

## REPORT

## Overyielding among plant functional groups in a long-term experiment

David U. Hooper<sup>1\*</sup> and Jeffrey S. Dukes<sup>2</sup>

<sup>1</sup>Department of Biology,  
Western Washington University,  
516 High St, Bellingham, WA  
98225-9160, USA

<sup>2</sup>Department of Biology,  
University of Utah, Salt Lake  
City, UT 84112-0840, USA  
Present address: Department  
of Biology, University of  
Massachusetts, Boston, MA  
02125-3393, USA

\*Correspondence: E-mail:  
hooper@biol.wvu.edu

### Abstract

A recent debate among ecologists has focused on mechanisms by which species diversity might affect net primary productivity. Communities with more species could use a greater variety of resource capture characteristics, leading to greater use of limiting resources (complementarity) and therefore greater productivity (overyielding). Recent experiments, however, have shown a variety of relationships between diversity and productivity. In an experiment on serpentine grassland communities spanning 8 years, we found that overyielding increased several years after plot establishment. Overyielding varied greatly depending on the functional characteristics of the species involved and the biotic and abiotic environment (particularly water availability). While functional differences among species led to strong complementarity and facilitation, these effects were not sufficient to cause significant transgressive overyielding or consistent increases in productivity with increased plant diversity. These results suggest that greater absolute production with greater diversity may be restricted to particular species combinations or environmental conditions.

### Keywords

Biodiversity, community composition, complementarity, ecosystem functioning, net primary production, overyielding, plant functional types, resource partitioning, sampling effect.

*Ecology Letters* (2004) 7: 95–105

### INTRODUCTION

Experiments studying how biodiversity affects ecosystem productivity have generated substantial debate about both the patterns observed and the mechanisms underlying these patterns (Loreau *et al.* 2001, 2002). Researchers studying the productivity of plant communities have found a variety of responses to increasing species or functional richness, including saturating increases, no differences and idiosyncratic differences as plant diversity increases (for a recent review; see Schmid *et al.* 2002). While relationships between diversity and productivity could arise through causal mechanisms such as complementary resource use, such patterns could also arise by chance, through the sampling effect (Aarssen 1997; Huston 1997; Tilman *et al.* 1997; Wardle 1999; Loreau 2000). A better understanding of the mechanisms underlying diversity–productivity relationships will shed light on how ecosystem processes and services might or might not change with changes in community composition and diversity (Loreau *et al.* 2001).

Definitions of complementarity have varied, which has led to confusion about whether complementarity is a

property of species, a mechanism leading to increased productivity in mixtures or the phenomenon of increased productivity itself (e.g. Petchey 2003). As have others (e.g. Trenbath 1974; Fridley 2001), we differentiate between overyielding and complementarity. The former refers to production in mixtures that exceeds expectations based on monoculture yields. Overyielding can be either transitive (mixture yields exceed yields from the most productive monocultures) or non-transitive (mixture yields exceed expectations but not absolute yields of the most productive monocultures). We treat complementary resource use and facilitation as two primary mechanisms leading to overyielding (Trenbath 1974; Harper 1977; Ewel 1986; Vandermeer 1990; Loreau 1998). Distinguishing between complementarity and facilitation is difficult in practice, in part because most metrics actually measure overyielding, rather than resource use *per se*. However, the underlying mechanisms can be differentiated using appropriate experiments (Vandermeer 1990).

How many and which species might be complementary and contribute to overyielding in diverse natural and seminatural communities remains unclear (Cardinale *et al.*

2000; Fridley 2001; Mouquet *et al.* 2002). Evidence from competition experiments and intercropping has shown that resource partitioning in space, time and resource type (e.g. ammonium vs. nitrate vs. amino acids) is greatest with species differing strongly in functional type, particularly in relatively simple two- or three-way mixtures (e.g. Harper 1977; Ewel 1986; Vandermeer 1990; McKane *et al.* 2002). However, because plants all need the same suite of basic resources, opportunities for complementarity may be relatively limited. For example, some authors (Huston *et al.* 2000) argue that if overyielding is involved in recent experiments (e.g. Loreau & Hector 2001; Tilman *et al.* 2001), it mainly results from effects of nitrogen-fixing species, a phenomenon already well known from intercropping (e.g. Trenbath 1974; Swift & Anderson 1993). Recent results have demonstrated overyielding without nitrogen fixers, but did not address the particular functional traits involved (van Ruijven & Berendse 2003). What determines whether overyielding occurs in complex mixtures and which species are involved?

Most studies assessing diversity/productivity relationships have been relatively short-term (<3 years, with many for only one growing season). Such experiments have shown evidence for overyielding in some systems, some of the time (Haggard & Ewel 1997; Hooper 1998; Dukes 2001; Loreau & Hector 2001; Engelhardt & Ritchie 2002; Fridley 2002, 2003; Hector *et al.* 2002; van Ruijven & Berendse 2003). However, in Minnesota grasslands (Tilman *et al.* 2001) and Swedish meadows (Mulder *et al.* 2002) the strength of overyielding and species richness effects increased several years following experimental establishment. While these results suggest that short-term experiments may underestimate the strength of complementarity or facilitation, it is not clear how common such patterns are.

Overyielding due to complementarity or facilitation does not necessarily lead to absolute increases in productivity with increasing plant diversity. Most common measures of overyielding for assessing complementarity/facilitation are rightly based on relative measures of production to correct for individual species differences in monoculture yields. However, for diversity to have a positive effect on productivity, it must lead to greater absolute production in mixture, which requires transitive overyielding (Tilman *et al.* 1997; Fridley 2001) – a phenomenon that is infrequently observed (Trenbath 1974; Jolliffe 1997; Hector *et al.* 2002; Fridley 2003; but see Tilman *et al.* 2001). Furthermore, recent experimental results call into question some of the assumptions of early models of the sampling effect: the most productive species in monoculture are not necessarily the most dominant in mixtures (Huston 1997; Tilman *et al.* 1997; Hector *et al.* 2002). These observations raise questions about the mechanisms by which increasing species richness is hypothesized to lead to greater net primary production in ecosystems.

We had three goals in this study. First, we sought to assess how overyielding varies with time, in response to both plot development and interannual variability. Second, we sought to assess which functional groups were responsible for overyielding and why. Third, we sought to better understand the relationship between complementarity and the response of absolute productivity to changing diversity. To address these questions, we investigated overyielding among four plant functional groups (groups of species with similar morphology and phenology) in synthetic serpentine grassland communities in California. Because the functional groups we used differ strongly in a variety of characteristics related to resource acquisition (most notably phenology, rooting depth, canopy architecture and nitrogen-fixation; Gulmon *et al.* 1983; Mooney *et al.* 1986; Hooper 1998), we hypothesized that increasing functional group richness would lead to complementary resource use and greater net primary productivity. In 1993, 2 years after these communities were sown, measurements indicated limited overyielding and no significant relationship of productivity with functional diversity (Hooper & Vitousek 1997; Hooper 1998). We resampled in 1998 and 1999, the seventh and eighth growing seasons since establishment, and once again tested for these relationships.

## METHODS

### Study site

Experimental plots containing representative species from a local serpentine grassland were established in the 1991–1992 growing season in south San Jose, California (Hooper 1998). Species were drawn from four plant functional groups: early-season annuals (E), late-season annuals (L), perennial bunchgrasses (P), and nitrogen-fixing legumes (N). The experimental treatments consisted of a full factorial combination of E, L and P functional groups, plus N-fixers planted alone and with all other groups. This gave a total of 10 treatments, including bare plots, although we will focus only on the vegetated plots here. All treatments were replicated six times in a randomized complete block design. Three species of early season annuals and two species of the other groups were planted wherever those functional groups were used. All treatments were planted with a target biomass of  $200 \text{ g m}^{-2}$ , the average from previous studies at nearby natural serpentine grasslands (see Hooper 1998 for details). Planting densities for single functional group (SFG) plots came from dividing the target biomass by the mean size of individuals and the proportion of seed germination for each species, aiming for equal proportions by biomass for each species. For mixtures, SFG planting densities were divided by the number of functional groups in the mixture, per replacement series design (Harper 1977). All species

established self-sustaining populations; in only a few plots did we not find all planted species. Densities of individual species were not manipulated after planting, and by 1998 they reflected population adjustments to resource availability and any ecological sorting that had taken place (Fig. 1). Analysis of overyielding by standard replacement series indices should therefore be robust and not dependent on the chosen planting density (Jolliffe 2000). Plots were weeded regularly to maintain composition, although average biomass of removed weeds never exceeded 5% of total biomass for any of the treatments, except the N treatment (see also Fig. 1). Therefore, weed removal had a very minor influence on total plot biomass or the amount of bare ground. (See Hooper (1998) for more details of experimental set-up.)

We measured aboveground net primary productivity (ANPP) in 1993, 1998 and 1999, the second, seventh and eighth years after plot establishment. We measured aboveground biomass at three times of the year: mid-April, mid-May and mid-August (peak biomass of the Es and Ns, Ps and Ls, respectively), as in Hooper (1998). None of the harvested areas were used again in subsequent harvests, and all reseeded naturally. As all aboveground tissue is new each year, ANPP is the sum of the peak biomasses of all species across sampling dates. In this highly seasonal system, this sampling strategy gives a much better estimate of total plot production than a one-time harvest.

Effects of functional group richness on ANPP were assessed by simple linear regression. As the experimental design led to unequal variances at different levels of functional group richness, the assumptions of linear regression, however, did not apply. Therefore, we used randomization analyses to compare the regression slope based on the productivity and functional group richness values of each plot with regression slopes created by randomly assigning each of the observed productivity values to each of the values for functional group richness (Dukes 2001). Actual slopes were compared with 999 randomly created slopes.

### Overyielding calculations

We selected several indices to address particular aspects of overyielding. We explain these indices in terms of species productivity (the most common usage), although we calculated all of them based on functional group productivity. We used monoculture productivity within blocks to generate expected values, as in Hooper (1998), rather than using mean monoculture values across blocks. To best reflect total resource use, weeds were included in the calculation of all indices when they could reasonably be classified into the relevant functional groups. Weeds not included in the calculations averaged at most 2–3% of total biomass in the mixture treatments in all years (Fig. 1).

Relative yield totals (RYT) measure overyielding by summing the relative yields (RY) for all species in a mixture:

$$\text{RYT} = \sum_{i=1}^s \text{RY}_i,$$

where  $s$  is the total number of species.  $\text{RY}_i = O_i/M_i$ , where  $O_i$  is the observed mixture yield of species  $i$  and  $M_i$  is the monoculture yield of species  $i$ .  $\text{RYT} > 1$  indicates overyielding. RYT is one of the most common metrics for assessing overyielding (see Fridley 2001 and references therein) and it is robust, as long as planting densities give constant final yield, sufficient time has been allowed for individuals to mature and community interactions to develop, and indices are calculated on a yield per area basis rather than as yield per individual (Hooper 1998; Jolliffe 2000). These criteria were met for all data in this study.

$D_{\max}$  assesses the degree to which transgressive overyielding occurs:

$$D_{\max} = \frac{O_T - \max(M_i)}{\max(M_i)},$$

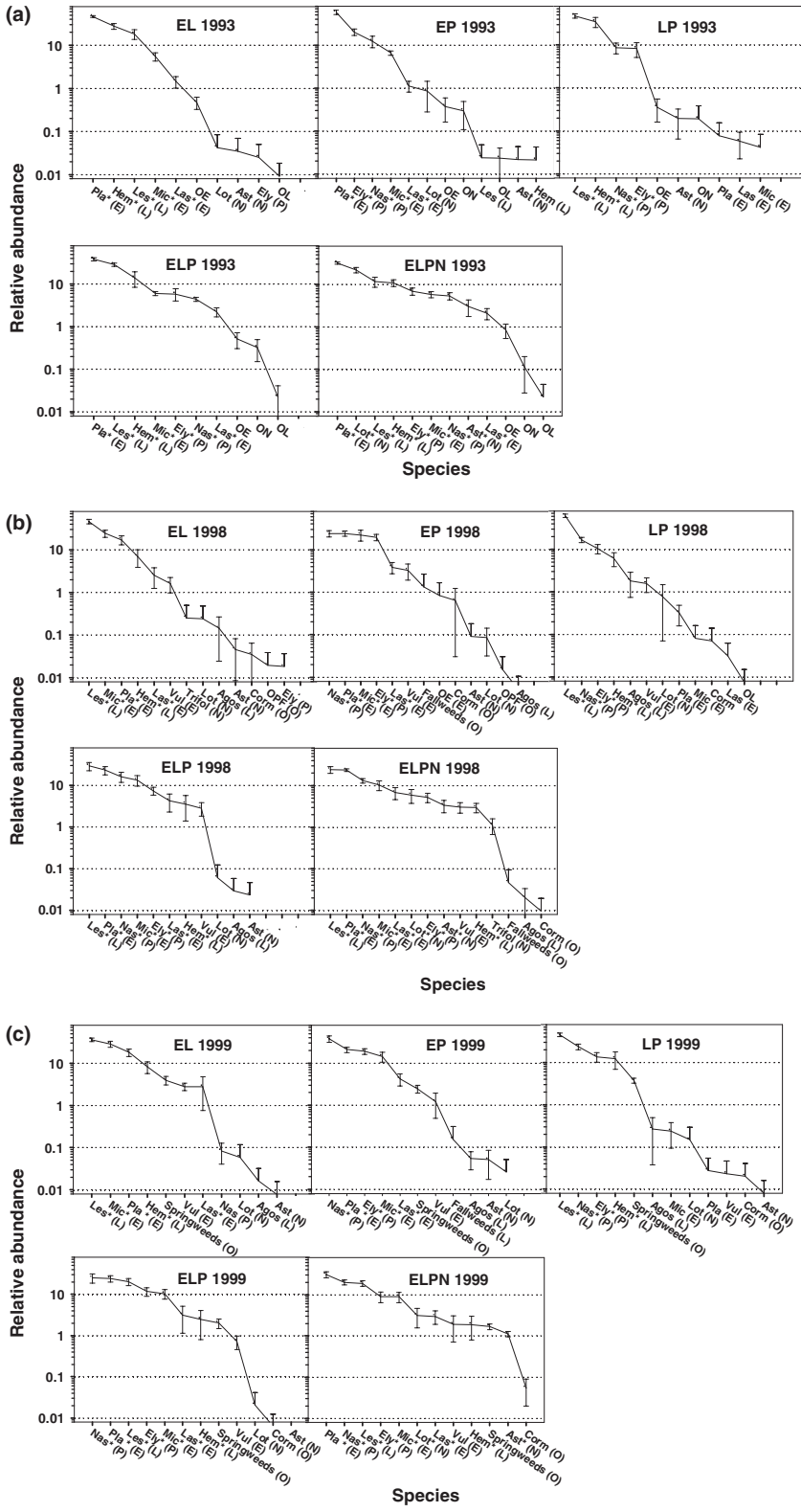
where  $O_T$  is the observed total yield of a given mixture, and  $\max(M_i)$  is the maximum monoculture yield of the species found in that mixture.  $D_{\max} > 0$  indicates transgressive overyielding (Loreau 1998).  $D_{\max}$  is an appropriate measure if one is interested not just in whether complementarity is occurring, but also if absolute yields in mixture are greater than for monocultures. It has been proposed as the 'acid test' for complementarity or facilitation, because the sampling effect alone cannot give significantly positive  $D_{\max}$  (Huston *et al.* 2000; Tilman *et al.* 2001).

We assessed the performance of individual functional groups in each mixture using  $D_i$ :

$$D_i = \frac{O_i - E_i}{E_i}.$$

$E_i$  is the expected yield of the individual species:  $E_i = M_i/S$ .  $D_i$  is positive when mixture yields for a given species are greater than expectations from monocultures.  $D_i > 0$  for all species indicates overyielding (Loreau 1998). Calculation of  $D_i$  helps to interpret RYT because different types and magnitudes of species interactions could lead to similar values of RYT (Jolliffe 2000).  $D_i$  gives the same information as RY of the individual species, but  $D_i$  is easier to interpret because the expected value (0) does not change with the number of species in mixture, as with RY.

We used the additive partitioning equation (APE) of Loreau & Hector (2001) to separate the productivity responses in mixtures into the complementarity effect (CE) and the selection effect (SE). CE gives the same information as RYT, although both CE and SE are in actual units of yield (Loreau & Hector 2001). The net difference in



**Figure 1** Dominance-diversity curves for mixture plots in (a) 1993, (b) 1998 and (c) 1999. Points show mean ( $\pm$ SE;  $n = 6$ ) for each treatment type. Species abbreviations are shown on the X-axis. An asterisk after the species abbreviation indicates that it was a planted species (i.e. not a weed) in that treatment type and letters in parentheses indicate functional group (E, L, P, N, or O for 'other') for both planted species and weeds. Abbreviations for species are as follows: Agos, *Agoseris heterophylla*; Ast, *Astragalus gambellianus*; Corm, bulb-forming species (*Brodiaea* spp., *Allium* spp., *Muilla* spp.) that we were unable to distinguish in their vegetative states; Ely, *Elymus multisetus*; Fallweeds, weeds removed from plot at least one month prior to August harvest, not separated by species; Hem, *Hemizonia congesta* ssp. *luzulifolia*; Las, *Lasthenia californica*; Les, *Lessingia micradenia*; Lot, *Lotus subpinnatus*; Mic, *Microseris douglasii*; Nas, *Nassella pulchra*; OE, other early season annuals (unidentifiable in vegetative phase); OL, other late season annuals (unidentifiable in vegetative phase); OPF, other perennial forbs; Pla, *Plantago erecta*; Springweeds, weeds removed from plots at least one month prior to April harvest, not separated by species; Trifol, *Trifolium* spp.; Vul, *Vulpia microstachys*.

yield for a mixture,  $\Delta Y$ , is the observed yield ( $Y_o$ ) minus the expected yield ( $Y_e$ ):

$$\Delta Y = Y_o - Y_e = CE + SE$$

where  $CE = s \times \text{mean}(\Delta RY_i) \times \text{mean}(M_i)$ ,  $s$  is the number of species, and  $\Delta RY_i$  is the difference between the observed and expected relative yield for species  $i$

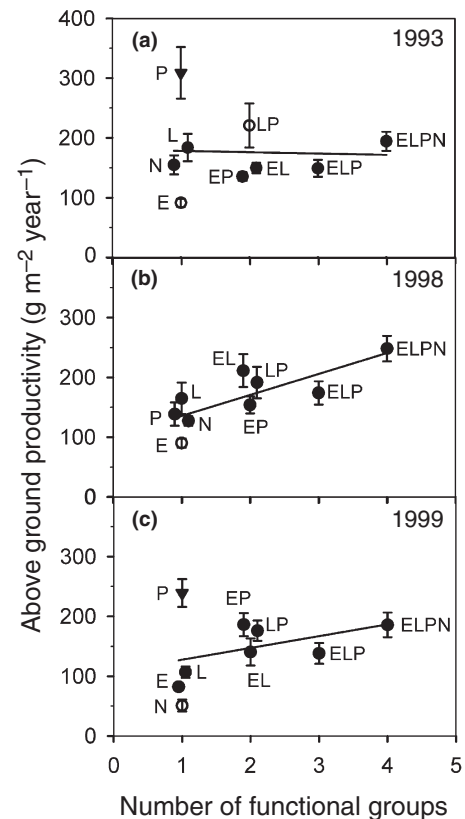
$$\Delta RY_i = RY_{oi} - RY_{ei}, \quad \text{where } RY_{ei} = 1/s.$$

$$SE = s \times \text{covariance}(\Delta RY_i, M_i) \quad (\text{Loreau \& Hector 2001}).$$

The selection effect is not strictly independent of complementarity (Petchey 2003) – if more productive species also tend to exhibit more resource partitioning (leading to a positive covariance between monoculture productivity and  $\Delta RY$ ), then some complementarity may be attributed to selection. As such, CE is likely to be conservative as a measure of complementarity. Similarly, SE and the original definition of the sampling effect do not correspond precisely, as pointed out by Loreau & Hector (2001) (see also Loreau 2000). Strong positive SE only indicates that the environment (biotic and abiotic) allows better than expected performance of (i.e. selects for) more productive species, whether by competitive dominance, as in the original sampling effects model (Huston 1997; Tilman *et al.* 1997), or improved resource partitioning or facilitation for more productive species. Finally, the magnitude of CE and SE cannot be compared across treatments because they are biased by the monoculture yields of the component species (Petchey 2003). For our treatment comparisons, we compare only whether these measures are significant.

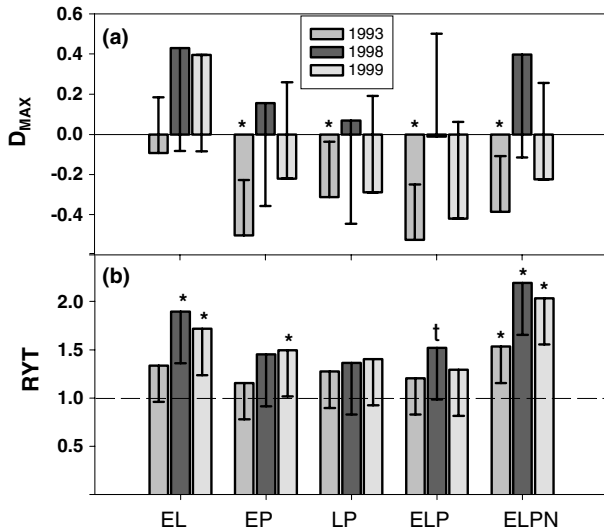
## RESULTS

The relationship between ANPP and functional group richness depended on the year. As previously reported (Hooper & Vitousek 1997), in 1993 functional group richness did not increase productivity (Fig. 2a). The perennial bunchgrasses were the most productive SFG treatment and the early season annuals the least. While the LP mixture had higher productivity than the other two-way mixtures, no mixtures exceeded the productivity of the perennial bunchgrasses alone. In contrast, ANPP increased significantly with functional group richness in 1998, and several mixtures out-produced the most productive SFG treatment, the late-season annuals – a pattern that suggests complementarity and transgressive overyielding (Fig. 2b). Again, the early season annuals were the least productive SFG treatment. The significance of the positive relationship between functional group richness and ANPP was affected somewhat by the distribution of N-fixers in the treatments. Eliminating both the N and ELPN treatments from the analysis reduced the slope of the response and the



**Figure 2** Response of productivity to plant functional group richness in 1993 (linear regression, with  $P$  values from randomization analyses:  $\text{ANPP} = -2.76x + 182$ ,  $r^2 = 0.001$ ,  $P = 0.752$ ), 1998 (linear regression:  $\text{ANPP} = 34.8x + 102$ ,  $r^2 = 0.303$ ,  $P = 0.002$ ) and 1999 (linear regression:  $\text{ANPP} = 19.0x + 110$ ,  $r^2 = 0.08$ ,  $P = 0.054$ ). Treatment means ( $\pm 1\text{SE}$ ) are shown, although regressions were performed on raw data. Significant differences among treatments within levels of functional group richness within years are indicated by different graph symbols (open circles, filled circles, triangles). Significance ( $P < 0.10$ ) was determined by *a priori* comparisons corrected for nine non-orthogonal tests by the sequential Dunn–Sidak method (Sokal & Rohlf 1995 p. 241) following ANOVA of natural log-transformed data using the methods of Hooper & Vitousek (1998). Data from 1993 reproduced from Hooper & Vitousek (1997) with permission of AAAS.

explanatory power, although the overall regression was still significant ( $m = 27.9$ ,  $r^2 = 0.108$ ,  $P = 0.032$ ). In 1999, the response of ANPP to functional group richness was intermediate between the responses seen in the other 2 years. The relationship was positive and borderline significant, but depended on the distribution of the N-fixer treatments ( $P = 0.868$  without N and ELPN) (Fig. 2c). As in 1993, the perennial bunchgrasses alone had the highest production among all single group treatments, and no mixtures exceeded this level. N-fixer production was substantially lower than in the other years.

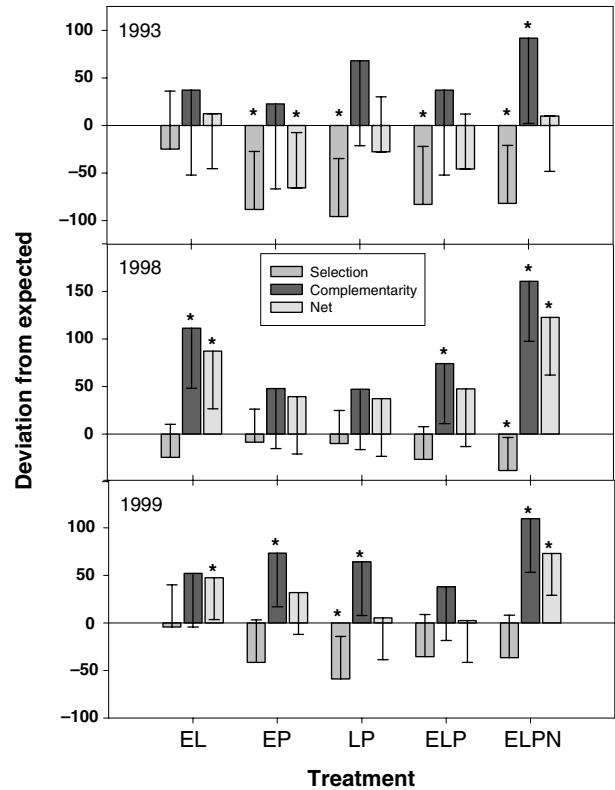


**Figure 3** Indices for assessing the degree of overyielding. (a)  $D_{max}$ , and (b) relative yield totals (RYT). Transgressive overyielding occurs when  $D_{max} > 0$ , and overyielding occurs when  $RYT > 1$ . Figures show means  $\pm$  95% confidence intervals, corrected for multiple comparisons by the Dunn–Sidak method, using  $t_{\alpha/[d.f.]}$  for five comparisons within each year,  $\alpha = 0.05$ , d.f. = 20 (Sokal & Rohlf 1995, Table C). Pooled standard errors were derived from one-way ANOVA, with functional group mixture as treatment, and with block effects included. An asterisk indicates significant ( $P < 0.05$ , corrected for multiple comparisons) differences from zero ( $D_{max}$ ) or 1 (RYT); †significance was borderline ( $0.10 > P > 0.05$ , corrected for multiple comparisons).

Several treatments overyielded, but not transgressively.  $D_{max}$  was not significantly greater than zero for any mixtures in any year (Fig. 3a). In 1993 and 1999,  $D_{max}$  in P-containing plots tended to be negative (significantly so in 1993), reflecting the production patterns in Fig. 2. In contrast, RYT indicated overyielding in the ELPN mixture in 1993, in the EL, ELP and ELPN mixtures in 1998, and in the EL, EP and ELPN mixtures in 1999 (Fig. 3b).

The functional groups that had the most positive production response in mixture (Es) were not the groups that were most productive in single-functional group treatments (Ls or Ps). All P-containing mixtures in 1993, the ELPN mixture in 1998 and the LP mixture in 1999 had significantly negative SE, although mean values for all treatments in all years were negative (Fig. 4). In some cases (ELPN in 1993, ELP 1998, EP and LP 1999), negative selection effects offset significantly positive CEs, so that mixture production did not differ significantly from the mean of monoculture productivities.

The most rigid criterion for complementarity using  $D_i$  is that all species or functional groups do better in mixture than expected from monoculture yields ( $D_i > 0$  for all). We found this pattern in only one mixture (EL) in



**Figure 4** Results of additive partitioning equation calculations for selection, complementarity, and net effects on total yield for mixture treatments (after Loreau & Hector 2001) in 1993, 1998 and 1999. Error bars are 95% confidence limits corrected for multiple comparisons as in Fig. 3. Symbols for significance are as in Fig. 3.

one year (1998; borderline significant in 1999) (Table 1). However, most mixtures had at least one group doing significantly better than expectations, even if other groups performed in mixture as expected from monocultures. In the LP mixture, the late season annuals did significantly better than expected, although the perennial bunchgrasses did worse (significantly so in 1993 and 1999), a pattern suggesting asymmetric competition despite an average (but not significant) RYT of 1.35 (Table 1, Fig. 3b). We saw a similar pattern in the ELPN mixture in 1993, with Es doing much better than expected and Ps doing significantly worse, despite RYT being significantly greater than one.

Functional groups differed in their responses to mixtures. Early season annuals exceeded expectations in all mixtures in all years (except EP in 1998; Table 1), despite having some of the lowest productivity in monoculture (Fig. 2). Performance of the late season annuals depended on mixture and year. While Ls and Es were strongly complementary in the EL mixture in 1998 and 1999, having perennial bunchgrasses in the community reduced the

**Table 1**  $D_i$  values for individual functional groups within mixtures,  $\pm 95\%$  confidence intervals. Confidence intervals are corrected for multiple comparisons as in Fig. 2. Significantly positive or negative ( $P < 0.05$ ) values are in bold (expected value for  $D_i = 0$ )

	EP			LP			ELP			ELPN		
	1993	1998	1999	1993	1998	1999	1993	1998	1999	1993	1998	1999
E	<b>0.82</b> $\pm$ 0.65	<b>1.07</b> $\pm$ 0.98	<b>0.92</b> $\pm$ 0.66	<b>1.00</b> $\pm$ 0.65	<b>0.83</b> $\pm$ 0.98	<b>0.99</b> $\pm$ 0.66	—	—	—	<b>1.29</b> $\pm$ 0.65	<b>1.36</b> $\pm$ 0.98	<b>1.10</b> $\pm$ 0.66
L	-0.15 $\pm$ 0.75	<b>0.74</b> $\pm$ 0.63	0.52 $\pm$ 0.53†	—	—	—	<b>1.31</b> $\pm$ 0.75	<b>0.94</b> $\pm$ 0.63	<b>1.24</b> $\pm$ 0.53	0.17 $\pm$ 0.75	0.22 $\pm$ 0.63	0.07 $\pm$ 0.53
P	—	—	—	<b>-0.69</b> $\pm$ 0.10	0.19 $\pm$ 0.44	0.004 $\pm$ 0.38	<b>-0.76</b> $\pm$ 0.10	<b>-0.21</b> $\pm$ 0.44	<b>-0.43</b> $\pm$ 0.38	<b>-0.84</b> $\pm$ 0.10	0.05 $\pm$ 0.44	-0.30 $\pm$ 0.38
N	—	—	—	—	—	—	—	—	—	—	—	0.34 $\pm$ 0.58

†Borderline significance:  $0.10 > P > 0.05$ , corrected for multiple comparisons.

performance of Ls in the ELP mixture, and adding N-fixers increased it again in the ELPN mixture in those years. The year 1993 was an exception in that Ls fared no better than expected in EL and ELPN mixtures. Except in 1993, N-fixers improved the productivity of all other groups, as reflected in high measures of overyielding (Fig. 3) and  $D_i$  in the ELPN treatment. In contrast, N-fixers themselves remained at expected levels of production (Table 1). Productivity by perennial bunchgrasses was significantly lower than expected in all treatments in 1993 and in later years came close to expectations in most mixtures. Despite greater overall productivity for the bunchgrasses in most plots in 1999 than 1998 (data not shown), they fared worse relative to monoculture expectations (lower  $D_i$ s) in all mixtures in 1999 than in 1998. In contrast, the Es, Ls and Ns produced more in monocultures and most mixtures in 1998 than 1999.

Total rainfall differed greatly among years: 659 mm in 1993, 885 mm in 1998 and 460 mm in 1999. The 1997–1998 growing season rainfall was the highest in the past 10 years recorded at this site, whereas the 1998–1999 rainfall was just above the long-term average of 427 mm (Simon 1994; California Department of Water Resources 2003).

## DISCUSSION

In this experiment, overyielding varied greatly depending on the functional characteristics of the species involved, the time since plot establishment, and the surrounding biotic and abiotic environment. We found long-term coexistence among all functional groups in our experimental plots. This coexistence, patterns of resource acquisition (Gulmon *et al.* 1983), and patterns of overyielding suggest that complementary resource use and/or facilitation are strong among the functional groups in this experiment. Similar to several other experiments (Hector *et al.* 1999; Troumbis *et al.* 2000; Loreau & Hector 2001; Engelhardt & Ritchie 2002), however, these effects were not sufficient to cause significant transgressive overyielding or consistent increases in productivity with increased plant diversity. The growing frequency of this pattern suggests, in contrast to some recent reviews (e.g. Schmid *et al.* 2002), that greater absolute production with greater species or functional diversity may be restricted to certain environmental conditions or particular species combinations.

### Patterns with time

Stronger overyielding in 1998 and 1999 compared with 1993 suggests that establishment of individuals and feedbacks from production of previous years can alter long-term patterns of overyielding compared with short-term results. Two other experiments also found a strengthening of

diversity effects over time (Tilman *et al.* 2001; Mulder *et al.* 2002), although the mechanisms were not clear. Previous results at our site pointed to a mix of complementary and competitive effects (Hooper 1998). Stronger overyielding in 1998 and 1999 was driven by higher relative yields of bunchgrasses in most mixtures and of late season annuals in the EL and ELPN mixtures. In 1993, perennial bunchgrasses had low  $D_i$  values because of exceptionally productive P monocultures and competition in mixtures from Es for nitrogen and from Ls for water (Hooper & Vitousek 1997; Hooper 1998; Dukes 2001). In subsequent years, however, Ps had both lower production in monocultures (Fig. 2) and better performance in mixtures. For those bunchgrass individuals surviving the first couple of years in mixtures, full establishment of deep perennial root systems apparently lessened the competitive effects of the other functional groups. Higher relative production of Ps and Ls in the ELPN mixture probably resulted from build-up of soil nitrogen from the presence of N-fixers. The higher relative production of Ls in the EL treatment and lower monoculture P production is harder to explain. These shifts could result from density-dependent litter feedbacks reducing yields of monocultures in later years, as both Ls and Ps leave substantial standing dead litter that persists for many years.

Following establishment, overyielding also varied across years for the perennial bunchgrasses, potentially because of differences in water availability. The Ps contributed more to overyielding in all mixtures in 1998 than in 1999 (Table 1). This pattern was not driven by absolute production of Ps, which was higher in monocultures and most mixtures in 1999 compared with 1998 (except ELP; data not shown). In 1999, however, the bunchgrasses in monocultures produced proportionally more than in mixtures, resulting in generally lower  $D_s$ . The opposite was true for all other functional groups: Es, Ls and Ns generally had lower productivity in 1999 compared with 1998, but monoculture and mixture yields changed roughly proportionally. Year-to-year climate variability, particularly water availability, can strongly influence relative abundances of species in natural serpentine grasslands, although direct relationships are not always clear (Hobbs & Mooney 1995). The patterns of overyielding and rainfall in our experiments suggest that greater moisture availability leads to greater complementarity and overyielding, as seen in other experiments with nutrients and light (e.g. Dukes 2001; Fridley 2002, 2003). Unfortunately, we do not have data for the years between 1993 and 1998 to test this relationship explicitly and to better understand the timing and trajectory of the shifts in overyielding in our plots.

### Mechanisms of overyielding

We found strong evidence for overyielding in a majority of mixtures once populations were well-established (three of

five mixtures in both 1998 and 1999). Furthermore, long-term coexistence among species in our plots contrasts with some short-term experiments having strong sampling effects, in which the experimenters postulate that competition will eliminate species over time (Engelhardt & Ritchie 2002; Fridley 2003). Reviews have found that the great majority of two and three-species mixtures tested have  $RYT < 1.3$  (96%, Trenbath 1974; 87%, Jolliffe 1997). In our experiment, average values for all mixtures equalled or exceeded 1.3 in both 1998 and 1999 (although not all were significantly greater than one). However, Trenbath's review omitted legume/non-legume mixes, and most experiments were with annual crops or grasses so species may have had limited resource use differentiation (Harper 1977, p. 265). Jolliffe (1997) found that those mixtures with greatest overyielding often included N-fixers, although some highly overyielding mixtures did not and some mixtures with legumes significantly underyielded.

In our experiment, N-fixers apparently facilitated, not just complemented, other functional groups in the ELPN mixture. The N-fixers themselves were at or near expected values of  $D_i$  in the ELPN mixture, however their presence led to higher relative production of the other functional groups compared with their production in the ELP mixture in both 1998 and 1999. Greater relative production could result from greater absolute production of individual functional groups in the presence of N-fixers or from lower expected yields because of decreased planting density at higher diversity. Both may have occurred in our experiment, but the former predominated: decreases in planting density alone were not sufficient to account for the strong increases in  $D_i$ . This pattern suggests that the primary effect of N-fixers on total production was through provision of additional nitrogen to other species (e.g. Mulder *et al.* 2002), not just by decreased competition for soil nitrogen.

While N-fixation clearly contributed to overyielding in the ELPN mixture, resource partitioning in time and space among non-N-fixing species also played a major role in this experiment. The number and identities of functional groups responding positively in mixtures depended on the mixture. By examining all  $D_i$  values in a mixture (Table 1) in addition to RYT, we developed a range of standards for overyielding. The strongest standard was met when every functional group in a given mixture had a positive  $D_i$  value (Loreau 1998). A 'relaxed' criterion for overyielding required that at least one group have a positive  $D_i$  and none have a significantly negative  $D_i$ . We focus here on results from 1998 and 1999, when the plots had fully established.

Not surprisingly, the greatest overyielding occurred among groups with strong differences in functional characteristics (Ewel 1986; Vandermeer 1990): the EL mixture met (1998) or came close to meeting (1999) the most rigorous



criterion. The ELPN mixture was intermediate in the degree of overyielding, with the highest RYT's but only two to three groups with  $D_i$  significantly greater than zero in both years. The EP and ELP mixtures had weaker overyielding, usually with only a single significantly positive  $D_i$  value. The species in the E and L groups differ strongly in rooting depth and timing of maximum resource use (Gulmon *et al.* 1983; Mooney *et al.* 1986), suggesting that overyielding resulted from complementary use of water and nitrogen. Temporal complementarity provides some of the strongest overyielding in intercropping studies (Vandermeer 1990). In our case, however, phenological and spatial (rooting depth) complementarity covary, so we cannot attribute our results solely to either mechanism. Strong performance of Es in mixture probably resulted from strong competition for nitrogen during the wet season (Hooper & Vitousek 1998; Hooper 1998). Interestingly, total root biomass was not a good predictor of success in mixtures, as in other low nutrient systems (e.g. Tilman & Wedin 1991). The early season annuals have a fine but dense root system compared with the perennial bunchgrasses, which have high total root biomass but a large proportion of fibrous roots. Complementarity between Es and Ls and significant increases in total biomass in 1998 even without N-fixers in the analysis (Fig. 2a) support other observations of positive diversity effects on productivity in the absence of legumes (van Ruijven & Berendse 2003). In contrast, the LP mixture exhibited asymmetric competition rather than complementarity, because the perennial bunchgrasses tended to do worse in mixture than expected (significantly so in 1999). Both groups in this mixture are deeply rooted and are likely to compete for late spring soil moisture (Dukes 2002).

Overyielding differed among the same groups in different mixtures. Although E and L functional groups were both strongly complementary in the EL mixture, only the early season annuals did better than expected in the ELP mixture (Table 1). This pattern contrasts with predictions of the potential for sampling effects for complementary sets of species (Loreau 2000; Fridley 2001; Loreau *et al.* 2001). In our case, a set of species could be strongly complementary in some mixtures but not in others. Clearly, species performances in complex mixtures are not simple functions of their performances in monocultures or less diverse mixtures, as implied in some sampling effects models (Huston 1997; Tilman *et al.* 1997).

Our results demonstrate that indices of overyielding based on relative yield totals (either RYT or CE) are best paired with more in-depth analyses using  $D_i$  or individual RY (see also Dukes 2001; Engelhardt & Ritchie 2002; Fridley 2003). Such analyses would help to resolve debates about how many and which species or functional groups contribute to overyielding (e.g. Hector *et al.* 1999; Huston *et al.* 2000; Tilman *et al.* 2001). They also provide a deeper

mechanistic understanding of which functional traits are likely to be complementary and how overyielding varies with the community mixture and abiotic conditions.

### Complementarity, overyielding and allocation

Relationships between complementarity and overyielding are complicated by different patterns of allocation of limiting resources among species. Despite the relatively strong indications of overyielding shown by RYT and  $D_b$ , we did not find a consistent, significantly positive relationship between average productivity and functional group richness (yes in 1998, no in 1993 and 1999). Similarly,  $D_{\max}$  was never significant, a pattern also observed in other experiments (Hector *et al.* 2002; Fridley 2003). In our experiment, the early-season annuals are strongly competitive for nutrients (Hooper & Vitousek 1998), but have a short growing season and a high proportional allocation to reproduction, leading to comparatively low relative growth rate and low total biomass accumulation compared with late season annuals or perennial bunchgrasses (Gulmon *et al.* 1983; Mooney *et al.* 1986; Armstrong 1991). However, the perennial bunchgrasses, which were the most productive in monoculture in 1993 and 1999, did relatively poorly in mixtures. These patterns of allocation and competition led to negative selection effects (Loreau & Hector 2001), with consequent reductions in total mixture biomass compared with expectations from complementary interactions. We would expect similar patterns where environmental conditions favour species with high allocation of nutrients to roots, reproduction or storage rather than to aboveground biomass increment (Chapin *et al.* 1986; Hector *et al.* 2002).

Negative selection effects that offset complementarity essentially decouple complementarity from absolute effects on net primary production. They have been reported in several other diversity/ecosystem functioning experiments (Hector *et al.* 1999, 2002; Troumbis *et al.* 2000; Engelhardt & Ritchie 2001; Špačková & Lepš 2001), and in competition experiments, particularly where fertility is low (e.g. Tilman & Wedin 1991). Complementarity will increase yields above the maximum monoculture yield only where the resource base is expanded enough above what the most productive species is able to garner to counterbalance differences in size and allocation among component species. The implications of this effect cut both ways: lack of transgressive overyielding does not rule out complementarity among species (Troumbis *et al.* 2000; Fridley 2001). Pure reliance on  $D_{\max}$  would miss the range of complementary interactions observed here. At the same time, complementarity does not necessarily lead to higher aboveground biomass accumulation with increasing diversity. While this point has been made before (e.g. Trenbath 1974; Fridley 2001; Hector *et al.* 2002), a common suggestion in both theoretical and empirical studies is that

complementarity could be a primary mechanism leading to increased productivity with increasing diversity, although differences in species' allocation are frequently ignored (e.g. Tilman *et al.* 1997; Mouquet *et al.* 2002). We found that even with strong complementarity and some facilitation among very different functional groups, absolute production only sometimes responded positively to plant functional diversity. That is, the expectation that complementarity will increase absolute productivity is not necessarily warranted. If negative selection effects are common, as they seem to be, strong facilitative interactions (as with N-fixers) may be necessary to generate consistent transgressive overyielding and overall positive effects of diversity on productivity in many ecosystems (e.g. Spehn *et al.* 2000; Tilman *et al.* 2001). Further understanding of the spatial, temporal and functional variation in the strength of complementarity, and how it links to community structure and ecosystem properties, will be critical to assessing the effects of diversity on ecosystem properties in both natural and managed ecosystems.

#### ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (grant DEB99-74159 to DH). JSD received support from NASA (Earth System Science Fellowship) and the US Department of Energy's OBER (Hollaender Postdoctoral Fellowship). Andy Hector and two anonymous reviewers gave invaluable comments on an earlier version of this manuscript. We thank Julia Verville, Sarah Hobbie, and Colleen Mohl for contributions in the field and the laboratory. Thanks also to P. Vitousek and H. Mooney for providing laboratory space, and to WMI and the personnel at the Kirby Canyon Landfill for allowing our research plots on-site.

#### REFERENCES

- Aarssen, L.W. (1997). High productivity in grassland ecosystems: affected by species diversity or productive species? *Oikos*, 80, 183–184.
- Armstrong, J. (1991). Rainfall variation, life form and phenology in California serpentine grassland. PhD Dissertation, Stanford University, Stanford, CA, USA.
- California Department of Water Resources (2003). *California Irrigation Management Information System*. URL <http://www.cimis.water.ca.gov/>
- Cardinale, B.J., Nelson, K. & Palmer, M.A. (2000). Linking species diversity to the functioning of ecosystems: on the importance of environmental context. *Oikos*, 91, 175–183.
- Chapin, F.S. III, Vitousek, P.M. & Van Cleve, K. (1986). The nature of nutrient limitation in plant communities. *Am. Nat.*, 127, 48–58.
- Dukes, J.S. (2001). Productivity and complementarity in grassland microcosms of varying diversity. *Oikos*, 94, 468–480.
- Dukes, J.S. (2002). Species composition and diversity affect grassland susceptibility and response to invasion. *Ecol. Appl.*, 12, 602–617.
- Engelhardt, K.A.M. & Ritchie, M.E. (2001). Effects of macrophyte species richness on wetland ecosystem functioning and services. *Nature*, 411, 687–689.
- Engelhardt, K.A.M. & Ritchie, M.E. (2002). The effect of aquatic plant species richness on wetland ecosystem processes. *Ecology*, 83, 2911–2924.
- Ewel, J.J. (1986). Designing agroecosystems for the humid tropics. *Annu. Rev. Ecol. Syst.*, 17, 245–271.
- Fridley, J.D. (2001). The influence of species diversity on ecosystem productivity: how, where and why? *Oikos*, 93, 514–526.
- Fridley, J.D. (2002). Resource availability dominates and alters the relationship between species diversity and ecosystem productivity in experimental plant communities. *Oecologia*, 132, 271–277.
- Fridley, J.D. (2003). Diversity effects on production in different light and fertility environments: an experiment with communities of annual plants. *J. Ecol.*, 91, 396–406.
- Gulmon, S.L., Chiarillo, N.R., Mooney, H.A. & Chu, C.C. (1983). Phenology and resource use in three co-occurring grassland annuals. *Oecologia (Berlin)*, 58, 33–42.
- Haggard, J.P. & Ewel, J.J. (1997). Primary productivity and resource partitioning in model tropical ecosystems. *Ecology*, 78, 1211–1221.
- Harper, J.L. (1977). *Population Biology of Plants*. Academic Press, London.
- 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.
- Hector, A., Bazely-White, E., Loreau, M., Otway, S. & Schmid, B. (2002). Overyielding in plant communities: testing the sampling effect hypothesis with replicated biodiversity experiments. *Ecol. Lett.*, 5, 502–511.
- Hobbs, R.J. & Mooney, H.A. (1995). Spatial and temporal variability in California annual grassland: results from a long-term study. *J. Vegetation Sci.*, 6, 43–56.
- Hooper, D.U. (1998). The role of complementarity and competition in ecosystem responses to variation in plant diversity. *Ecology*, 79, 704–719.
- Hooper, D.U. & Vitousek, P.M. (1997). The effects of plant composition and diversity on ecosystem processes. *Science*, 277, 1302–1305.
- Hooper, D.U. & Vitousek, P.M. (1998). Effects of plant composition and diversity on nutrient cycling. *Ecol. Monogr.*, 68, 121–149.
- Huston, M.A. (1997). Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia*, 110, 449–460.
- Huston, M.A., Aarssen, L.W., Austin, M.P., Cade, B.S., Fridley, J.D., Garnier, E. *et al.* (2000). No consistent effect of plant diversity on productivity. *Science*, 289, 1255.
- Jolliffe, P.A. (1997). Are mixed populations of plant species more productive than pure stands? *Oikos*, 80, 595–602.
- Jolliffe, P.A. (2000). The replacement series. *J. Ecol.*, 88, 371–385.
- Loreau, M. (1998). Separating sampling and other effects in biodiversity experiments. *Oikos*, 82, 600–602.
- Loreau, M. (2000). Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos*, 91, 3–17.

- Loreau, M. & Hector, A. (2001). Partitioning selection and complementarity in biodiversity experiments. *Nature*, 412, 72–76.
- 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, UK.
- McKane, R.B., Johnson, L.C., Shaver, G.R., Nadelhoffer, K.J., Rastetter, E.B., Fry, B. *et al.* (2002). Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. *Nature*, 415, 68–71.
- Mooney, H.A., Hobbs, R.J., Gorham, J. & Williams, K. (1986). Biomass accumulation and resource utilization in co-occurring grassland annuals. *Oecologia (Berlin)*, 70, 555–558.
- Mouquet, N., Moore, J.L. & Loreau, M. (2002). Plant species richness and community productivity: why the mechanism that promotes coexistence matters. *Ecol. Lett.*, 5, 56–65.
- Mulder, C.P.H., Jumpponen, A., Högberg, P. & Huss-Danell, K. (2002). How plant diversity and legumes affect nitrogen dynamics in experimental grassland communities. *Oecologia*, 133, 412–421.
- Petchey, O.L. (2003). Integrating methods that investigate how complementarity influences ecosystem functioning. *Oikos*, 101, 323–330.
- van Ruijven, J. & Berendse, F. (2003). Positive effects of plant species diversity on productivity in the absence of legumes. *Ecol. Lett.*, 6, 170–175.
- Schmid, B., Joshi, J. & Schläpfer, F. (2002). Empirical evidence for biodiversity–ecosystem functioning relationships. In: *Functional Consequences of Biodiversity: Experimental Progress and Theoretical Extensions* (eds Kinzig, A., Tilman, D. & Pacala, S.). Princeton University Press, Princeton, NJ, pp. 120–150.
- Simon, R.L. (1994). *Kirby Canyon RDF Meteorological Program Annual Summary Report for July 1993–June 1994*. Kirby Canyon Recycling and Disposal Facility, San Jose, CA.
- Sokal, R.R. & Rohlf, F.J. (1995). *Biometry*, 3rd edn. W.H. Freeman and Co., New York, NY.
- Špaèková, I. & Lepš, J. (2001). Procedure for separating the selection effect from other effects in diversity–productivity relationships. *Ecol. Lett.*, 4, 585–594.
- Spehn, E.M., Joshi, J., Schmid, B., Diemer, M. & Körner, C. (2000). Above-ground resource use increases with plant species richness in experimental grassland ecosystems. *Funct. Ecol.*, 14, 326–337.
- Swift, M.J. & Anderson, J.M. (1993). Biodiversity and ecosystem function in agricultural systems. In: *Biodiversity and Ecosystem Function* (eds Schulze, E.-D. & Mooney, H.A.). Springer-verlag, Berlin, Germany, pp. 15–41.
- Tilman, D. & Wedin, D. (1991). Plant traits and resource reduction for five grasses growing on a nitrogen gradient. *Ecology*, 72, 685–700.
- Tilman, D., Lehman, C.L. & Thomson, K.T. (1997). Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl Acad. Sci. USA*, 94, 1857–1861.
- Tilman, D., Reich, P.B., Knops, J., Wedin, D., Mielke, T. & Lehman, C. (2001). Diversity and productivity in a long-term grassland experiment. *Science*, 294, 843–845.
- Trenbath, B.R. (1974). Biomass productivity of mixtures. *Adv. Agron.*, 26, 177–210.
- Troumbis, A.Y., Dimitrakopoulos, P.G., Siamantziouras, A.S. & Memtsas, D. (2000). Hidden diversity and productivity patterns in mixed Mediterranean grasslands. *Oikos*, 90, 549–559.
- Vandermeer, J.H. (1990). Intercropping. In: *Agroecology* (eds Carrol, C.R., Vandermeer, J.H. & Rosset, P.M.). McGraw and Hill, New York, NY, pp. 481–516.
- Wardle, D.A. (1999). Is “sampling effect“ a problem for experiments investigating biodiversity – ecosystem function relationships? *Oikos*, 87, 403–407.

Editor, David Wardle

Manuscript received 7 July 2003

First decision made 11 August 2003

Manuscript accepted 12 November 2003