

## LETTER

# Conventional functional classification schemes underestimate the relationship with ecosystem functioning

Justin P. Wright,<sup>1,2\*</sup> Shahid Naeem,<sup>2</sup> Andy Hector,<sup>3</sup> Clarence Lehman,<sup>4</sup> Peter B. Reich,<sup>5</sup> Bernhard Schmid<sup>3</sup> and David Tilman<sup>4</sup>

<sup>1</sup>Department of Biology, Duke University, Box 90338, Durham, NC, USA

<sup>2</sup>Department of Ecology, Evolution, and Environmental Biology, Columbia University, 1200 Amsterdam Ave MC5557, New York, NY, USA

<sup>3</sup>Institute of Environmental Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

<sup>4</sup>Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, MN 55108, USA

<sup>5</sup>Department of Forest Resources, University of Minnesota, St Paul, MN 55108, USA

\*Correspondence: E-mail: jw67@duke.edu

## Abstract

Studies linking the functional diversity of a biota to ecosystem functioning typically employ *a priori* classifications of species into hypothetically complementary groups. However, multiple alternate classifications exist in which the number of functional groups, the number of species per functional group, and the grouping of species differ from the *a priori* scheme. Without assessing the relative precision, or ability of an *a priori* scheme to accurately predict ecosystem functioning relative to its many alternatives, the validity and utility of analyses based on a single *a priori* classification scheme remains unclear. We examine the precision of *a priori* classifications used in 10 experimental grassland systems in Europe and the United States that have found evidence for a significant role of functional plant diversity in governing ecosystem function. The predictive precision of the *a priori* classifications employed in these studies was seldom significantly higher than the precision of random classifications. *Post-hoc* classification schemes that performed well in predicting ecosystem function resembled each other more with regard to species composition than average classifications, but there was still considerable variability in the manner in which these classification schemes grouped species. These results suggest that we need a more nuanced understanding of how the diversity of functional traits of species in an assemblage affects ecosystem functioning.

## Keywords

Biodiversity, classification, complementarity, ecosystem function, functional groups, sampling effect.

*Ecology Letters* (2006) 9: 111–120

## INTRODUCTION

Widespread declines in biodiversity at both global and local scales have motivated considerable research directed towards understanding how changes in biological diversity may affect ecosystem functioning and derived ecosystem services (Loreau *et al.* 2001; Hooper *et al.* 2005). There is a growing consensus that functional diversity is likely to be the component of biodiversity most relevant to ecosystem functioning (Diaz & Cabido 2001; Hooper *et al.* 2002; Naeem & Wright 2003; Reich *et al.* 2004), where functional diversity comprises the diversity and range of functional traits possessed by the biota of an ecosystem. One of the

primary challenges in such research is determining appropriate methods for quantifying functional diversity.

The most commonly used technique for quantifying functional diversity consists of clustering species with shared taxonomic, physiological and morphological traits into functional groups, assuming that groups with similar traits differ in their response to and effect on resources (Chapin *et al.* 1996; Lavorel & Garnier 2002; Petchey & Gaston 2002). The number of functional groups, or functional group richness (FGR), can then be used as an approximation of functional diversity in an ecosystem. This procedure constitutes constructing an *a priori* functional classification which contrasts with a *post hoc* or null approach

in which all possible functional classification schemes are examined, irrespective of the ecological or evolutionary relationships that cluster species, and the scheme that best predicts ecosystem response to biodiversity loss is considered the most appropriate functional classification. While there are merits to both methods, the relative precision of an *a priori* scheme to capture functional diversity can only be assessed by comparing it against its many alternatives.

Most studies examining the effects of FGR on ecosystem functioning have focused on plant diversity, and have employed a functional classification scheme based on the well-established plant functional groups of grasses, non-leguminous forbs, and legumes (Tilman *et al.* 1997; Hector *et al.* 1999; Naeem *et al.* 1999; Symstad 2000; Reich *et al.* 2001), with occasional refinements such as separating grasses into C<sub>3</sub> and C<sub>4</sub> grasses (Tilman *et al.* 1997; Reich *et al.* 2001), separating forbs into tall or short forbs (Roscher *et al.* 2004), or separating forbs into woody and non-woody plants (Tilman *et al.* 1997). While not universally used (Hooper & Vitousek 1997; Hooper & Dukes 2004), these functional types encompass a variety of trait differences and are assumed to represent groups that differentially influence most terrestrial ecosystem functions including net primary productivity, carbon sequestration, nitrogen retention, decomposition, and other processes that affect carbon and nitrogen cycling. We will refer to this widely used grass–forb–legume functional classification scheme as the *a priori* grass/forb/legume (GFL) classification.

Although widely used, this *a priori* GFL classification scheme represents only one of many possible schemes and its precision remains unknown. For example,  $S$  species could be classified into anywhere from 1 to  $S - 1$  groups and for each level of  $F_{\max}$  (i.e. number of groups into which the species are divided) there exists a large number of ways species can be classified. For example, given the 34 species used in the Silwood Park site of the BIODEPTH experiment (Hector *et al.* 1999), one could construct over  $10^{28}$  possible classification schemes. Classification schemes will vary in their ability to account for covariance between FGR and ecosystem functioning in experimental plots. The higher the covariance explained, the higher the precision of the classification scheme.

The *a priori* GFL classification has been widely used in grassland systems (Naeem & Wright 2003) because it is believed to classify plants by their impacts on ecosystem functions relatively effectively, a possibility supported by regression analyses of results from combinatorial manipulative experiments of grassland plant diversity (Tilman *et al.* 1997; Hector *et al.* 1999; Reich *et al.* 2001). However, given the lack of empirical support that explicitly demonstrates that species within these functional groups possess complementary traits, and even some evidence to the contrary (Craine *et al.* 2002; but see Roscher *et al.* 2004), the GFL

classification might best be considered a ‘candidate’ grouping (Vitousek & Hooper 1993). Support for the *a priori* GFL classification is provided by Petchey (2004) who bootstrapped the  $F$  ratio of the change in deviance caused by removing FGR from regression models that included both species richness ( $S$ ) and FGR as the independent variables. By randomizing species assigned to three functional groups in the bootstrapped  $F$  ratios, Petchey (2004) confirmed that FGR, assessed using the GFL classification, was a significant determinant of ecosystem functioning at two to three of the eight sites in the BIODEPTH experiment. These results were mirrored by an analysis of the BioCON experiment which showed that FGR, again assessed using the GFL classification, had an effect on ecosystem functioning independent of species richness (Reich *et al.* 2004). However, to date, the ability of the ‘candidate’ GFL classification scheme to predict ecosystem functioning relative to alternative classification schemes with the same or different numbers of functional groups has not been tested.

Regardless of whether the *a priori* GFL classification scheme is the best of all possible classifications, it is still an open question as to whether any classification scheme that groups species together can effectively describe the functional diversity of an assemblage. Grouping species assumes that the traits of importance are discrete rather than continuously distributed among species, that the variance in traits is smaller within than between species (Chapin *et al.* 1996), and that if multiple traits are responsible for controlling ecosystem functions that these traits tend to be correlated within species, presumably due to trade-offs. To date, most evidence shows that within assemblages there tend to be fairly continuous distributions of traits (Craine *et al.* 2001, 2002; Reich *et al.* 2003; Diaz *et al.* 2004). While at global scales there is evidence for trade-offs between different traits thought to modify ecosystem functioning (Grime *et al.* 1997; Diaz *et al.* 2004; Wright *et al.* 2004), whether these trade-offs exist within assemblages that exist in a common environment is unknown (Grime 1998). Furthermore, using the same classification scheme to predict ecosystem functioning assumes that the same traits are responsible for regulating different ecosystem functions. Thus it is unclear whether or not, an ‘optimal’ classification exists and whether, even within a given site, the effect of functional diversity on different ecosystem functions is best captured by using a single functional classification scheme.

Here, we compare the success of the *a priori* GFL classification scheme in predicting three ecosystem functions in 10 experimental grassland ecosystems to randomly assembled functional classifications. We use a calculation of the similarity of the top *post hoc* classification schemes generated for each site and each ecosystem function to determine whether they are converging on an ‘optimal’ classification scheme. We also compare the similarity of the

top *post hoc* classification schemes for different ecosystem functions at a site to determine the degree to which a 'universal' classification scheme, i.e. a scheme that works equally well for all ecosystem functions, exists.

## METHODS

We examined the relationship between functional diversity and ecosystem functioning using data from grassland diversity-functioning experiments conducted at 10 sites (Table 1). Each of these experiments manipulated the species richness of plots and assessed the effect of functional diversity on ecosystem functioning by regressing plot-level measures of productivity, as estimated by above- and below-ground peak biomass, and nutrient retention, as estimated by soil nitrogen concentrations, against the number of functional groups (as defined by the *a priori* GFL scheme) present in a plot. The BioCON experiment

was conducted at Cedar Creek Natural History Area, Minnesota, USA (for details see Reich *et al.* 2001). We analysed data from the 1999 growing season in the ambient plots (i.e. no added nitrogen or CO<sub>2</sub>). The Cedar Creek experiment was also conducted at Cedar Creek Natural History Area (for details see Tilman *et al.* 1997). We analysed data from the 1997 growing season (CC) and the average of data from the 2001–2002 growing season (CC2). CC2 data were averaged over 2 years to reduce subsampling variance (Lambers *et al.* 2004). The BIODEPTH experiment was conducted at eight different sites across Europe (for details see Hector *et al.* 1999; Spehn *et al.* 2005). We analysed data from the second year of the experiment at each site. Note that relationships between *a priori* FGR and ecosystem functioning may differ from values previously reported because: our analyses use data from different years than previously reported analyses, in some cases the data sets provided contained different numbers of plots than in

**Table 1** The relationship between functional diversity and ecosystem functioning in grassland studies

Site	Species	Plots	Functional classification	Ecosystem function	$R^2$ ( $F$ vs. EF) <i>a priori</i>	<i>A priori</i> percentile
BioCON	16	74	C <sub>3</sub> , C <sub>4</sub> , F, L	Above	0.043	48.2
				Below	0.001	19.1
				Soil N	<b>0.114</b>	<b>99.3</b>
Cedar Creek (1997)	18	163	C <sub>3</sub> , C <sub>4</sub> , F, L, W	Above	<b>0.081</b>	40.7
				Below	<b>0.095</b>	<b>99.9</b>
				Soil N	<b>0.050</b>	87.3
Cedar Creek 2 (2001–2002)	18	163	C <sub>3</sub> , C <sub>4</sub> , F, L, W	Above	<b>0.359</b>	43.4
				Below	<b>0.386</b>	90.2
				Soil N	0.016	<b>96.4</b>
Germany	31	60	G, F, L	Above	<b>0.504</b>	<b>99.9</b>
				Below	0.016	<b>100</b>
Portugal	14	41	G, F, L	Above	<b>0.143</b>	40.8
				Below	<b>0.120</b>	0.4
Switzerland	47	64	G, F, L	Above	<b>0.577</b>	<b>100</b>
				Below	0.006	89.2
Greece	23	52	G, F, L	Above	0.002	66.5
				Below	0.001	54.8
Ireland	12	70	G, F, L	Above	<b>0.134</b>	4.6
				Below	0.038	<b>99.5</b>
Sweden	12	54	G, F, L	Above	<b>0.269</b>	44.0
				Below	<b>0.095</b>	2.6
Gr. Britain A (Sheffield)	12	54	G, F, L	Above	<b>0.432</b>	7.7
				Below	<b>0.124</b>	70.6
Gr. Britain B (Silwood Park)	34	66	G, F, L	Above	<b>0.181</b>	93.6
				Below	<b>0.178</b>	<b>100</b>

Functional Classification lists the functional groups used in the *a priori* functional classification (C<sub>3</sub> = C<sub>3</sub> grass, C<sub>4</sub> = C<sub>4</sub> grass, F = forb, G = grass, N = legume, and W = woody). Ecosystem functions are peak above-ground biomass (Above), peak below-ground biomass (Below), and soil nitrogen concentrations (Soil N).  $R^2$  values for relationships between *a priori* functional group richness ( $F$ ) and ecosystem functioning (EF) that are significant at  $p < 0.05$  are listed in bold. *A priori* percentile indicates where the  $R^2$  from the *a priori* relationship falls within the distribution of  $R^2$  values obtained using random classification, with values greater than 95% indicated in bold.

earlier analyses, plots with no species planted were excluded from our analysis, and, for consistency, only linear regressions were used in our analyses while some of the previously published studies log-transformed FGR (Tilman 1997).

We developed *post hoc* classifications using Monte Carlo methods to randomly classify species from these experiments at values for  $F_{\max}$  (i.e. number of groups into which the species are divided) for each site from 2 to  $S - 1$ , where  $S$  is the number of species employed in each experiment. Values of 1 and of  $S$  were not tested because they correspond to the intercept-only or a species richness model, respectively. There are justifications for comparing the *a priori* classifications only to *post hoc* classifications with the same  $F_{\max}$  (Petchey & Gaston 2002). However, we were explicitly testing whether the *a priori* classifications could outperform any other functional classification scheme, as we felt the assumption that the optimum classification scheme contains three or four groups was unsupported. Clearly, our results will depend on the selection of our particular null model (Gotelli & Graves 1996). However, in 17 out of the 25 comparisons (data not shown), the highest performing *post hoc* classification had an  $F_{\max}$  below 4. Therefore, the bias of our particular null model, if any, was to support the *a priori* classification scheme by including comparisons with *post hoc* classifications with a high  $F_{\max}$ . For each site we constructed 50 000 randomized classifications. In each classification, one species, chosen at random, was assigned to each functional group from 1 to  $F_{\max}$ . Any remaining species were assigned to groups at random. To ensure that all possible levels of  $F_{\max}$  were sampled, we first constructed three random classifications for each level of  $F_{\max}$  between 2 and  $S - 1$ . For all remaining classifications the level of  $F_{\max}$  was determined randomly such that the probability of any given level of  $F_{\max}$  equalled the proportion of the number of combinations with that level of  $F_{\max}$  within the population of all possible combinations. After creating each classification, we counted the number of groups present in each plot and performed ordinary least-squares regression between the number of groups and the level of ecosystem functioning measured in each experimental plot. The precision of *post hoc* classification schemes was estimated as the  $R^2$  of the linear regression between FGR (the number of groups into which the *post hoc* scheme classified the species present in each experimental plot) and ecosystem functioning measurements for each experimental plot. Repeated runs of 50 000 iterations yielded similar distributions of  $R^2$  values, suggesting that this level of replication is sufficient to estimate the distribution of the entire population of classifications.

We determined the relative precision of the *a priori* scheme by comparing the  $R^2$  of the regression in the original study to the distribution of  $R^2$  values generated by the

*post hoc* classifications. We examined three commonly assessed ecosystem functions: (1) above-ground plant biomass, (2) below-ground plant biomass, and (3) soil nitrogen concentrations.

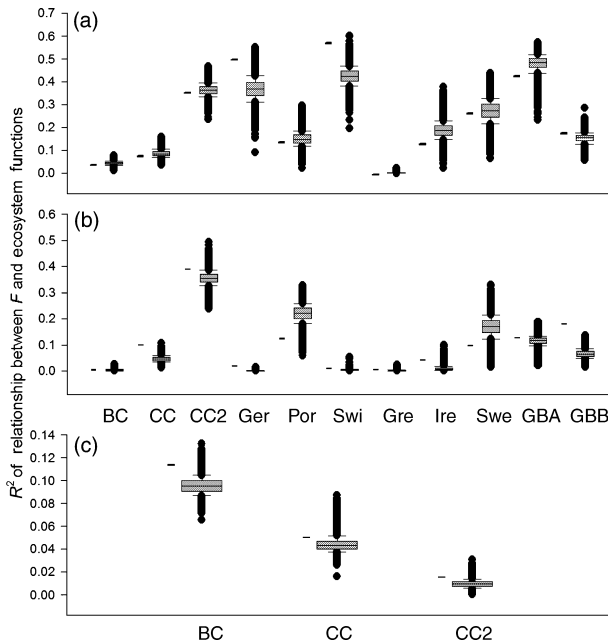
To assess the degree to which the top-performing *post hoc* classification schemes for each ecosystem function at each site grouped species together, we calculated a similarity index derived from Jaccard's similarity index for pairs of classification schemes (Magurran 1988): Similarity =  $j/(a + b - j)$ , where  $j$  is the number of species pairs that are classified together in both classification schemes,  $a$  is the number of species pairs classified together in Classification 1 and  $b$  is the number of species pairs classified together in Classification 2. This similarity score ranges from 0 when the two classification schemes do not group any species pairs in the same manner to 1 when all of the species pairs grouped together in one classification are also grouped together in the other classification. We arbitrarily selected the 50 classification schemes with the highest  $R^2$  between FGR and ecosystem function, representing the top 0.01% of all *post hoc* classification schemes and calculated the similarity index for each pair-wise comparison within this group. We then selected a random set of 50 from the remaining 49 950 classifications and calculated all pair-wise similarity scores as a null comparison. The Jaccard's similarity index is biased by the number of species in a community (in this case, the number of species in an experiment) (Magurran 1988), so to standardize the similarity of classification schemes between experiments with different numbers of species, we calculated an Associative Similarity Index ( $\text{Sim}_{\text{TOP}} - \text{Sim}_{\text{RAN}})/\text{Sim}_{\text{RAN}}$ , where  $\text{Sim}_{\text{TOP}}$  is the mean pair-wise similarity of the top 50 classification schemes for a given site and ecosystem function and  $\text{Sim}_{\text{RAN}}$  is the mean pair-wise similarity of the randomly selected 50 classification schemes for the same site and ecosystem function. This index provides an estimate of the degree to which the top 50 classifications are more similar to each other in the manner in which they grouped species together than are 50 random classifications, e.g. an ASI of 1 indicates that the average similarity of pairs of top classifications is 100% greater than the average similarity of pairs of randomly selected classifications.

If the most precise classification schemes for different ecosystem functions at a given site grouped species together in a similar fashion, this would provide evidence for the existence of a single classification scheme that works equally well for all ecosystem functions at that site. To test this hypothesis, we used the Associative Similarity Index described above to calculate the mean of all pair-wise similarity scores between the top 50 classification schemes from two different ecosystem functions at a given site and the mean of all pair-wise comparisons of 50 random classification schemes from each ecosystem function. We

then calculated the Associative Similarity Index for each combination of two ecosystem functions at each site as described above. Because no sites used the same set of species, we were unable to test the similarity of functional classifications across sites.

## RESULTS

For 11 of the 25 cases, the *a priori* GFL functional scheme had a level of precision that was at or below the median for the *post hoc* classification schemes (Fig. 1, Table 1). Although at most sites (17 of 25), the relationships between  $F$  and ecosystem functioning using the *a priori* GFL functional classification were statistically significant, these relationships were typically weak, with the  $R^2$  of the regression exceeding 0.2 in only six instances. Furthermore, in only five instances did the  $R^2$  of statistically significant relationships between



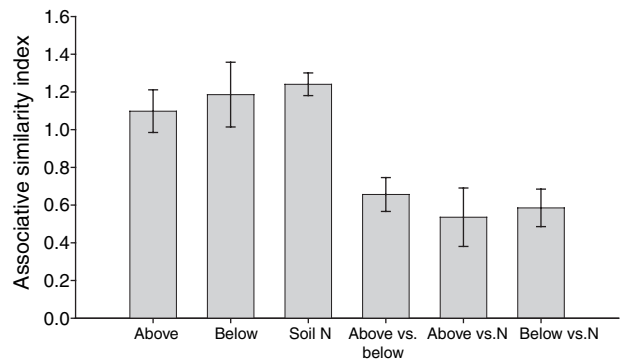
**Figure 1** The distribution of  $R^2$  values between functional group richness and (a) above-ground biomass, (b) below-ground biomass, and (c) soil nitrogen concentrations using random classifications at 10 experiments (BC = BioCON, CC = Cedar Creek 1997, CC2 = Cedar Creek 2001–2002, Ger = Germany, Por = Portugal, Swi = Switzerland, Gre = Greece, Ire = Ireland, Swe = Sweden, GBA = Sheffield, GBB = Silwood Park). The middle bar of each box represents the median values of the distribution, the upper and lower edges of each box, the 75th and 25th percentile respectively, the whiskers the 90th and 10th percentile, and additional dots, outliers. The  $R^2$  value of the relationship between functional group richness and ecosystem functioning using *a priori* GFL functional classifications at each site is indicated by the dash.

FGR and ecosystem functioning using the *a priori* GFL classification fall above the 95th percentile of the distribution of  $R^2$  values generated using *post hoc* classifications.

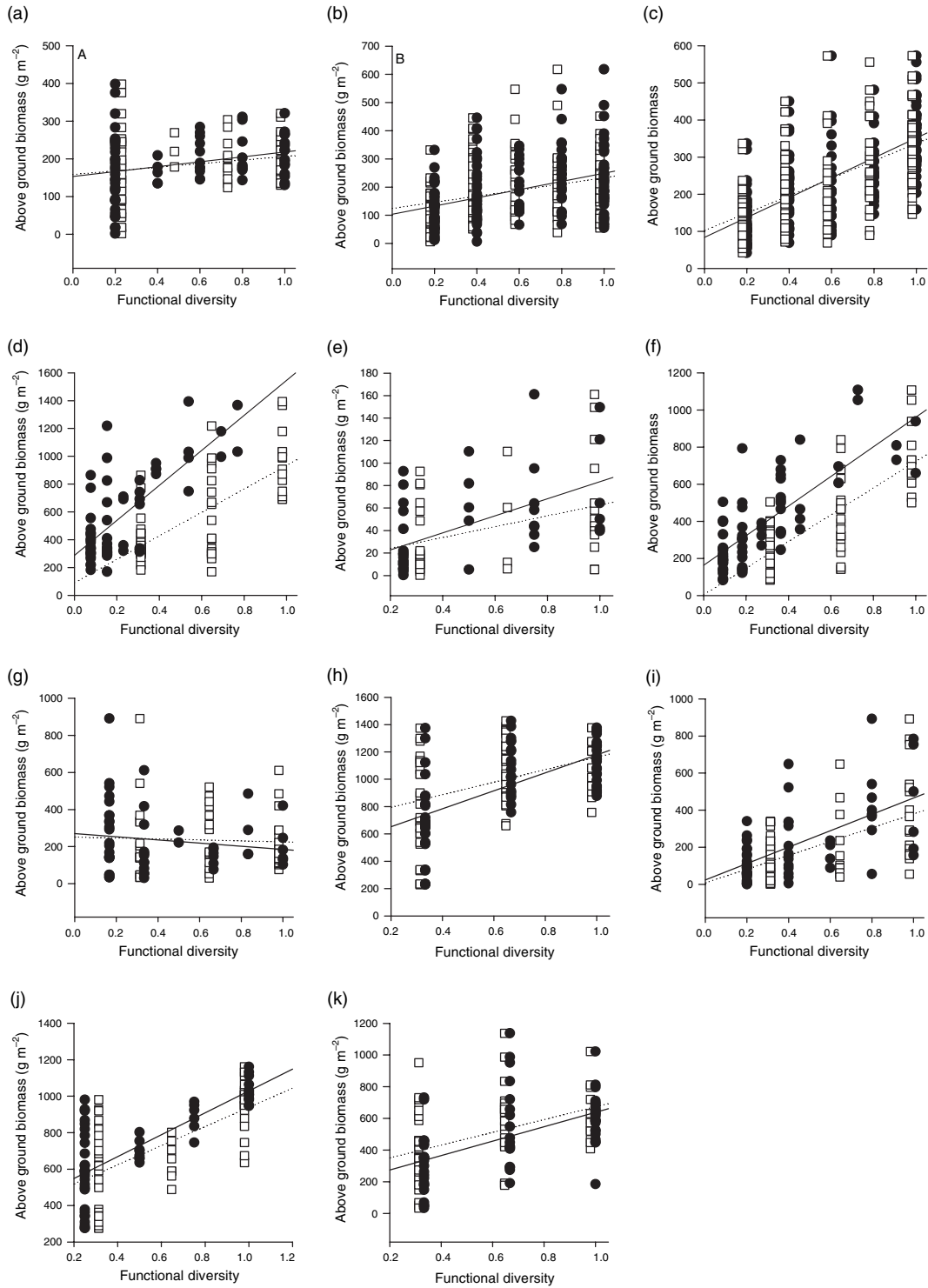
The *a priori* GFL classification scheme performed best when predicting below-ground biomass, outperforming *post hoc* classification schemes at four of the sites. At one site (Germany), the *a priori* GFL functional classification scheme outperformed random classifications for both above- and below-ground biomass. However, across all sites, there was no correlation between the relative success of *a priori* functional groups in predicting above- and below-ground productivity at a site (Pearson correlation coefficient = 0.224,  $n = 11$ ,  $p = 0.51$ ).

We analysed data from the Cedar Creek experiment at two different time points (1997 and 2001–2002), and although the precision of the *a priori* GFL classification scheme increased dramatically over time for both above- and below-ground biomass, the rank of the *a priori* GFL classification scheme relative to the random classifications remained similar.

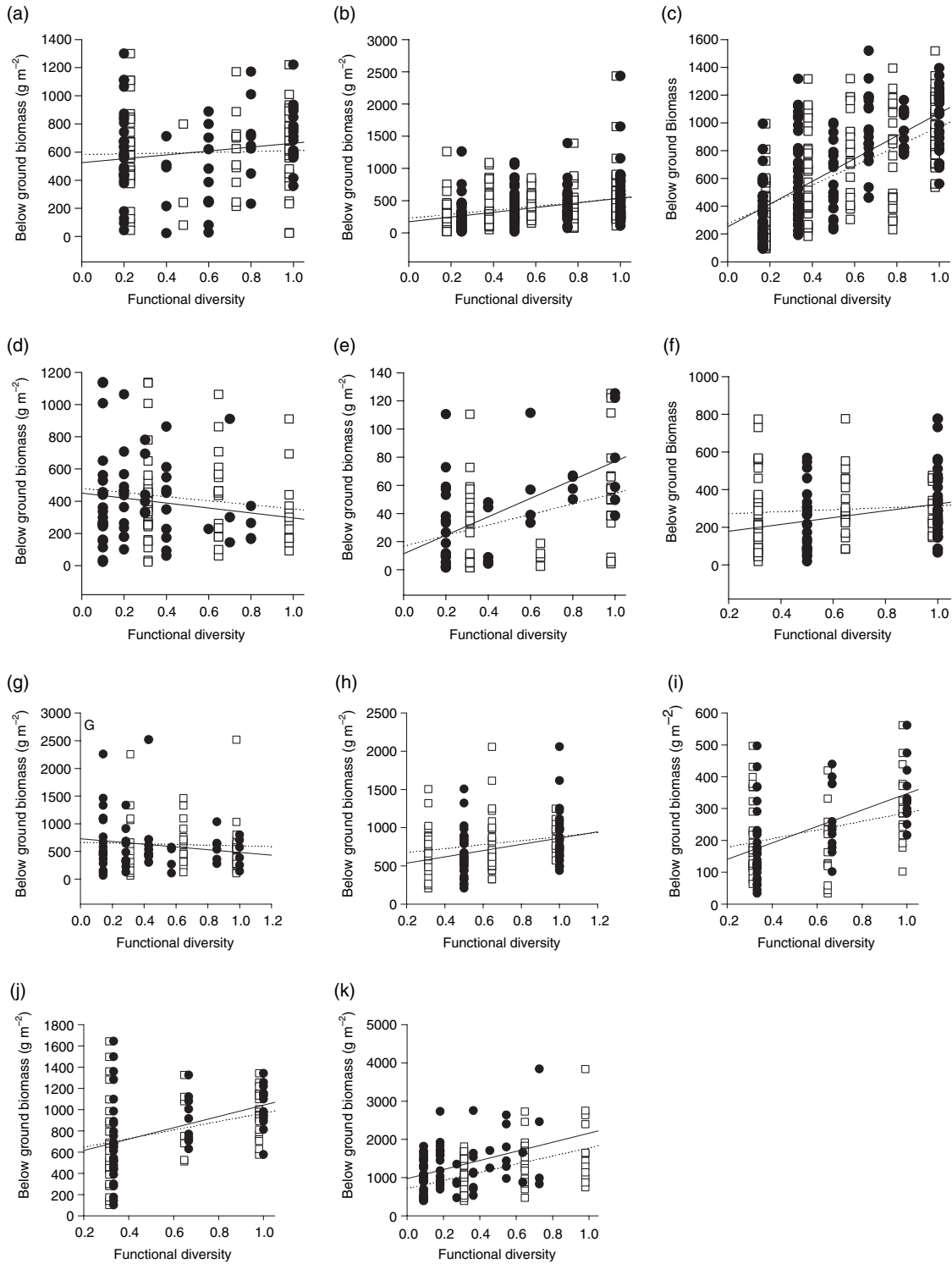
The average pair-wise similarity of the 50 top-performing classification schemes created by randomization for a given ecosystem function was generally low, ranging from 0.043 (Switzerland, above-ground biomass) to 0.242 (Sweden, below-ground biomass). However, across all sites, the average similarity of the top 50 classifications was approximately double the similarity between 50 classifications drawn at random from the remaining 49 950 classifications (Fig. 2). Overall, at a given site, the average pair-wise similarity of 50 top-performing classification schemes for



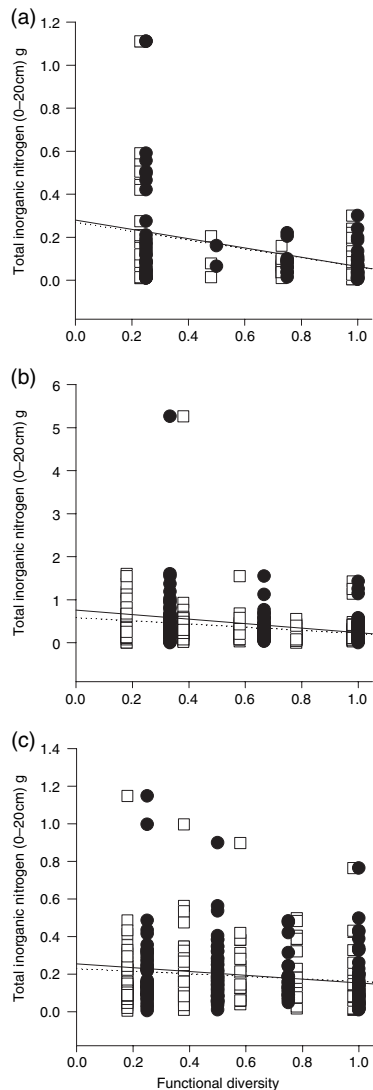
**Figure 2** Similarity of the top 50 *post hoc* classifications in grouping species together relative to a random selection of 50 *post hoc* classifications within and between ecosystem functions. An Associative Similarity Index (see Methods for details of calculation) of 0 represents no difference in mean pair-wise similarity between the top 50 classifications and the randomly selected 50 classifications while a score of 1 represents a 100% increase in mean pair-wise similarity between the top 50 classifications and the randomly selected 50 classifications. Mean ASI are shown with 1 standard error.



**Figure 3** The relationship between functional diversity as estimated using *a priori* GFL classification scheme (open squares) and the best performing *post hoc* classification scheme (closed circles) and above-ground biomass for all sites. Functional diversity is the fraction of all possible functional groups planted in each plot. Note that functional diversity values for the *a priori* classification have been slightly offset for clarity. Best-fit linear relationships between functional diversity and ecosystem functions using *a priori* classification scheme (dotted line) and *post hoc* classification scheme (solid line) are also shown. (a) BioCON, (b) Cedar Creek 1997, (c) Cedar Creek 2001–2002, (d) Germany, (e) Portugal, (f) Switzerland, (g) Greece, (h) Ireland, (i) Sweden, (j) Sheffield, (k) Silwood Park.



**Figure 4** The relationship between functional diversity as estimated using *a priori* GFL classification scheme (open squares) and the best performing *post hoc* classification scheme (closed circles) and below-ground biomass for all sites. Functional diversity is the fraction of all possible functional groups planted in each plot. Note that functional diversity values for the *a priori* classification have been slightly offset for clarity. Best-fit linear relationships between functional diversity and ecosystem functions using *a priori* classification scheme (dotted line) and *post hoc* classification scheme (solid line) are also shown. (a) BioCON, (b) Cedar Creek 1997, (c) Cedar Creek 2001–2002, (d) Germany, (e) Portugal, (f) Switzerland, (g) Greece, (h) Ireland, (i) Sweden, (j) Sheffield, (k) Silwood Park.



**Figure 5** The relationship between functional diversity as estimated using *a priori* GFL classification scheme (open squares) and the best performing *post hoc* classification scheme (closed circles) and soil nitrogen for all sites. Functional diversity is the fraction of all possible functional groups planted in each plot. Note that functional diversity values for the *a priori* classification have been slightly offset for clarity. Best-fit linear relationships between functional diversity and ecosystem functions using *a priori* classification scheme (dotted line) and *post hoc* classification scheme (solid line) are also shown. (a) BioCON, (b) Cedar Creek 1997, (c) Cedar Creek 2001–2002.

different ecosystem functions was lower than the similarity of the top 50 classification schemes within an ecosystem function, ranging from 0.048 (Switzerland above-ground biomass vs. below-ground biomass) to 0.18 (Ireland, above-ground biomass vs. below-ground biomass). However, the average similarity of the top 50 classifications across ecosystem functions was still higher than the similarity of 50 random classifications across functions (Fig. 2).

In general, the ability of the best *post-hoc* classification to predict ecosystem functioning was higher when predicting above-ground biomass than when predicting below-ground biomass and soil nitrogen (Figs 3–5).

## DISCUSSION

These findings show that for key ecosystem functions many possible functional classification schemes exist that can potentially predict ecosystem response to changes in biodiversity and many of these have greater explanatory power than the *a priori* GFL classification commonly in use. In spite of its wide use, due to the appeal of its biological foundation, the GFL classification scheme often has low explanatory power, and may, in many instances, be no more effective than classifying species into completely random groups. Alternative schemes for assessing functional diversity based explicitly on ecophysiological and morphological traits of species (Craine *et al.* 2002; Petchey & Gaston 2002; Reich *et al.* 2003; Mouillot *et al.* 2005), while potentially more system-specific (but see Diaz *et al.* 2004; Wright *et al.* 2004) and dependent on initial trait selection and weighting, may capture more of the functional variation that leads to diversity effects than traditional functional classifications (Petchey 2004). The *post hoc* methods outlined here are another potential solution to the challenge of identifying the functional groups responsible for maintaining ecosystem functioning. However, it is important to note that *post hoc* methods identify best-fit functional groups for existing experiments. As such, to avoid risks of non-independence or circularity, they require independent empirical tests that can confirm the efficacy of such groupings. In addition, although the *post hoc* method can identify groupings that perform well in predicting ecosystem functions, without further analysis, it still does not reveal which functional traits are important for establishing a mechanistic understanding of the biotic controls on the ecosystem function of interest.

The ability of the optimal *post hoc* classification schemes to predict ecosystem functioning was much higher for above-ground biomass than for either below-ground biomass or nitrogen retention (as estimated by soil nitrogen concentrations). It is possible that this result reflects a flaw in theory that predicts that functional differentiation should be equally important in regulating primary production and nutrient cycling. However, it seems more likely that in annual grasslands, peak above-ground biomass represents an adequate surrogate for primary productivity, while below-ground biomass and soil nitrogen concentrations are poor surrogates for primary productivity and nutrient retention respectively, because of year-to-year carryover of below-ground biomass and rapid nitrogen cycling in these frequently nitrogen limited systems. It is also possible that the low degree of explanatory power in some of our relationships was



due to significant non-linearity in the response of ecosystem functioning to changes in functional diversity. However, such non-linearities were not readily observed.

Furthermore, while for a given ecosystem function at a particular site, the top-performing *post hoc* classifications did group species together in a more similar fashion than a random selection of *post hoc* classification schemes, the overall similarity was quite low. This suggests that there are multiple ways to group species that still result in a relatively high ability to explain the covariance between FGR and ecosystem functioning. The lack of a single 'optimal' classification scheme could be because there are only a few species that are driving ecosystem function (Grime 1998) and as long as they are classified appropriately, the classification of the remaining species is irrelevant. Alternatively, this result could be due to the fact that because traits are distributed in a continuous fashion within an assemblage or there is little correlation between traits, there is no single optimal classification. If this is the case, functional diversity will be more appropriately characterized by a multivariate index of what volume of 'trait-space' a group of species occupies (Walker *et al.* 1999; Petchey & Gaston 2002; Mouillot *et al.* 2005) than a simple count of arbitrary functional groups. Without further data on the distribution of traits among species and a better understanding of which traits affect ecosystem functioning, it is difficult to distinguish between these two hypotheses.

Given that top-performing *post hoc* classification schemes were less similar between than within ecosystem functions, the existence of a 'universal' classification scheme, i.e. one that performs equally well in capturing the effects of functional diversity on multiple ecosystem functions, seems unlikely. Different ecosystem functions are quite likely to be strongly affected by different traits, and the form of the relationship between functional diversity and ecosystem functioning can vary across ecosystem functions, and ecosystems. For example, nitrogen retention will be driven primarily by traits that affect nitrogen uptake (e.g. tissue C:N, root distribution, symbiotic nitrogen fixation) and decomposition, while productivity will be driven by traits related to carbon acquisition (e.g. photosynthetic rate, water use efficiency, plant architecture). While some of these traits are likely to be correlated, representing differing resource acquisition strategies (Wright *et al.* 2004), large databases analysing the distribution of traits among species tend to find an even distribution of species in trait-space rather than the distinct clustering that would indicate unique resource acquisition strategies (Craine *et al.* 2001; Diaz *et al.* 2004). To the extent that the top-performing *post hoc* classification schemes for predicting different ecosystem functions at a site are more similar to each other than are random classifications, one could argue that classifications that work well for one ecosystem function work well for another.

However, given the low overall similarity, it is likely to be more profitable to investigate exactly which plant traits are responsible for particular ecosystem functions and to only include relevant traits in future calculations of functional diversity.

Collectively, our findings suggest that the use of the *a priori* GFL classification by current studies has underestimated the role of functional diversity for two reasons. First, the *a priori* GFL functional classification has often been an ineffective scheme compared to alternatives identified by the Monte Carlo *post hoc* method. Second, given the analysis of the similarity of classification schemes both within and between ecosystem functions at a site, it appears that there is unlikely to be any single grouping of species that accurately captures the functional diversity of an assemblage. Given global declines in biodiversity due to habitat transformation, biological invasions, and overexploitation (Wilcove *et al.* 1998), understanding the ecosystem consequences of such widespread change remains one of the major challenges of contemporary ecological research (Loreau *et al.* 2001, 2002). To the extent that biodiversity loss involves losses in functional diversity, identification of effective measures of functional diversity are necessary to predict changes in ecosystem functions and the services derived from them (Daily *et al.* 1997).

## ACKNOWLEDGEMENTS

The authors would like to thank the many people responsible for conducting the experiments from which these data were collected. This manuscript was improved by comments from E. Bernhardt, D. Hooper, O. Petchey, and an anonymous referee. This work was supported by NSF DEB 0130289 and is a product of the BioMERGE network.

## REFERENCES

- Chapin, F.S., III, Reynolds, H.L., D'Antonio, C.M. & Eckhart, V.M. (1996). The functional role of species in terrestrial ecosystems. In: *Global Change and Terrestrial Ecosystems* (eds Walker, B. & Steffen, W.) Cambridge University Press, Cambridge, pp. 403–430.
- Craine, J., Froehle, J., Tilman, D., Wedin, D. & Chapin, F.S., III (2001). The relationship among root and leaf traits of 76 grassland species and relative abundance along fertility and disturbance gradients. *Oikos*, 93, 274–285.
- Craine, J.M., Tilman, D., Wedin, D., Reich, P., Tjolkner, M. & Knops, J. (2002). Functional traits, productivity and effects on nitrogen cycling of 33 grassland species. *Funct. Ecol.*, 16, 563–574.
- Daily, G., Alexander, S., Ehrlich, P.R., Gouler, L., Lubchenco, J., Matson, P.A. *et al.* (1997). Ecosystem services: benefits supplied to human societies by natural ecosystems. *Issues Ecol.*, 2, 1–18.
- Diaz, S. & Cabido, M. (2001). Vive la difference: plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.*, 16, 646–655.

- Diaz, S., Hodgson, J.G., Thompson, K., Cabido, M., Cornelissen, J.H.C., Jalili, A. *et al.* (2004). The plant traits that drive ecosystems: evidence from three continents. *J. Veg. Sci.*, 15, 295–304.
- Gotelli, N.J. & Graves, G.R. (1996). *Null Models in Ecology*. Smithsonian Institution Press, Washington.
- Grime, J.P. (1998). Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.*, 86, 902–910.
- Grime, J.P., Thompson, K., Hunt, R., Hodgson, J.G., Cornelissen, J.H.C., Rorison, I.H. *et al.* (1997). Integrated screening validates primary axes of specialisation in plants. *Oikos*, 79, 259–281.
- Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M.C., Diemer, M., Dimitrakopoulos, P.G. *et al.* (1999). Plant diversity and productivity experiments in European grasslands. *Science*, 286, 1123–1127.
- Hooper, D.U. & Dukes, J.S. (2004). Overyielding among plant functional groups in a long-term experiment. *Ecol. Lett.*, 7, 95–105.
- Hooper, D.U. & Vitousek, P.M. (1997). The effects of plant composition and diversity on ecosystem processes. *Science*, 277, 1302–1305.
- Hooper, D.U., Solan, M., Symstad, A., Diaz, S., Gessner, M.A., Buchman, N. *et al.* (2002). Species diversity, functional diversity, and ecosystem functioning. In: *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (eds Loreau, M., Naeem, S. & Inchausti, P.). pp. 195–208. Oxford University Press, New York.
- Hooper, D.U., Chapin, F.S. III, Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S. *et al.* (2005). Effects of biodiversity on ecosystem functioning: a consensus of current knowledge and needs for future research. *Ecol. Monogr.*, 75, 3–36.
- Lambers, J.H.R., Harpole, W.S., Tilman, D., Knops, J. & Reich, P.B. (2004). Mechanisms responsible for the positive diversity–productivity relationship in Minnesota grasslands. *Ecol. Lett.*, 7, 661–668.
- Lavorel, S. & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Funct. Ecol.*, 16, 545–556.
- Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A. *et al.* (2001). Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, 294, 804–808.
- Loreau, M., Naeem, S. & Inchausti, P. (2002). *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives*. Oxford University Press, Oxford.
- Magurran, A.E. (1988). *Ecological Diversity and Its Measurement*. Princeton University Press, Princeton, NJ.
- Mouillot, D., Mason, W.H.N., Dumay, O. & Wilson, J.B. (2005). Functional regularity: a neglected aspect of functional diversity. *Oecologia*, 142, 353–359.
- Naeem, S. & Wright, J.P. (2003). Disentangling biodiversity effects on ecosystem functioning: deriving solutions to a seemingly insurmountable problem. *Ecol. Lett.*, 6, 567–579.
- Naeem, S., Byers, D., Tjossem, S.F., Bristow, C. & Li, S. (1999). Plant neighborhood diversity and production. *Ecoscience*, 6, 355–365.
- Petchey, O.L. (2004). On the statistical significance of functional diversity effects. *Funct. Ecol.*, 18, 297–303.
- Petchey, O.L. & Gaston, K.J. (2002). Functional diversity (FD), species richness and community composition. *Ecol. Lett.*, 5, 402–411.
- Reich, P., Knops, J.M.H., Tilman, D., Craine, J., Ellsworth, D., Tjoelker, M. *et al.* (2001). Plant diversity enhances ecosystem responses to elevated CO<sub>2</sub> and nitrogen deposition. *Nature*, 410, 809–812.
- Reich, P.B., Buschena, C., Tjoelker, M.G., Wrage, K., Knops, J., Tilman, D. *et al.* (2003). Variation in growth rate and ecophysiology among 34 grassland and savanna species under contrasting N supply: a test of functional group differences. *New Phytol.*, 157, 617–631.
- Reich, P., Tilman, D., Naeem, S., Ellsworth, D., Knops, J., Craine, J. *et al.* (2004). Species and functional group diversity independently influence biomass accumulation and its response to CO<sub>2</sub> and N. *Proc. Natl. Acad. Sci. USA*, 101, 10101–10106.
- Roscher, C., Schumacher, J., Baade, J., Wilcke, W., Gleixner, G., Weisser, W.W. *et al.* (2004). The role of biodiversity for element cycling and trophic interactions: an experimental approach in a grassland community. *Basic Appl. Ecol.*, 5, 107–121.
- Spehn, E.M., Hector, A., Joshi, J., Scherer-Lorenzen, M., Schmid, B., Bazeley-White, E. *et al.* (2005). Ecosystem effects of the manipulation of plant diversity in European grasslands. *Ecol. Monogr.*, 75, 37–63.
- Symstad, A.J. (2000). A test of the effects of functional group richness and composition on grassland invasibility. *Ecology*, 81, 99–109.
- Tilman, D. (1997). Community invasibility, recruitment limitation, and grassland biodiversity. *Ecology*, 78, 81–92.
- Tilman, D., Knops, J.M.H., Wedin, D., Reich, P., Ritchie, M. & Siemann, E. (1997). The influence of functional diversity and composition on ecosystem processes. *Science*, 277, 1300–1302.
- Vitousek, P.M. & Hooper, D.U. (1993). Biological diversity and terrestrial ecosystem biogeochemistry. In: *Biodiversity and Ecosystem Function* (eds Schulze, E.D. & Mooney, H.A.). pp. 3–14. Springer-Verlag, Berlin.
- Walker, B., Kinzig, A. & Langridge, J. (1999). Plant attribute diversity, resilience, and ecosystem function: the nature and significance of dominant and minor species. *Ecosystems*, 2, 95–113.
- Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. (1998). Quantifying threats to imperiled species in the United States. *Bioscience*, 48, 607–615.
- Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F. *et al.* (2004). The worldwide leaf economics spectrum. *Nature*, 428, 821–827.

Editor, Peter Morin

Manuscript received 27 June 2005

First decision made 5 August 2005

Second decision made 3 October 2005

Manuscript accepted 5 October 2005