alteration (Supplementary Table 15). For all but FRic, the abundance-weighted versions of these indices were computed. Communities comprising fewer than three species were assigned a value of zero for FRic, FEve, FDiv, PD, MPD and MNTD, as their computation is not possible for such communities.

Multivariate analysis of experiment and real-world intersection. *Multivariate comparison.* All analyses were carried out in R version 3.4.2 (ref. [98]). Here, a multivariate PCA approach was employed, on the basis of numerous plant-community properties, to assess the distribution, similarities and differences between plant communities of biodiversity experiments and real-world systems. Our approach is based on the relative distribution of plant communities in multidimensional, multivariate space. As this distribution is highly dependent on the community properties entering the PCAs and the information they carry, we took care to avoid multicollinearity[99] among these community properties and thus avoid overamplifying information shared by several community properties.

To quantitatively assess which variables carried redundant information, we tested for multicollinearity of community properties by calculating vif (R function corvif, provided by ref. [100]). In the German and US datasets, we sequentially removed the variables with the highest vif value until all vif values were <3. Only the last of the eight variables removed differed between the German and US datasets, so for comparability between regional datasets, we removed all nine variables from both datasets (Supplementary Tables 16 and 17). Specifically, Shannon's diversity index, FDis, species richness, leaf area, D1, PD, MPD, RaoQ and FDiv were removed (in order of sequential removal), and only the following 12 community properties were employed in the PCAs: D2, SEve, FRic, FEve, SLA, leaf dry mass, leaf N, leaf P, seed mass, height, LDMC and MNTD (Fig. 1b,f; see Extended Data Figs. 6 and 7 for the variance explained by all PCA axes and scores of the 12 community properties for the first two axes, respectively). This vif-justified removal of community properties that were highly correlated with species richness also helps with the issue of differences in species richness being caused by differing vegetation-survey areas in the German real-world and Jena Experiment communities (see above). To test what impact the selection of community properties entering the PCA had on our results, we re-ran our analysis using various subsets of community properties or all of them (see below, Supplementary Information on Sensitivity Analyses 1 and Supplementary Tables 1–3). Separate community-property PCAs were computed for the German and US data subsets using the rda function in R package vegan (with the variables scaled to avoid bias due to different range sizes of the properties), and the data were visualized in biplots with 95% confidence ellipses (Fig. 1a,e; see ref. [101] for the full dataset entering the PCAs).

Intersection-calculation methods. The intersection between the experimental and real-world plots was calculated using three methods of differing complexity, all based on the community-property PCAs presented in Fig. 1a,e. Intersections were calculated between two groups of data per geographic region: (1) all experimental communities across all years and (2) a subset of the most comparable and data-rich real-world datasets (combined real-world datasets). As described above, for Jena, the related combined real-world communities used in this intersection analysis were only the German real-world communities (Biodiversity Exploratories) and the Jena real-world communities. For BioDIV, only the Fertilization 1 and Fertilization 2 plots were used as the combined real-world counterparts when calculating the intersections. First, the first two PCA axes were used to assess the two-dimensional intersection of the 95% confidence ellipses for the experimental and real-world data using the functions ellipse and point.in.polygon in R packages car[102] and sp[103,104], respectively (Supplementary Fig. 1). Second, the first three PCA axes were employed to compute the intersection of three-dimensional convex hull volumes using the functions convhulln and tsearchn in R package geometry[105] (Fig. 1c,g shows two-dimensional representations of the three-dimensional convex hull volumes). Third, using the first three PCA axes, three-dimensional hypervolumes were computed using the hypervolume package in R (ref. [106]). The intersection hypervolume of the experimental and real-world hypervolumes was then calculated, and the function hypervolume_inclusion_test was used to assess which communities fall in the intersection hypervolume (Supplementary Fig. 1). For the subsequent analysis of BEF relationships, experimental plots were defined as realistic if their plant communities fell inside the intersection in at least one of the years present in the dataset. Higher thresholds (for example, 90% of the years inside the intersection) may be inappropriate given that the early years of the experiment see the establishment of sown communities, and would have rendered too few Jena Experiment plots realistic to adequately assess BEF relationships in the constrained datasets (Supplementary Information on Sensitivity Analyses 1 and Supplementary Tables 2 and 3). As such, the inclusion criterion used resulted in the selection of the most realistic experimental plots, while also providing a sufficient number of realistic plots to compare BEF relationships in the constrained and unconstrained datasets. Given this threshold, each plot in the experiments was defined as either realistic (that is, the plot's plant community was within the intersection in at least one year) or unrealistic. Calculating the intersection on the basis of three different methods of different complexity demonstrated that the selection of realistic communities was largely insensitive to the underlying methodology (Supplementary Tables 2 and 3). We therefore focus our analyses on using three-dimensional convex-hull volumes, a method of intermediate complexity, and present results for the other methods in the Supplementary Information.

Measurement of ecosystem-function variables. A range of above- and belowground ecosystem process rates and state variables was selected as ecosystem functions from the Jena Experiment and BioDIV in such a way that the functions of these experiments were as comparable as possible. Only function data obtained between 2006 and 2015 (at least four years after the initiation of the experiments) were used because BEF relationships shortly after the initial establishment of experiments are often unrepresentative of longer-term trends[16,107]. These selection criteria resulted in the following functions: plant aboveground biomass (biomass), aboveground plant biomass C:N ratio (plant C:N), soil organic carbon (C) and root biomass were available for both experiments. As soil inorganic C should not play a role in BioDIV due to the sandy soil, measurements of total C can be considered representative of organic C stocks here (Supplementary Methods). Herbivory rate, soil microbial biomass C, phosphatase activity and pollinator abundance were available only for Jena. For details regarding the measurement of these ecosystem functions in the Jena Experiment and BioDIV, please refer to the Supplementary Methods.

Statistical analysis of unconstrained and constrained experimental BEF relationships. To assess whether and how much BEF relationships change when excluding unrealistic plots from the analysis, each relationship was first analysed in the unconstrained dataset with all experimental plots. The biodiversity experiment datasets were then constrained to include only realistic plots, and the models were re-run. For ecosystem function variables with multiple years of data, the values were averaged across years, and simple linear models were fit that tested for the effect of realized target species richness ($\log_2$, averaged per plot between 2006 and 2015) on the individual functions. Where necessary, square-root or $\log_{10}$ transformation was applied to the response variables to meet the model assumptions of normality and homoscedasticity of variances. For each of the resulting relationships, slope estimates and their 95% confidence intervals (function confint in R) were calculated. The slopes and confidence intervals of each pair of constrained and unconstrained relationships were compared to decide whether the slope or sign of the relationship had changed. If the confidence intervals of unconstrained and constrained slopes included each other's mean value, we concluded that they were not significantly different. Additionally, a paired *t*-test directly comparing the slope values estimated from unconstrained and constrained data subsets (for the 12 BEF relationships in Fig. 2, $n = 12$) was performed.

Sensitivity analyses. Since our analysis involved many decisions on which variables to include and what exact analytical pathway to follow, and these decisions might affect our results, several sensitivity analyses were performed regarding different aspects of our analysis.

To test whether different subsets of community properties entering the PCA affected our results, our analysis was re-run for combinations of (1) different subsets of community properties—that is, (a) the vif-selected 12 community properties (presented in the main text), (b) all available 21 community properties and (c) four subsets excluding one class of community properties (taxonomic, phylogenetic, functional diversity or CWM functional traits, respectively)—and (2) three methods to compute the intersection between the biodiversity experiment and real-world plots, described above (Supplementary Figs. 1 and 3). These community-property subsets were used to demonstrate how strongly the results were influenced by each class of community properties. To keep the number of sensitivity analyses manageable given the high number of possible combinations of community properties and overlap calculation methods, only the vif-selected subset and the set containing all 21 community properties were tested with all three methods. Additionally, we conducted a series of sensitivity analyses that assessed the impact of other methodological changes on the PCA-based selection of realistic biodiversity experiment plots. They include using more subsets of community properties (sensitivity analysis A), including more principal components (axes) of the PCA to define realistic plots based on higher-dimensional space (B), including all available real-world datasets (not just the most methodologically comparable ones, C), using different inclusion criteria to define experimental plots as realistic (D), using species-abundance-based NMDS rather than community-property-based PCAs to assess the intersections of different datasets (E, Supplementary Fig. 2) and including only those German real-world plots in the PCAs that resemble the Jena Experiment in their land use (F). The details on the methodology and results of these sensitivity analyses are described in Supplementary Information on Sensitivity Analyses 1, Supplementary Tables 1–3 and Supplementary Figs. 1–3.

To test whether shifts in the significance of BEF relationships in Fig. 2 simply resulted from the strong reduction of error degrees of freedom associated with using data subsets, we performed a sensitivity analysis in which we randomly selected the same proportion of plots as realistic as that in our PCA-driven selection of realistic plots, 500 times for each relationship (Supplementary Information on Sensitivity Analyses 2 and Supplementary Fig. 4). We also performed an alternative version of this sensitivity analysis that restricted the random draws of Jena Experiment plots to only those with a species richness falling within the truncated species-richness gradient of the realistic Jena plots (Supplementary Fig. 5).

To gain further insight into our findings at Jena, data from the experimental plots that were abandoned and allowed to undergo natural succession (Jena invasion plots) were more closely analysed. Over time, these migrated towards the multivariate community-property space occupied by real-world communities (Extended Data Fig. 2).

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

Data availability

The data supporting the findings of our study are available at https://doi.org/10.25829/idiv.1869-11-3082.

Code availability

The R code to reproduce the findings and figures of our study is available at https://doi.org/10.25829/idiv.1869-11-3082.

Received: 31 July 2019; Accepted: 17 July 2020;
Published online: 24 August 2020

References

1. Cardinale, B. J. et al. Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
2. Tilman, D., Isbell, F. & Cowles, J. M. Biodiversity and ecosystem functioning. *Annu. Rev. Ecol. Evol. Syst.* **45**, 471–493 (2014).
3. Isbell, F. et al. Linking the influence and dependence of people on biodiversity across scales. *Nature* **546**, 65–72 (2017).
4. van der Plas, F. Biodiversity and ecosystem functioning in naturally assembled communities. *Biol. Rev.* **94**, 1220–1245 (2019).
5. Schulze, E.-D. & Mooney, H. *Biodiversity and Ecosystem Functioning* (Springer, 1993).
6. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
7. Balvanera, P. et al. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9**, 1146–1156 (2006).
8. Hines, J. et al. Mapping change in biodiversity and ecosystem function research: food webs foster integration of experiments and science policy. *Adv. Ecol. Res.* **61**, 297–322 (2019).
9. Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
10. Roscher, C., Schumacher, J. & Baade, J. The role of biodiversity for element cycling and trophic interactions: an experimental approach in a grassland community. *Basic Appl. Ecol.* **121**, 107–121 (2004).
11. Tilman, D. et al. Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
12. Hooper, D. U. et al. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
13. Cardinale, B. J. et al. The functional role of producer diversity in ecosystems. *Am. J. Bot.* **98**, 572–592 (2011).
14. O'Connor, M. I. et al. A general biodiversity–function relationship is mediated by trophic level. *Oikos* **126**, 18–31 (2017).
15. Loreau, M. et al. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**, 804–808 (2001).
16. Reich, P. B. et al. Impacts of biodiversity loss escalate through time as redundancy fades. *Science* **336**, 589–592 (2012).
17. Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
18. Grime, J. P. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* **86**, 902–910 (1998).
19. Wardle, D. A. et al. Biodiversity and ecosystem function: an issue in ecology. *Bull. Ecol. Soc. Am.* **81**, 235–239 (2000).
20. Leps, J. What do the biodiversity experiments tell us about consequences of plant species loss in the real world? *Basic Appl. Ecol.* **5**, 529–534 (2004).
21. Srivastava, D. S. & Vellend, M. Biodiversity–ecosystem function research: is it relevant to conservation? *Annu. Rev. Ecol. Evol. Syst.* **36**, 267–294 (2005).
22. Duffy, J. E. Why biodiversity is important to the functioning of real-world ecosystems. *Front. Ecol. Environ.* **7**, 437–444 (2008).
23. Duffy, J. E. Biodiversity effects: trends and exceptions—a reply to Wardle and Jonsson. *Front. Ecol. Environ.* **8**, 11–12 (2010).
24. Wardle, D. A. & Jonsson, M. Biodiversity effects in real ecosystems—a response to Duffy. *Front. Ecol. Environ.* **8**, 10–11 (2010).
25. Wardle, D. A. Do experiments exploring plant diversity–ecosystem functioning relationships inform how biodiversity loss impacts natural ecosystems? *J. Veg. Sci.* **27**, 646–653 (2016).
26. Manning, P. et al. Transferring biodiversity-ecosystem function research to the management of 'real-world' ecosystems. *Adv. Ecol. Res.* **61**, 323–356 (2019).
27. Wilsey, B. J. & Potvin, C. Biodiversity and ecosystem functioning: importance of species evenness in an old field. *Ecology* **81**, 887–892 (2000).
28. Wilsey, B. J. & Polley, H. W. Realistically low species evenness does not alter grassland species-richness–productivity relationships. *Ecology* **85**, 2693–2700 (2004).
29. Hillebrand, H., Bennett, D. & Cadotte, M. Consequences of dominance: a review of evenness effects on local and regional ecosystem processes. *Ecology* **89**, 1510–1520 (2008).
30. Schmitz, M. et al. Consistent effects of biodiversity on ecosystem functioning under varying density and evenness. *Folia Geobot.* **48**, 335–353 (2013).
31. Finn, J. A. et al. Ecosystem function enhanced by combining four functional types of plant species in intensively managed grassland mixtures: a 3-year continental-scale field experiment. *J. Appl. Ecol.* **50**, 365–375 (2013).
32. Weisser, W. W. et al. Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: patterns, mechanisms, and open questions. *Basic Appl. Ecol.* **23**, 1–73 (2017).
33. Schmid, B. & Hector, A. The value of biodiversity experiments. *Basic Appl. Ecol.* **5**, 535–542 (2004).
34. Eisenhauer, N. et al. Biodiversity–ecosystem function experiments reveal the mechanisms underlying the consequences of biodiversity change in real world ecosystems. *J. Veg. Sci.* **27**, 1061–1070 (2016).
35. Isbell, F. et al. Nutrient enrichment, biodiversity loss, and consequent declines in ecosystem productivity. *Proc. Natl Acad. Sci. USA* **110**, 11911–11916 (2013).
36. 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).
37. Buchmann, T. et al. Connecting experimental biodiversity research to real-world grasslands. *Perspect. Plant Ecol. Evol. Syst.* **33**, 78–88 (2018).
38. Tilman, D. et al. The influence of functional diversity and composition on ecosystem processes. *Science* **277**, 1300–1302 (1997).
39. 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).
40. Isbell, F. et al. Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature* **526**, 574–577 (2015).
41. Fischer, M. et al. Implementing large-scale and long-term functional biodiversity research: the biodiversity exploratories. *Basic Appl. Ecol.* **11**, 473–485 (2010).
42. Soliveres, S. et al. Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality. *Nature* **536**, 456–459 (2016).
43. Tilman, D. Secondary succession and the pattern of plant dominance along experimental nitrogen gradients. *Ecol. Monogr.* **57**, 189–214 (1987).
44. Clark, C. M. & Tilman, D. Loss of plant species after chronic low-level nitrogen deposition to prairie grasslands. *Nature* **451**, 712–715 (2008).
45. Inouye, R. et al. Old-field succession on a Minnesota sand plain. *Ecology* **68**, 12–26 (1987).
46. Díaz, S. et al. The global spectrum of plant form and function. *Nature* **529**, 167–171 (2015).
47. Craven, D. et al. Multiple facets of biodiversity drive the diversity–stability relationship. *Nat. Ecol. Evol.* **2**, 1579–1587 (2018).
48. Nakamura, G., Gonçalves, L. O. & da Silva Duarte, L. Revisiting the dimensionality of biological diversity. *Ecography (Cop.)* **43**, 539–548 (2020).
49. Stevens, R. D. & Tello, J. S. On the measurement of dimensionality of biodiversity. *Glob. Ecol. Biogeogr.* **23**, 1115–1125 (2014).
50. Manning, P. et al. Simple measures of climate, soil properties and plant traits predict national-scale grassland soil carbon stocks. *J. Appl. Ecol.* **52**, 1188–1196 (2015).
51. Adler, D. & Kelly, T. vioplot: Violin plot. R package version 0.3.0 (2018).
52. Loreau, M. & Hector, A. Partitioning selection and complementarity in biodiversity experiments. *Nature* **412**, 72–76 (2001).
53. Allan, E. et al. Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecol. Lett.* **18**, 834–843 (2015).
54. Le Bagousse-Pinguet, Y. et al. Phylogenetic, functional, and taxonomic richness have both positive and negative effects on ecosystem multifunctionality. *Proc. Natl Acad. Sci. USA* **116**, 8419–8424 (2019).
55. Venail, P. et al. Species richness, but not phylogenetic diversity, influences community biomass production and temporal stability in a re-examination of 16 grassland biodiversity studies. *Funct. Ecol.* **29**, 615–626 (2015).
56. Hillebrand, H. & Matthiessen, B. Biodiversity in a complex world: consolidation and progress in functional biodiversity research. *Ecol. Lett.* **12**, 1405–1419 (2009).
57. Grace, J. B. et al. Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature* **529**, 390–393 (2016).
58. Liang, J. et al. Positive biodiversity–productivity relationship predominant in global forests. *Science* **354**, aaf8957 (2016).
59. Oehri, J., Schmid, B., Schaepman-Strub, G. & Niklaus, P. A. Biodiversity promotes primary productivity and growing season lengthening at the landscape scale. *Proc. Natl Acad. Sci. USA* **114**, 10160–10165 (2017).
60. Díaz, S. et al. Incorporating plant functional diversity effects in ecosystem service assessments. *Proc. Natl Acad. Sci. USA* **104**, 20684–20689 (2007).
61. Lavorel, S. et al. Using plant functional traits to understand the landscape distribution of multiple ecosystem services. *J. Ecol.* **99**, 135–147 (2011).
62. Schmid, B. The species richness–productivity controversy. *Trends Ecol. Evol.* **17**, 113–114 (2002).
63. Loreau, M. Biodiversity and ecosystem functioning: a mechanistic model. *Proc. Natl Acad. Sci. USA* **95**, 5632–5636 (1998).
64. Maestre, F. T. et al. Plant species richness and ecosystem multifunctionality in global drylands. *Science* **335**, 214–218 (2012).
65. van der Plas, F. et al. Jack-of-all-trades effects drive biodiversity–ecosystem multifunctionality relationships in European forests. *Nat. Commun.* **7**, 11109 (2016).

66. Socher, S. A. et al. Direct and productivity-mediated indirect effects of fertilization, mowing and grazing on grassland species richness. *J. Ecol.* **100**, 1391–1399 (2012).
67. Hobbs, R. J., Higgs, E. & Harris, J. A. Novel ecosystems: implications for conservation and restoration. *Trends Ecol. Evol.* **24**, 599–605 (2009).
68. Klaus, V. H. et al. Do biodiversity–ecosystem functioning experiments inform stakeholders how to simultaneously conserve biodiversity and increase ecosystem service provisioning in grasslands? *Biol. Conserv.* **245**, 108552 (2020).
69. Roscher, C. et al. Convergent high diversity in naturally colonized experimental grasslands is not related to increased productivity. *Perspect. Plant Ecol. Evol. Syst.* **20**, 32–45 (2016).
70. Ellenberg, H. & Leuschner, C. *Vegetation Mitteleuropas mit den Alpen: In Ökologischer, Dynamischer und Historischer Sicht* (UTB, 2010).
71. Blüthgen, N. et al. A quantitative index of land-use intensity in grasslands: integrating mowing, grazing and fertilization. *Basic Appl. Ecol.* **13**, 207–220 (2012).
72. Tilman, D., Reich, P. B. & Knops, J. M. H. Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature* **441**, 629–632 (2006).
73. Tilman, D. Community invasibility, recruitment limitation, and grassland biodiversity. *Ecology* **78**, 81–92 (1997).
74. Catford, J. A. et al. Traits linked with species invasiveness and community invasibility vary with time, stage and indicator of invasion in a long-term grassland experiment. *Ecol. Lett.* **22**, 593–604 (2019).
75. Fargione, J. et al. From selection to complementarity: shifts in the causes of biodiversity–productivity relationships in a long-term biodiversity experiment. *Proc. R. Soc. B* **274**, 871–876 (2007).
76. Londo, G. The decimal scale for releves of permanent quadrats. *Vegetatio* **33**, 61–64 (1976).
77. Roscher, C. et al. What happens to the sown species if a biodiversity experiment is not weeded? *Basic Appl. Ecol.* **14**, 187–198 (2013).
78. Kattge, J. et al. TRY—a global database of plant traits. *Glob. Change Biol.* **17**, 2905–2935 (2011).
79. Cayuela, L., Stein, A. & Oksanen, J. Taxonstand: Taxonomic standardization of plant species names. R package version 2.1 (2017).
80. *The Plant List* version 1.1 (2013); http://www.theplantlist.org/
81. Qian, H. & Jin, Y. An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *J. Plant Ecol.* **9**, 233–239 (2016).
82. Martins, W. S., Carmo, W. C., Longo, H. J., Rosa, T. C. & Rangel, T. F. SUNPLIN: simulation with uncertainty for phylogenetic investigations. *BMC Bioinform.* **14**, 324 (2013).
83. Rangel, T. F. et al. Phylogenetic uncertainty revisited: implications for ecological analyses. *Evolution* **69**, 1301–1312 (2015).
84. Cornelissen, J. H. C. et al. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Aust. J. Bot.* **51**, 335–380 (2003).
85. Goolsby, E. W., Bruggeman, J. & Ane, C. Rphylopars: Phylogenetic comparative tools for missing data and within-species variation. R package version 0.2.9 (2016).
86. Penone, C. et al. Imputation of missing data in life-history trait datasets: which approach performs the best? *Methods Ecol. Evol.* **5**, 961–970 (2014).
87. Oksanen, J. et al. Vegan: Community ecology package. R package version 2.3-4 (2016).
88. Hill, M. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**, 427–432 (1973).
89. Smith, B. & Wilson, J. B. A consumer's guide to evenness indices. *Oikos* **76**, 70–82 (1996).
90. Magurran, A. *Measuring Biological Diversity* (Blackwell, 2004).
91. Morris, E. K. et al. Choosing and using diversity indices: insights for ecological applications from the German biodiversity exploratories. *Ecol. Evol.* **4**, 3514–3524 (2014).
92. Tucker, C. M. et al. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol. Rev.* **92**, 698–715 (2017).
93. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
94. Villéger, S., Mason, N. W. H. & Mouillot, D. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology* **89**, 2290–2301 (2008).
95. Laliberte, E. & Legendre, P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91**, 299–305 (2010).
96. Mouchet, M. A., Villéger, S., Mason, N. W. H. & Mouillot, D. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Funct. Ecol.* **24**, 867–876 (2010).
97. Laliberté, E., Legendre, P. & Shipley, B. FD: Measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-12 (2014).
98. *R: A Language and Environment for Statistical Computing* v.3.4.2 (R Core Team, 2019); https://doi.org/10.1007/978-3-540-74686-7
99. Dormann, C. F. et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop.)* **36**, 27–46 (2013).
100. Zuur, A. F., Ieno, E. N. & Elphick, C. S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **1**, 3–14 (2010).
101. Jochum, M. et al. R-code and aggregated data from: The results of biodiversity-ecosystem functioning experiments are realistic. *iDiv Data Repository* https://doi.org/10.25829/idiv.1869-11-3082 (2020).
102. Fox, J. & Weisberg, S. *An R Companion to Applied Regression* (SAGE, 2011).
103. Pebesma, E. & Bivand, R. Classes and methods for spatial data in R. *R News* **5**, 9–13 (2005).
104. Bivand, R. S., Pebesma, E. & Gomez-Rubio, V. *Applied Spatial Data Analysis with R* (Springer, 2013).
105. Habel, K., Grasman, R., Gramacy, R. B., Stahel, A. & Sterratt, D. C. geometry: Mesh generation and surface tessellation. R package version 0.4.1 (2019).
106. Blonder, B. & Harris, D. hypervolume: High dimensional geometry and set operations using kernel density estimation, support vector machines, and convex hulls. R package version 2.0.11 (2018).
107. Meyer, S. T. et al. Effects of biodiversity strengthen over time as ecosystem functioning declines at low and increases at high biodiversity. *Ecosphere* **7**, e01619 (2016).
108. Brownrigg, R. mapdata: Extra map databases. R package version 2.3.0 (2018).

Acknowledgements

We thank the establishers, maintainers, coordinators, technical and research staff, and data owners of all involved projects, as well as the TRY initiative. We thank S. Soliveres and E. Allan for discussion; S. Thiel, G. Luo, D. Bahauddin and F. Schneider for help with data extraction and handling; and R. Junker and B. Blonder for assistance with the calculation of multidimensional hypervolumes. This study was funded through Jena Experiment SP 7 (Swiss National Science Foundation grant no. 310030E-166017/1). Further support came from the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (grant no. FZT 118). The Jena Experiment was funded by the Deutsche Forschungsgemeinschaft (grant nos FOR 456 and FOR 1451) with additional support from Friedrich Schiller University Jena, the Max Planck Institute for Biogeochemistry in Jena and the Swiss National Science Foundation. All Cedar Creek studies are funded by the US National Science Foundation's Long-Term Ecological Research (LTER) programme (grant no. DEB-1234162). F.I. acknowledges funding from the LTER Network Communications Office (grant no. DEB-1545288). We thank the iDiv Data Repository for hosting our R code and aggregated datasets and for performing the related quality checks.

Author contributions

M.J., P.M., M.F. and F.v.d.P. conceived and designed the study. M.J., M.F., F.I., C.R., S.B., G.B., N.B., J.A.C., J.C.-B., A.E., N.E., G.G., N.H., J.K., V.H.K., T.K., M.L., G.L.P., S.T.M., L.M., Y.O., D.P., P.B.R., D.S., S.S., B.S., D.T., T.T., A.V., C.W., A.W., W.W.W., W.W. and P.M. contributed data. M.J. developed the analytical framework and analysed the data. R.M.-V. constructed the phylogenetic hypothesis trees. M.J. and P.M. wrote the manuscript. M.J., M.F., F.I., C.R., F.v.d.P., S.B., G.B., N.B., J.A.C., J.C.-B., A.E., N.E., G.G., N.H., J.K., V.H.K., T.K., M.L., G.L.P., S.T.M., R.M.-V., L.M., Y.O., C.P., D.P., P.B.R., A.R., D.S., S.S., B.S., D.T., T.T., A.V., C.W., A.W., W.W.W., W.W. and P.M. contributed to the discussion of the results and writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

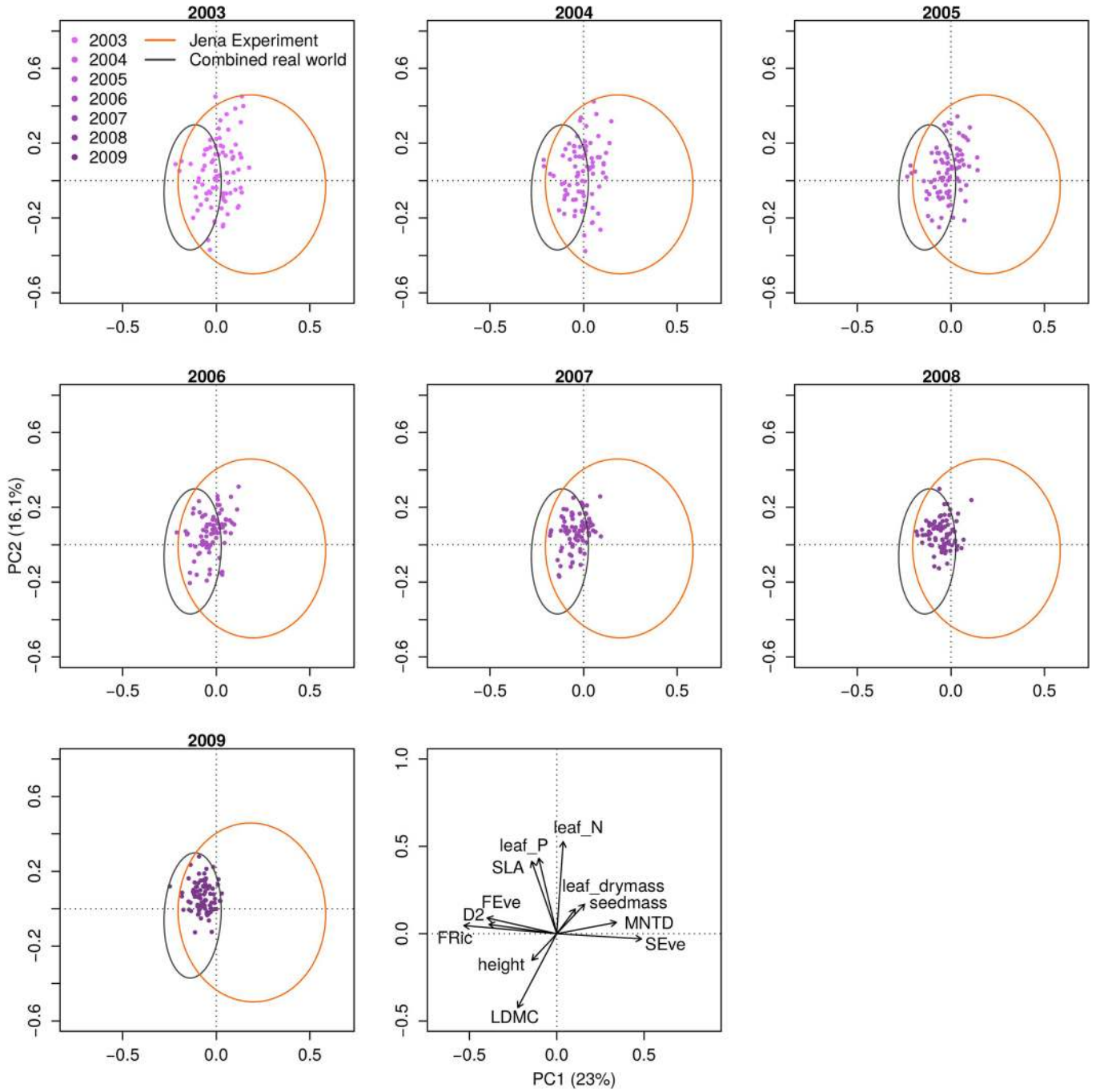
Extended data is available for this paper at https://doi.org/10.1038/s41559-020-1280-9.

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-020-1280-9.

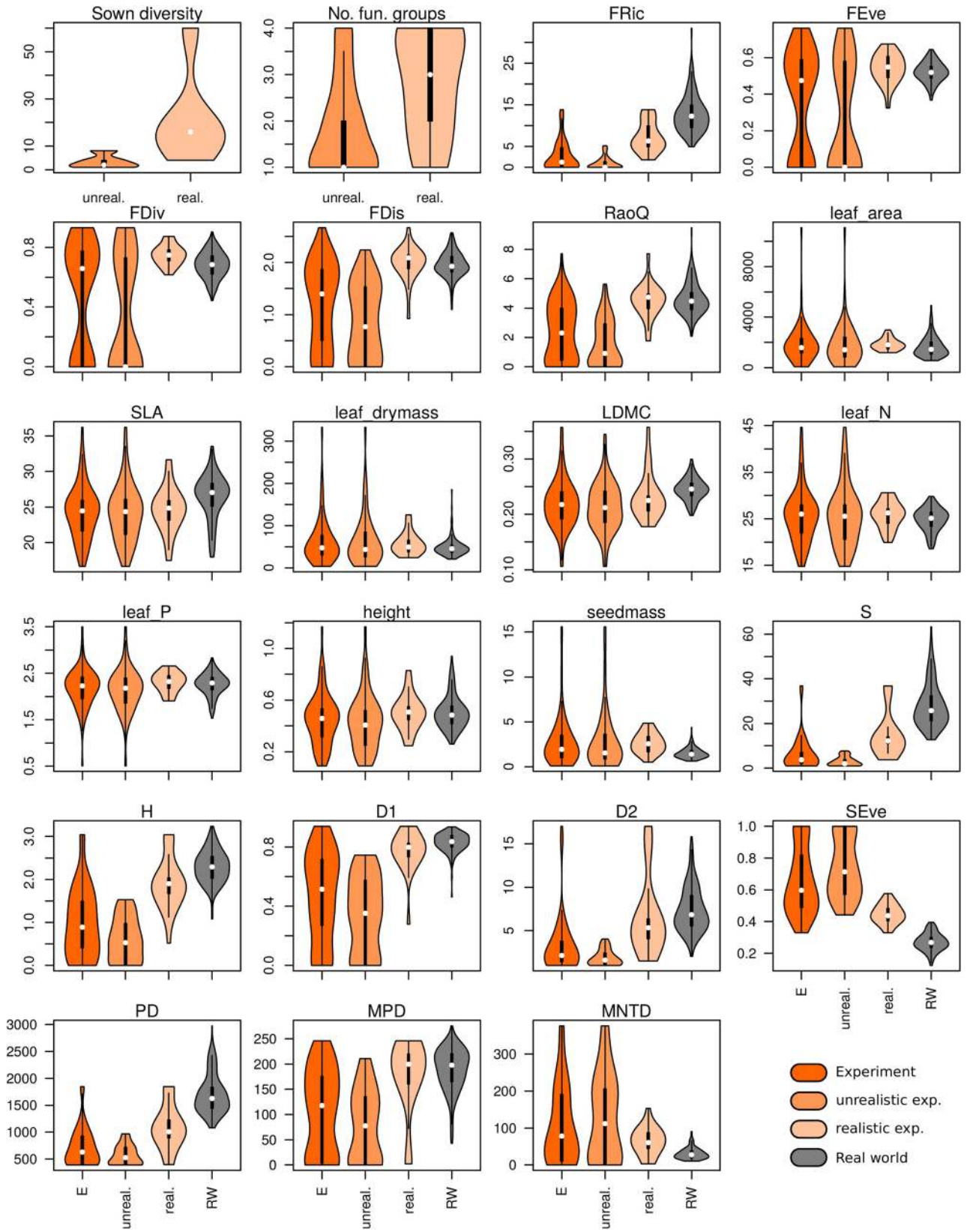
Correspondence and requests for materials should be addressed to M.J.

Country	Project name	Project code	Main reference	Number of plots used	Vegetation data years	Ecosystem functions & years
G	Jena Experiment	Jena Experiment	[10]	82	2003-2015	plant aboveground biomass (2006-15), plant CN (2007-12), soil organic C (2008, 2011, 2014), root biomass (2011, 2014), herbivory (2010-12), soil microbial biomass C (2010), phosphatase activity (2013), pollinator abundance (2010, 2012)
G	Biodiversity Exploratories	German real world	[41]	150	2008-2015	NA
G	Saale grasslands	Jena real world	Roscher unpubl.	14	2011	NA
G	Jena invasion sub-experiments	Jena invasion	[32, 69]	82	2003-2015	NA
G	Jena mown succession plots	Jena succession	[69]	2	2003-2009	NA
U	e120	BioDIV	[38]	159	1996-2015 (not 2009)	plant aboveground biomass (2006-15, not 2009), plant CN (2006), total soil C (2006), root biomass (2010)
U	e001	Fertilization 1	[43]	207	1982-2004	NA
U	e002	Fertilization 2	[43]	162	1982-1991	NA
U	e014	Old field succession chronosequence	[45]	23	1983,1989, 1994, 1997, 2002, 2006, 2011	NA
U	e093	Oak savannah	[73, 74]	30	1991	NA

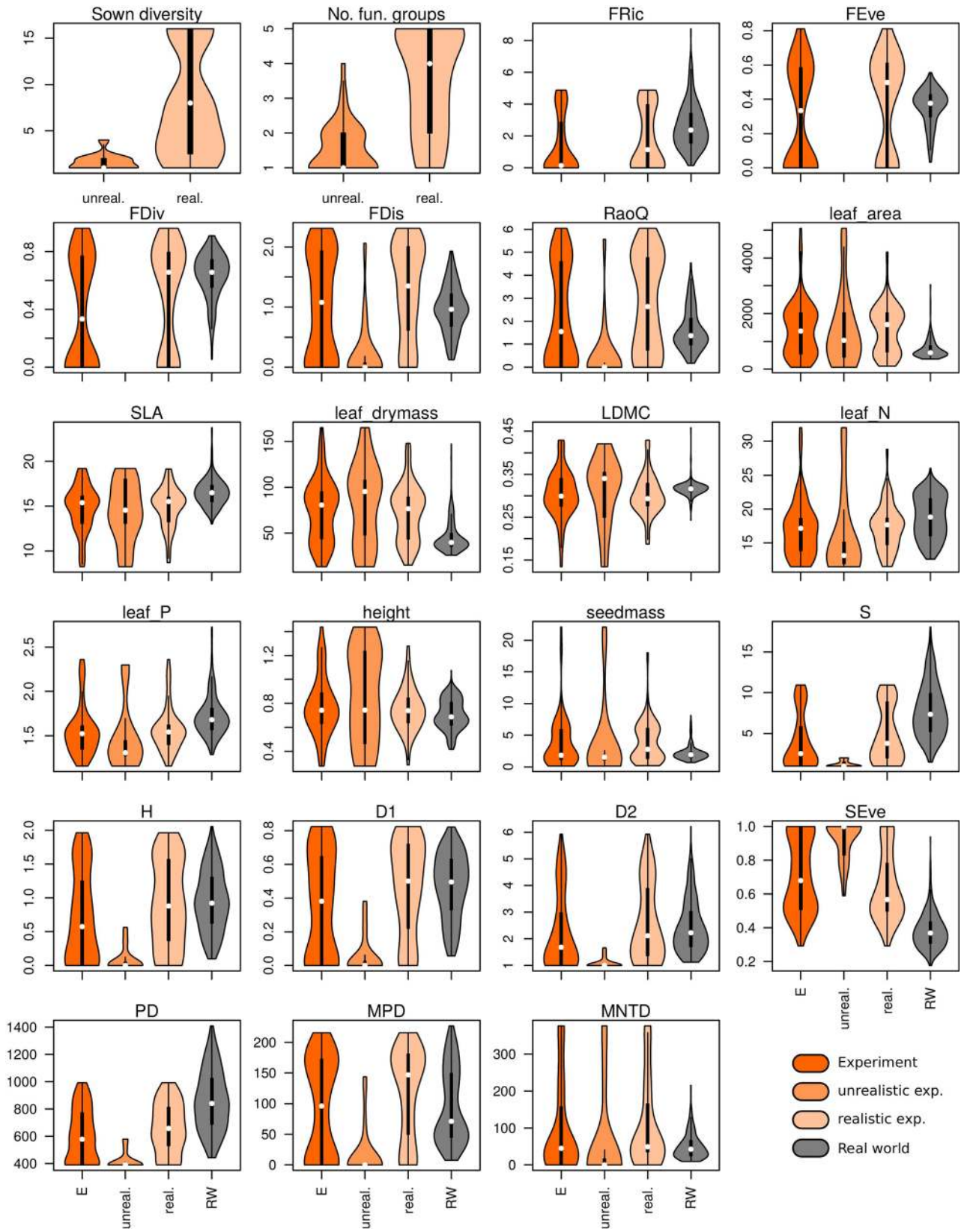
Extended Data Fig. 1 | List of German and US datasets for vegetation and ecosystem function data. Lists country, project name, project code used in this paper, main reference, number of plots we used, years we have vegetation data for, functions we used including years. Most of the raw data is openly available in various online repositories: Jena Experiment (http://jenaexperiment.uni-jena.de/index.php/data/), Biodiversity Exploratories (https://www.bexis.uni-jena.de/Login/Account.aspx), Cedar Creek (https://www.cedarcreek.umn.edu/research/data). Data from the Saale grasslands (Jena real world) were provided by Christiane Roscher and are currently not openly available. Aggregated datasets used for this study are now available online[101].



Extended Data Fig. 2 | Temporal movement of Jena invasion communities into the real-world realm. Based on the PCA in Fig. 1a. Different shades of purple show Jena invasion communities across the years from 2003-2009. Orange and gray ellipses show 95% confidence intervals for Jena Experiment and combined real-world plots (but their communities are not plotted here), respectively. Note that while the points in different panels are from single years, the ellipses are fixed to the across-year comparison in Fig. 1a. The last panel shows the PCA factor loadings for the 12 vif-selected community properties (arrows scaled to improve visibility - "const=25" in R vegan "biplot" function[87]). Within six years of succession, the plant communities of Jena invasion plots fully "moved" into the core of the community property space defined by the combined real-world plots (German real world and Jena real world, respectively).



Extended Data Fig. 3 | Violin plots of all 21 community properties of German data. Experimental (E, Jena Experiment, strong orange, 82 plots), unrealistic experimental (unreal., intermediate orange, 59 plots), selected realistic experimental (real., weak orange, 23 plots) and combined real-world plots (German real world, Jena real world, gray, 164 plots), all averaged across years per plot. Combination of boxplot and rotated kernel density plot (R package "vioplot"[51]). Realistic plots were calculated based on the 12 vif-selected community properties and the convex hull volume method. Units: leaf area (mm^2), leaf dry mass (mg), leaf dry matter content (LDMC, g/g), leaf nitrogen concentration (leaf N, mg/g), leaf phosphorus concentration (leaf P, mg/g), plant height (m), specific leaf area (SLA, mm^2/mg) and seed mass (dry mass in mg). Other community properties are dimensionless.



Extended Data Fig. 4 | Violin plots of all 21 community properties of US data. Experimental (E, BioDIV, strong orange, 159 plots), unrealistic experimental (unreal., intermediate orange, 37 plots), selected realistic experimental (real., weak orange, 122 plots) and combined real-world plots (Fertilization 1 & 2, gray, 369 plots), all averaged across years per plot. Combination of boxplot and rotated kernel density plot (R package "vioplot"[51]). Realistic plots were calculated based on the 12 vif-selected community properties and the convex hull volume method. Units: leaf area (mm^2), leaf dry mass (mg), leaf dry matter content (LDMC, g/g), leaf nitrogen concentration (leaf N, mg/g), leaf phosphorus concentration (leaf P, mg/g), plant height (m), specific leaf area (SLA, mm^2/mg) and seed mass (dry mass in mg). Other community properties are dimensionless.

model	u_n	c_n	u_slop	c_slop	u_low	c_low	u_upp	c_upp	u_p	c_p	u_R2	c_R2
J_biomass	82.00	23.00	2.39	2.03	1.87	0.45	2.91	3.61	0.00	0.01	0.51	0.22
J_plantCN	82.00	23.00	1.02	-2.12	-0.15	-5.92	2.20	1.67	0.09	0.26	0.02	0.02
J_soilorgC	82.00	23.00	0.09	0.06	0.04	-0.10	0.13	0.23	0.00	0.42	0.14	-0.01
J_rootbiomass	80.00	23.00	0.10	0.04	0.07	-0.05	0.14	0.13	0.00	0.37	0.34	-0.01
J_herbivory	80.00	23.00	0.07	0.24	0.02	0.11	0.12	0.38	0.01	0.00	0.08	0.36
J_micBMC	80.00	23.00	48.40	-8.18	23.35	-92.06	73.46	75.71	0.00	0.84	0.15	-0.05
J_phosphatase	80.00	23.00	103.17	16.33	70.38	-98.35	135.97	131.01	0.00	0.77	0.33	-0.04
J_pollinators	79.00	23.00	0.10	0.13	0.06	0.03	0.14	0.23	0.00	0.02	0.27	0.21
BioDIV_biomass	159.00	122.00	0.24	0.26	0.20	0.22	0.28	0.31	0.00	0.00	0.44	0.49
BioDIV_plantCN	158.00	122.00	-0.20	1.22	-1.24	0.16	0.84	2.28	0.71	0.02	-0.01	0.03
BioDIV_soilC	158.00	122.00	0.03	0.03	0.01	0.01	0.04	0.05	0.00	0.00	0.08	0.08
BioDIV_rootbiomass	150.00	117.00	261.03	266.43	221.71	216.44	300.35	316.41	0.00	0.00	0.53	0.49

Extended Data Fig. 5 | Model parameters for BEF relationships presented in Fig. 2. Values are presented for unconstrained (u) and constrained (c) models of Jena (J) and BioDIV BEF relationships. Constraining was done using the 12 vif-selected community properties and the convex hull method. Sample size (n), slope estimates (slop), lower (low) and upper (upp) 95% confidence intervals, p-values (p) and adjusted R^2 values (R2). All values are rounded to two decimal places.

PC	GER percent var. explained	US percent var. explained
PC1	0.23	0.24
PC2	0.16	0.18
PC3	0.13	0.12
PC4	0.11	0.09
PC5	0.09	0.07
PC6	0.07	0.07
PC7	0.05	0.06
PC8	0.04	0.05
PC9	0.04	0.04
PC10	0.03	0.04
PC11	0.02	0.02
PC12	0.02	0.01

Extended Data Fig. 6 | Variance explained by 12 PCA axes (12 vif-selected community properties). Percentage of total variance explained by each of the 12 PCA axes (PC's, see Fig. 1) for each region (GER = Germany and US = USA). Rounded to two decimal places.

	Germany		USA	
	PC1	PC2	PC1	PC2
FRic	-3.47	0.28	2.90	-3.24
FEve	-2.60	0.54	2.19	-2.33
SLA	-0.93	2.47	3.19	-0.14
leaf drymass	0.69	0.84	0.76	1.50
LDMC	-1.45	-2.53	-3.32	-1.86
leaf N	0.24	3.15	2.18	3.21
leaf P	-0.67	2.58	3.69	1.41
height	-0.93	-0.91	-2.71	0.22
seed mass	1.04	1.01	2.80	3.04
D2	-2.53	0.32	3.13	-2.39
SEve	3.16	-0.18	-2.16	3.15
MNTD	2.22	0.39	0.83	2.24

Extended Data Fig. 7 | PCA scores for 12 vif-selected community properties of PCA's in Fig. 1. Scores have been produced using the scores() command of the "vegan" package[87] in R and have been rounded to two decimal places.

nature research

Corresponding author(s): Malte Jochum

Last updated by author(s): July 1st, 2020

Reporting Summary

Nature Research 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 Research 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	No software has been used to collect data.
Data analysis	We used the SUNPLIN software (Martins, 2013, BMC Bioinformatics) to add species lacking from the used megaphylogeny to our phylogenetic trees at random (see Methods section). All other analyses were performed in R. R-code and data are available as Jochum, M. et al. R-code and aggregated data from: The results of biodiversity-ecosystem functioningexperiments are realistic. iDiv Data Repository. DOI: https://doi.org/10.25829/idiv.1869-11-3082,2020.

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Data and Code Availability statements are included.

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	Plant community data on 3329 plot-year combinations in the German and 9954 plot-year combinations in the US comparison. Ecosystem function data for 12 functions / processes from two different biodiversity experiments over several decades. Where permits were necessary, the respective guidelines have been followed by the original data contributors.
Research sample	Original research data from the Jena Experiment, BioDIV (Cedar Creek Ecosystem Science Reserve experiment e120), the Biodiversity Exploratories and other Cedar Creek experiments, the TRY Initiative and comparatively small contributions from single co-authors.
Sampling strategy	For plant community properties, the sample sizes are 3329 and 9945 for the German and US comparison. For the BEF relationships, full dataset sample sizes are 82 plots for the Jena Experiment and 159 plots for BioDIV.
Data collection	Data was collected by original data providers from the major projects involved (Jena Experiment, Cedar Creek Ecosystem Science Reserve, Biodiversity Exploratories, TRY initiative), most of whom are now listed as co-authors on our manuscript.
Timing and spatial scale	Vegetation data: several decades, two distinct geographical regions; Function data: one to many years for the respective functions in each experiment (two distinct places: one in Jena, Germany, one in Cedar Creek, Minnesota, USA).
Data exclusions	Ecosystem function data was only used for the last 10 years (2006-2015).
Reproducibility	Large-scale and long-term biodiversity experiments and real-world observations are hard to replicate. Therefore, we used two sets of experiment and real-world studies. Data and R-code are publicly available.
Randomization	NA
Blinding	NA

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Field-work has been done over decades in the involved projects by original data owners.
Location	Germany: Jena and three regions across Germany for the Biodiversity Exploratories. USA: Cedar Creek, Minnesota, USA
Access & import/export	According to project guidelines of the involved projects (Jena Experiment, Cedar Creek Ecosystem Science Reserve, Biodiversity Exploratories).
Disturbance	According to project guidelines of the involved projects (Jena Experiment, Cedar Creek Ecosystem Science Reserve, Biodiversity Exploratories).

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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