Multispecies grasslands produce more yield from lower nitrogen inputs across a climatic gradient

Article

Accepted Version

O'Malley, J., Finn, J. A., Malisch, C. S., Suter, M., Meyer, S. T., Peratoner, G., Thivierge, M.-N., Abalos, D., Adler, P. R., Bezemer, T. M., Black, A. D., Ergon, Å., Golińska, B., Grange, G., Hakl, J., Hoekstra, N. J., Huguenin-Elie, O., Jing, J., Jungers, J. M., Lajeunesse, J., Loges, R., Louarn, G., Luscher, A., Moloney, T., Reynolds, C. K. ORCID: https://orcid.org/0000-0002-4152-1190, Sturite, I., Khan, A. S., Vishwakarma, R., Zhang, Y., Zhu, F. and Brophy, C. (2025) Multispecies grasslands produce more yield from lower nitrogen inputs across a climatic gradient. Science. ISSN 1095-9203 doi: 10.1126/science.ady0764 Available at https://centaur.reading.ac.uk/127102/

To link to this article DOI: http://dx.doi.org/10.1126/science.ady0764

Publisher: American Association for the Advancement of Science version was published in Science on 4 December 2025, DOI: 10.1126/science.ady0764."

Title: Multispecies grasslands produce more yield from lower nitrogen inputs across a climatic gradient

Authors: James O'Malley[1], John A. Finn[2]*, Carsten S. Malisch[3], Matthias Suter[4], Sebastian T. Meyer[5], Giovanni Peratoner[6], Marie-Noëlle Thivierge[7], Diego Abalos[3], Paul R. Adler[8], T. Martijn Bezemer[9], Alistair D. Black[10], Åshild Ergon[11], Barbara Golińska[12], Guylain Grange[1,2], Josef Hakl[13], Nyncke J. Hoekstra[14], Olivier Huguenin-Elie[4], Jingying Jing[15], Jacob M. Jungers[16], Julie Lajeunesse[17], Ralf Loges[18], Gaëtan Louarn[19], Andreas Lüscher[4], Thomas Moloney[20], Christopher K. Reynolds[21], Ievina Sturite[22], Ali Sultan Khan[2], Rishabh Vishwakarma[1], Yingjun Zhang[15], Feng Zhu[23], Caroline Brophy[1]

Affiliations:

[1]School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland.

[2]Environment, Soils and Land Use Dept., Teagasc, Johnstown Castle, Co. Wexford, Ireland.

[3]Department of Agroecology, Aarhus University, Tjele, Denmark.

[4]Agroscope, Reckenholzstrasse, Zurich, Switzerland.

[5]School of Life Sciences, Technical University of Munich, Munich, Germany.

[6]Laimburg Research Centre, Laimburg 6, Vadena/Pfatten, Italy.

[7]Agriculture and Agri-Food Canada, Quebec Research and Development Centre, Quebec City, Quebec, Canada.

[8]United States Department of Agriculture, Agricultural Research Service, PA, USA.

[9]Institute of Biology, Leiden University, Leiden, The Netherlands.

[10]Department of Agricultural Sciences, Lincoln University, Christchurch, New Zealand.

[11]Dept. of Plant Sciences, Norwegian University of Life Sciences, Norway.

[12]Department of Grassland and Natural Landscape Sciences, Poznan University of Life Sciences, Poland.

[13]Department of Agroecology and Crop Production, Czech University of Life Sciences Prague, Czech Republic.

[14]Louis Bolk Institute, Bunnik, The Netherlands.

[15]College of Grassland Science and Technology, China Agricultural University, Beijing, China.

[16]Department of Agronomy and Plant Genetics, University of Minnesota, MN, USA.

[17]Agriculture and Agri-Food Canada, Research farm, Normandin, Quebec, Canada.

[18]Institute of Animal Nutrition and Physiology, Group of Organic Farming, Kiel University, Kiel, Germany.

[19]INRAE UR4 URP3F, Lusignan, France.

[20]DLF, Ballymountain, Ferrybank, Co. Waterford, Ireland.

[21]School of Agriculture, Policy and Development, University of Reading, Early Gate, UK.

[22]Department of Grassland and Livestock, Norwegian Institute of Bioeconomy Research (NIBIO), Ås station Steinkjer, Norway.

[23]Hebei Key Laboratory of Soil Ecology, Key Laboratory of Agricultural Water Resources, Center for Agricultural Resources Research, Institute of Genetic and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang, China.

*Corresponding author. Email: John.Finn@teagasc.ie

Abstract

High-yielding forage grasslands frequently comprise low species diversity and receive high inputs of nitrogen fertilizer. To investigate multispecies mixtures as an alternative strategy, the 26-site international 'LegacyNet' experiment systematically varied the diversity of sown grasslands using up to six high-yielding forage species (grasses, legumes, and herbs), managed under moderate nitrogen inputs. Multispecies mixtures outyielded two widely used grassland practices: a grass monoculture with higher nitrogen fertilizer, and a two-species grass-legume community. High yields in multispecies mixtures were driven by strong positive grass-legume and legume-herb interactions. In warmer sites, the yield advantage of legume-containing multispecies mixtures over monocultures and the high-nitrogen grass increased. Improved design of grassland mixtures can inform more environmentally sustainable forage production and may enhance adaptation of productive grasslands to a warming climate.

One Sentence Summary

High-yielding mixtures of grasses, legumes and herbs outperform communities managed under common grassland practices.

Main Text:

To feed a growing global population, agricultural production systems have undergone significant simplification and intensification to increase productivity and food/feed quality (*1*). In the context of ongoing resource depletion and the global climate and biodiversity crises, innovative agricultural practices are needed to balance food security with environmental sustainability (*2-5*). Ruminant livestock production is a major contributor to the global agricultural economy, and associated grassland forage production varies from high-input, intensively managed monocultures to more extensive, low-input systems (*6*). Sown grasslands in temperate regions (hereafter 'productive grasslands') typically comprise a small number of high-yielding perennial species, often grass monocultures reliant on high nitrogen fertilizer inputs or two species grass-legume communities, typically comprising 70% *Lolium perenne* (perennial ryegrass) and 30% *Trifolium repens* (white clover). Despite promoting yield increases, high-nitrogen inputs contribute to several agri-environmental problems, including greenhouse gas emissions, reduced soil and water quality, and biodiversity loss (*1*). The use of multispecies grasslands has been proposed to achieve high yields and increase sustainability (*7, 8*). However, critical gaps remain in understanding whether higher plant diversity with lower nitrogen inputs can match yields of productive grasslands with high inputs across varying climates and environmental conditions.

Biodiversity and ecosystem function is well-studied in semi-natural grassland experiments, where higher plant diversity generally increases yield (e.g., (*8-10*)) and enhances resistance and resilience to environmental disturbance (*11*). However, most semi-natural grasslands have low yields and/or limited forage quality, and many species mixtures fail to out-yield the most productive monoculture (*12*). Based on ecological principles, combining productive species with complementary traits has potential to deliver high yields with moderate inputs (*8, 13*). Mixtures of grasses and legumes in productive grasslands at low to medium nitrogen fertilizer levels can outyield the component monocultures (at the same level of nitrogen fertilizer) and achieve the same or higher yields as highly fertilized grass monocultures due to the nitrogen-fixing abilities of legume species (*14-17*). The inclusion of herb species (defined as non-leguminous forbs) in grass-legume mixtures increases functional diversity, which can increase overall resource utilization and ecosystem function (*10, 18*). However, the nitrogen-saving and yield-related benefits in productive grasslands and mixtures with more than two functional groups are largely unexplored.

Investigations of the addition of herbs to grass or grass-legume communities in productive grasslands are difficult to synthesize because they vary considerably in their experimental design (*19-26*). Systematic investigation across varying conditions is required for the development of robust agricultural practices intended for wide implementation. We established LegacyNet, a voluntary, coordinated 26-site network of grassland experiments spanning a broad climatic gradient within temperate regions across North America, Europe, Asia, and New Zealand (Fig. 1A, Table S1). LegacyNet sites implemented a common field experiment (*27*), where the composition and sown proportions of six high-yielding forage species (two grasses, two legumes, and two herbs, denoted G1, G2, L1, L2, H1 and H2) were manipulated at plot scale (Table S2), to assess the suitability of the wide scale adoption of multispecies mixtures in productive grasslands. Given the range of environmental conditions across sites, each site selected species suited to the locality (*27*). Nevertheless, there was high overlap in species across sites (Table S1); the species most commonly used across sites were G1 *L. perenne*, G2 *Phleum pratense* (timothy), L1 *Trifolium pratense* (red clover), L2 *T. repens*, H1 *Cichorium intybus* (chicory), and H2 *Plantago lanceolata* (plantain). Sites applied none (two sites) or moderate levels of nitrogen fertilizer, over an average duration of 24 months (Tables S1 and S2). Mixture and monoculture performances were compared to a grass monoculture (*L. perenne* at 21 of 26 sites) with a higher level of nitrogen fertilizer (at least twice the moderate rate), and a two-species community of 70:30 *L. perenne* and *T. repens*. Thus, LegacyNet investigated (a) whether diversity of plant species and/or functional groups affects the yields of productive grasslands; (b) whether higher-diversity mixtures can equal or exceed the yields of conventional management practices; and (c) the extent to which climate alters potential yield benefits of multispecies mixtures.

Mixing of functional groups underpinned a high degree of overyielding

Yield per growing season (*27*) varied substantially across sites, with median values for plots at each site ranging from 4.65 – 14.89 t ha^{-1}, reflecting the range of climates, soil conditions and management practices. Many mixture plots yielded higher than monoculture plots at individual sites (Fig. 1B).

Across sites, yields of the six forage species in monoculture were highest for L1, and all other species were similarly high to each other (Fig. 2; predicted from a Diversity-Interactions model (*28-30*), Tables S3 and S4). Overyielding occurs when a mixture outyields the weighted average of the component monocultures ('weighted average monoculture' hereafter) (*31*). The six-species grass-legume-herb equi-proportional mixture (GLH) and four-species equi-proportional mixtures of grass-legume (GL), legume-herb (LH), and grass-herb (GH) all displayed overyielding by outperforming their weighted average monoculture by 3.13 (GLH, +34%), 2.90 (GL, +31%), 2.46 (LH, +26%), and 0.78 (GH, +9%) t ha^{-1} respectively (each $P < 0.0001$, testing whether the net interactions (Fig. 2) differed from zero). Strong overyielding occurred across a wide range of sown proportions of grasses, legumes, and herbs (Figs. 3A and B). We identified other six-species mixtures that outperformed the equi-proportional six-species mixture (e.g., Fig. 3A: the optimal mixture, G:L:H = 0.24: 0.59: 0.17, yielded 12.83 (+4%) compared to 12.31 t ha^{-1} for the equi-proportional mixture, cyan versus magenta circle; $P < 0.0001$). Thus, we found overwhelming evidence of overyielding in mixtures that was robust across a wide range of sown functional group proportions; this was due to strong interaction effects (Fig. 3A, Table S3), with interactions being important even when some functional groups had low sown proportions (partially due to the non-linear term theta (*29, 32*) in Eq. 3 (*27*) and Table S3, Fig. S1).

Transgressive overyielding occurs when a mixture outperforms the best-performing monoculture (*31*). Averaged across sites, a wide range of sown proportions of grasses, legumes and herbs performed better than (Fig. 3C, the blue region) or comparably to (Fig. 3C, the two light grey regions) the best-performing monoculture. This high degree of transgressive overyielding is only possible due to positive net effects of interspecific interactions (quantified in Table S3 and illustrated in grey in Fig. 2). There were strong positive interactions between grasses and legumes, between herbs and legumes, and to a lesser degree between grasses and herbs (Table S3). The interaction between the two legumes species was moderate, but was non-significant between the two grasses and between the two herbs. Although we cannot discern among the various mechanisms that underpin these interactions, they likely include the ability of mixtures to utilize a wider range of soil, nutrient and light resources, especially nitrogen (*33*), as well as a reduced susceptibility to pests and diseases. Strong grass-legume interactions are consistent with previous research at single sites (*16, 33, 34*), and across environmental conditions (*14*), and reflect the importance of nitrogen fixation by legumes and the ability of other functional groups to transform the additional available nitrogen into biomass. Some evidence for interactions of herbs with grasses or legumes has been shown in single-site studies (*22, 23*) but our results provide large-scale evidence that multispecies mixtures that include herb species can also deliver high yields.

Strong legume-herb and moderate grass-herb interaction effects contributed to a wide range of six-species mixtures outyielding two- and four-species grass-legume mixtures when comparing across communities with the same proportion of legumes (Fig S2).

Increasing species diversity within functional groups can benefit yield

When two species are functionally redundant (*28*), they are interchangeable at sowing without any impact on yield. At the 26 individual sites, the two species within each functional group were only functionally redundant at seven sites for grasses, four sites for legumes and seven sites for herbs (Fig. S3). Functional redundancy, however, may have even lower incidence if multiple responses are considered, and/or over different timescales (*35*).

We further investigated whether yields could be optimized by sowing one rather than two species per functional group. Across sites, equi-proportional mixtures with three species spanning three functional groups yielded on average 0.91 t ha^{-1} lower (-7%, Fig. S4) than the equi-proportional six-species mixture, clearly indicating a yield advantage of the latter. As richness increased from three to six species, any increase in yield strongly depended on species composition (Fig. S5). At individual sites, the majority of equi-proportional three-species combinations yielded lower (125 of 208 comparisons) or comparably (81 of 208 comparisons) to the equi-proportional six-species mixture and yielded higher in only two comparisons (Fig. S4). In the absence of prior knowledge on local species' performance, sowing two species within each functional group, rather than one, is a more reliable strategy to optimize yield and reduce sensitivity to changes in functional group proportions (Fig. S6). Thus, even with a targeted choice of high-performing forage species, six-species GLH mixtures are better than three-species GLH mixtures.

Multispecies mixtures with lower nitrogen outperformed a high N grass monoculture

To compare mixture performance with a grass monoculture with high nitrogen, the LegacyNet design included a high-nitrogen application (on average across sites 2.4 times higher than the moderate rate; Table S1) on additional monocultures of G1 ('high N grass' hereafter). Across sites, yields of multispecies mixtures across a wide range of sown functional group proportions and managed with moderate nitrogen fertilizer either outyielded or were comparable to the high N grass yield (Fig. 3D, blue plus two light grey regions). The high N grass yielded 11.07 t ha^{-1}, and the equi-proportional six-species mixture was 1.23 t ha^{-1} higher ($P = 0.0002$, +11%; Fig. 2: six-species equi-proportional mixture versus the dashed red line), despite the average nitrogen rate being 152 kg ha^{-1} lower for the multispecies mixture. Therefore, functionally-diverse mixtures with much lower nitrogen application can outyield a high N grass monoculture. Mixtures that contained little to no legumes had lower yields than the high N grass (Fig. 2, grass-grass, herb-herb and grass-herb predictions, and Fig. 3D, black region). At 21 of 26 sites, the yield of the six-species equi-proportional mixture was higher (12) or comparable (nine) to the high N grass, and significantly lower only at five sites (Fig. S7A, Table S5). The extent of these effects provides evidence that farmers can utilize multispecies mixtures in lieu of highly fertilized grass monocultures to reduce their financial costs and environmental impacts while simultaneously maintaining or improving their yields.

Multispecies mixtures outperformed a two-species grass-legume community

Twenty-one of 26 LegacyNet sites sowed *L. perenne* as their G1 species, and 20 of 26 sowed *T. repens* as their L2 species (Table S1). Therefore, we use our model to estimate the performance of a two-species community of 70% *L. perenne* and 30% *T. repens* ('70:30 G1:L2 community' hereafter). Across sites, the equi-proportional six-species mixture and the equi-proportional four-species GL and LH mixtures, outyielded the 70:30 G1:L2 community (10.44 t ha^{-1}; Fig. 2 dashed black line) by 1.86 t ha^{-1} (+18%), 1.79 t ha^{-1} (+17%), and 1.53 t ha^{-1} (+15%), respectively ($P <$ 0.0001 each) (Fig. 2). Similar results could also be seen at the individual site level (Fig. S7B, Table S5). A wide range of four- and six-species mixtures of grasses, legumes, and herbs (where species within each functional group were sown in equal proportion) significantly outyielded the two-species 70:30 G1:L2 community (Fig. 3E). These results underline that increases in species richness (from two to four to six) and manipulation of relative abundances of four-species grass-legume and six-species grass-legume-herb mixtures can result in higher yields than those from the two-species 70:30 G1:L2 community. While legumes play a key role (Fig. 3A), six-species mixtures of grasses, legumes, and herbs always out-performed two- and four-species grass-legume communities when comparing across communities with the same sown legume percentage, and for values less than 70% (Fig. S2).

Mixtures with legumes increased yield advantages across a temperature gradient

Yield generally showed a quadratic relationship with average daily temperature across sites (Fig. 4A, Table S3). Yields were reduced considerably at lower temperatures (Fig. 4A); for example, the equi-proportional six-species mixture yielded 10.47 t ha^{-1} at 5°C compared to 12.50 t ha^{-1} at 11°C. Although the effect of temperature was similar on all monocultures and grass-herb mixtures (Fig. 4A), any mixture containing legumes yielded proportionately more as temperature increased, because the grass-legume and legume-herb interactions were stronger at higher temperatures (Table S3, climate terms). Thus, yield patterns across the equi-proportional grass-legume, legume-herb and grass-legume-herb communities were quite consistent from 9°C to 13°C (Fig. 4A and Fig. S8A); e.g., at 9°, 11°, and 13°C degrees in Fig. 4A, the six-species equi-proportional mixture yielded 12.21, 12.50 and 12.39 t ha^{-1} respectively. Importantly, at higher temperatures, these legume-related interactions caused an increased yield difference between the six-species equi-proportional mixture and both the high N grass monoculture (Fig. 4B) and all the other monocultures (Fig. 4A). There were combinations of sown grass-legume-herb proportions with clear evidence of i) overyielding (Fig. S8B) across the temperature gradient, ii) transgressive overyielding at and above 9°C (Fig. S8C), iii) outperformance of the high N grass at and above 7°C (Fig. S8D), and iv) outperformance of the 70:30 G1:L2 community at all temperatures (Fig. S8E).

Implications for agronomic practice

To optimize yield and nitrogen-saving benefits, we identified a range of functional group proportions defined by: 30 – 70% of legumes, a minimum of 15% of grasses, and a minimum of 10% of herbs (Fig. 4C, shaded region). Many diverse mixtures within this region exhibited transgressive overyielding by outperforming the highest-yielding monoculture (L1, red clover, in 23 of 26 sites). In practice, red clover monocultures would not be sown in productive grasslands, given the risk of disease in fields repeatedly cropped with it, bloat risk for grazing ruminants, and negative environmental effects on nitrous oxide emissions and nitrate leaching (*36, 37*). Despite being an overly stringent comparison, it is all the more impressive that many mixtures outperformed red clover, the highest-yielding monoculture.

We retrospectively identified the optimal-yielding communities at each site (Fig. S9), which were generally within or close to the optimal region of community proportions across sites (Fig. 4C). However, the site-specific sown proportions of these optimal communities would not be known in advance and, furthermore, would likely vary depending on the weather conditions in the year of establishment and subsequent years. Therefore, aiming for sown proportions of the functional groups within the specified region across sites (Fig. 4C) is a more dependable approach to optimize yield. The wide range of species proportions in this region also allows more targeted design of grassland communities to match different farming systems, without any compromise in yields, e.g., lower legume proportions (~30%) would be more appropriate for a grassland grazed by ruminants, whereas higher legume proportions may better suit an ungrazed grassland ley in a rotation.

Productive grasslands are expected to deliver more than just forage yield, and communities in the optimal region (Fig. 4C) can also reduce agri-environmental harms associated with nitrogenous gaseous emissions and leaching (*36-39*), and lead to benefits across multiple ecosystem functions (multifunctionality), including agronomic (*40, 41*), livestock (*21, 23, 42*) and other environmental (*13, 43*) responses.

The duration of the LegacyNet experiments generally matches well with that of grassland leys as implemented within rotational cropping systems, where they can enhance soil carbon, disrupt pest and weed life cycles, and reduce GHG emissions and nitrate leaching (*44*). Even within a short multispecies grassland ley, savings from reduced nitrogen fertilizer far exceed additional seed costs. Productive grassland systems can have a four- to eight-year duration that exceeds our experimental length. Although we show yield benefits of mixtures in the initial years, species' proportions may shift away from the optimal region (Fig. 4C) over several years; however, adaptive farming management (e.g. over-sowing or slot-seeding of seeds of declining species, rather than ploughing and reseeding (e.g. (*45*)) can direct grassland composition back to the optimal region. The insights provided by this study on the value of multispecies mixtures can incentivize longer-term agronomic and grazing studies, as well as improved selection of forage plant species and cultivars that are more persistent, maintain functional diversity over time, and improve synergistic interactions among species in mixtures.

Grassland production contributes greatly to livelihoods in rural regions and is threatened by the climate crisis (*46, 47*). Our study strengthens the evidence that legume-containing multispecies mixtures are an adaptation option for future climate scenarios (*48, 49*). Climate change also increases the incidence of extreme weather events, for which our results do not necessarily provide insight. However, multispecies mixtures can enhance yield resilience to weather-related stresses, including drought (*8, 24, 50*).

More diverse multispecies mixtures can generally deliver higher yields than either a grass monoculture with higher nitrogen input or a two-species grass-legume mixture. Overall, we highlight the strong potential of combining high-performing plants from different functional groups to enhance forage yields across a wide gradient of temperate climates and soil conditions, while achieving reductions in nitrogen fertilizer inputs.

References and Notes:

1. G. Lemaire, F. Gastal, A. Franzluebbers, A. Chabbi, Grassland–cropping rotations: an avenue for agricultural diversification to reconcile high production with environmental quality. *Environmental Management* **56**, 1065-1077 (2015).
2. M. B. Cole, M. A. Augustin, M. J. Robertson, J. M. Manners, The science of food security. *npj Science of Food* **2**, 14 (2018).
3. H. C. J. Godfray *et al.*, Food security: the challenge of feeding 9 billion people. *Science* **327**, 812-818 (2010).
4. H. C. J. Godfray, T. Garnett, Food security and sustainable intensification. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20120273 (2014).
5. Y. Yang *et al.*, Climate change exacerbates the environmental impacts of agriculture. *Science* **385**, eadn3747 (2024).
6. FAO, "Transforming the livestock sector through the Sustainable Development Goals," (Food and Agriculture Organization of the United Nations, 2018).
7. G. Tamburini *et al.*, Agricultural diversification promotes multiple ecosystem services without compromising yield. *Science Advances* **6**, eaba1715 (2020).
8. F. Isbell *et al.*, Benefits of increasing plant diversity in sustainable agroecosystems. *Journal of Ecology* **105**, 871-879 (2017).
9. L. Zheng *et al.*, Effects of plant diversity on productivity strengthen over time due to trait-dependent shifts in species overyielding. *Nature Communications* **15**, 2078 (2024).
10. W. W. Weisser *et al.*, Biodiversity effects on ecosystem functioning in a 15-year grassland experiment: Patterns, mechanisms, and open questions. *Basic and Applied Ecology* **23**, 1-73 (2017).
11. F. Isbell *et al.*, Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature* **526**, 574-577 (2015).
12. B. J. Cardinale *et al.*, Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proceedings of the National Academy of Sciences* **104**, 18123-18128 (2007).
13. J. A. Finn *et al.*, Design principles for multi-species productive grasslands: Quantifying effects of diversity beyond richness. *Journal of Ecology* **112**, 2471-2479 (2024).
14. J. A. Finn *et al.*, Ecosystem function enhanced by combining four functional types of plant species in intensively managed grassland mixtures: a 3-year continental-scale field experiment. *Journal of Applied Ecology* **50**, 365-375 (2013).
15. M. Jørgensen, A. K. Bakken, L. Østrem, C. Brophy, The effects of functional trait diversity on productivity of grass-legume swards across multiple sites and two levels of nitrogen fertiliser. *European Journal of Agronomy* **151**, 126993 (2023).
16. D. Nyfeler *et al.*, Strong mixture effects among four species in fertilized agricultural grassland led to persistent and consistent transgressive overyielding. *Journal of Applied Ecology* **46**, 683-691 (2009).
17. M. Suter *et al.*, Nitrogen yield advantage from grass–legume mixtures is robust over a wide range of legume proportions and environmental conditions. *Global Change Biology* **21**, 2424-2438 (2015).
18. M. W. Cadotte, K. Carscadden, N. Mirotchnick, Beyond species: functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology* **48**, 1079-1087 (2011).
19. K. M. McCarthy, C. G. McAloon, M. B. Lynch, K. M. Pierce, F. J. Mulligan, Herb species inclusion in grazing swards for dairy cows—A systematic review and meta-analysis. *Journal of Dairy Science* **103**, 1416-1430 (2020).
20. C. Grace *et al.*, The effect of increasing pasture species on herbage production, chemical composition and utilization under intensive sheep grazing. *Grass and Forage Science* **73**, 852-864 (2018).
21. C. Grace *et al.*, Grazing multispecies swards improves ewe and lamb performance. *Animal* **13**, 1721-1729 (2019).
22. S. Baker *et al.*, A comparison of the nutritive value of perennial ryegrass, perennial ryegrass and white clover and multispecies-based farmlet systems. *Grassland Research* **2**, 129-139 (2023).
23. A. I. Roca-Fernández, J. L. Peyraud, L. Delaby, R. Delagarde, Pasture intake and milk production of dairy cows rotationally grazing on multi-species swards. *Animal* **10**, 1448-1456 (2016).
24. G. Grange, J. A. Finn, C. Brophy, Plant diversity enhanced yield and mitigated drought impacts in intensively managed grassland communities. *Journal of Applied Ecology* **58**, 1864-1875 (2021).
25. J. B. Pinxterhuis *et al.*, Implementing plantain (*Plantago lanceolata*) to mitigate the impact of grazing ruminants on nitrogen losses to the environment: A review. *Grass and Forage Science* **79**, 144-157 (2024).

26. M. A. Sanderson *et al.*, Plant species diversity and management of temperate forage and grazing land ecosystems. *Crop Science* **44**, 1132-1144 (2004).
27. Materials and methods.
28. L. Kirwan *et al.*, Diversity–interaction modeling: estimating contributions of species identities and interactions to ecosystem function. *Ecology* **90**, 2032-2038 (2009).
29. J. Connolly *et al.*, An improved model to predict the effects of changing biodiversity levels on ecosystem function. *Journal of Ecology* **101**, 344-355 (2013).
30. R. A. Moral *et al.*, Going beyond richness: Modelling the BEF relationship using species identity, evenness, richness and species interactions via the DImodels R package. *Methods in Ecology and Evolution* **14**, 2250-2258 (2023).
31. B. R. Trenbath, in *Advances in Agronomy,* N. C. Brady, Ed. (Academic Press, 1974), vol. 26, pp. 177-210.
32. R. Vishwakarma, L. Byrne, J. Connolly, R. de Andrade Moral, C. Brophy, Estimation of the non-linear parameter in Generalised Diversity-Interactions models is unaffected by change in structure of the interaction terms. *Environmental and Ecological Statistics* **30**, 555-574 (2023).
33. D. Nyfeler, O. Huguenin-Elie, M. Suter, E. Frossard, A. Lüscher, Grass-legume mixtures can yield more nitrogen than legume pure stands due to mutual stimulation of nitrogen uptake from symbiotic and non-symbiotic sources. *Agriculture, Ecosystems & Environment* **140**, 155-163 (2011).
34. E. Sturludóttir *et al.*, Benefits of mixing grasses and legumes for herbage yield and nutritive value in Northern Europe and Canada. *Grass and Forage Science* **69**, 229-240 (2014).
35. F. Isbell *et al.*, High plant diversity is needed to maintain ecosystem services. *Nature* **477**, 199-202 (2011).
36. S. Cummins *et al.*, Beneficial effects of multi-species mixtures on N_2O emissions from intensively managed grassland swards. *Science of the Total Environment* **792**, 148163 (2021).
37. D. Nyfeler, O. Huguenin-Elie, E. Frossard, A. Lüscher, Effects of legumes and fertiliser on nitrogen balance and nitrate leaching from intact leys and after tilling for subsequent crop. *Agriculture, Ecosystems & Environment* **360**, 108776 (2024).
38. I. Feigenwinter, L. Hörtnagl, N. Buchmann, N_2O and CH_4 fluxes from intensively managed grassland: The importance of biological and environmental drivers vs. management. *Science of The Total Environment* **903**, 166389 (2023).
39. A. Egan, T. Moloney, J. B. Murphy, P. J. Forrestal, Ribwort plantain inclusion reduces nitrate leaching from grass-clover swards; A multi-year five soil study. *Agriculture, Ecosystems & Environment* **380**, 109376 (2025).
40. M. Suter, O. Huguenin-Elie, A. Lüscher, Multispecies for multifunctions: combining four complementary species enhances multifunctionality of sown grassland. *Scientific Reports* **11**, 3835 (2021).
41. G. Grange, C. Brophy, R. Vishwakarma, J. A. Finn, Effects of experimental drought and plant diversity on multifunctionality of a model system for crop rotation. *Scientific Reports* **14**, 10265 (2024).
42. L. M. Cranston, P. R. Kenyon, S. T. Morris, P. D. Kemp, A review of the use of chicory, plantain, red clover and white clover in a sward mix for increased sheep and beef production. *Journal of New Zealand Grasslands* **77**, 89-94 (2015).
43. I. Ikoyi, G. Grange, J. A. Finn, F. P. Brennan, Plant diversity enhanced nematode-based soil quality indices and changed soil nematode community structure in intensively-managed agricultural grasslands. *European Journal of Soil Biology* **118**, 103542 (2023).
44. C. S. Malisch *et al.*, The importance of multi-species grassland leys to enhance ecosystem services in crop rotations. *Grass and Forage Science* **79**, 120-134 (2024).
45. M. Dodd, T. Rhodes, Multi-year performance of white clover oversown into eastern North Island hill country. *Journal of New Zealand Grasslands* **86**, 129-134 (2024).
46. A. K. Knapp *et al.*, Rainfall variability, carbon cycling, and plant species diversity in a mesic grassland. *Science* **298**, 2202-2205 (2002).
47. D. Loka *et al.*, Impacts of abiotic stresses on the physiology and metabolism of cool-season grasses: A review. *Food and Energy Security* **8**, e00152 (2019).
48. Å. Ergon *et al.*, How can forage production in Nordic and Mediterranean Europe adapt to the challenges and opportunities arising from climate change? *European Journal of Agronomy* **92**, 97-106 (2018).
49. M.-N. Thivierge *et al.*, Perennial forages in cold-humid areas: Adaptation and resilience-building strategies toward climate change. *Agronomy Journal* **115**, 1519-1542 (2023).
50. E. Haughey *et al.*, Higher species richness enhances yield stability in intensively managed grasslands with experimental disturbance. *Scientific Reports* **8**, 15047 (2018).
51. N. T. Longford, *Random coefficient models*. (Oxford University Press, New York, 1993).

52. B. T. West, K. B. Welch, A. T. Gałecki, *Linear mixed models: a practical guide using statistical software, second edition*. (Chapman and Hall/CRC, New York, ed. 2, 2022).
53. M. Loreau *et al.*, Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**, 804-808 (2001).
54. Q. H. Vuong, Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307-333 (1989).
55. R Core Team, R: a language and environment for statistical computing, version 4.4.2. https://www.R-project.org/ (2024).
56. SAS Institute Inc., in *SAS/STAT® 15.2 User's Guide*. (Cary, NC: SAS Institute Inc., 2023).
57. H. Wickham, ggplot2: Elegant graphics for data analysis, version 3.5.1. https://ggplot2.tidyverse.org (2016).
58. D. Kahle, H. Wickham, ggmap: spatial visualization with ggplot2. *The R Journal* **5**, 144-161 (2013).
59. R. Vishwakarma, C. Brophy, L. Byrne, C. Hurley, DImodelsVis: visualising and interpreting statistical models fit to compositional data, version 1.0.1. https://CRAN.R-project.org/package=DImodelsVis (2024).
60. R. Vishwakarma, C. Brophy, C. Hurley, PieGlyph: axis invariant scatter pie plots, version 1.0.0. https://CRAN.R-project.org/package=PieGlyph (2024).

Acknowledgements:

Funding:

Research Ireland Frontiers for the Future programme, grant number 19/FFP/6888 (CB, JOM)

TCD Postgraduate Research Studentship, code 1252 (JOM)

Multi4More (project ref. 21R456) funded by the Department of Agriculture, Food & the Marine and the Department of Agriculture, Environment and Rural Affairs (JAF, RV)

European Union's Horizon 2021 doctoral network programme under the Marie Skłodowska-Curie grant agreement No. 101072579 (LegumeLegacy) (CB, JAF, CSM, BG, CKR, NJH)

Author contributions

Conceptualization: JOM, CB, JAF, CSM, MS, GG, AL

Data curation: JOM, CB, RV

Formal Analysis: JOM, CB, JAF, MS, RV

Funding acquisition: JOM, CB, JAF, AL

Investigation: JAF, CSM, STM, GP, MNT, DA, PRA, TMB, ADB, ÅE, BG, GG, JH, NJH, OHE, JJ, JMJ, JL, RL, GL, AL, TM, CKR, IS, ASK, YZ, FZ

Methodology: JOM, CB, JAF, MS, AL, RV

Project administration: CB, JAF, CSM

Software: JOM, CB, RV

Supervision: CB, JAF, CSM, MS

Visualization: JOM, CB, JAF, MS, RV, AL

Writing – original draft: JOM, CB, JAF, CSM, MS, STM, GP, MNT, AL

Writing – review and editing: JOM, CB, JAF, CSM, MS, STM, GP, MNT, DA, PRA, TMB, ADB, ÅE, BG, GG, JH, NJH, OHE, JJ, JMJ, JL, RL, GL, AL, TM, CKR, IS, ASK, RV, YZ, FZ

Competing interests: Authors declare that they have no competing interests

Data and materials availability: All data and code required to perform the analysis has been made available on Dryad at the time of submission.

Supplementary materials

Materials and methods

Tables S1 – S5

Figures S1 – S9

References (*51*-*60*)

Figures

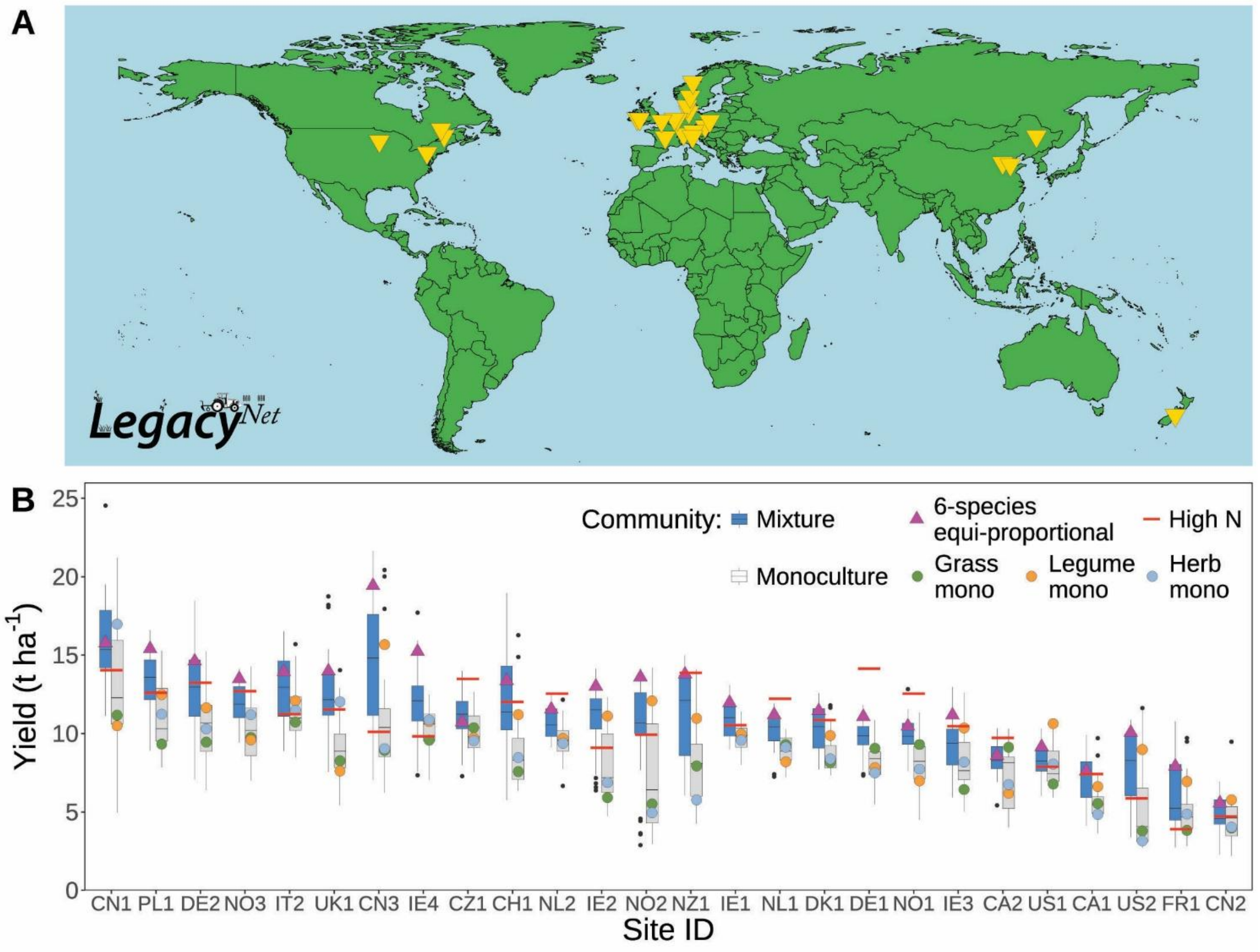


Fig. 1. Yields varied substantially across the LegacyNet sites. (A) The locations of the 26 LegacyNet sites indicated by inverted yellow triangles (see Table S1 for the site location details and ID codes). **(B)** The yield per growing season (t ha^{-1}) based on plot-level measurements for each site are shown in boxplots that are grouped by mixtures (blue; 29 plots per site) and monocultures (grey; 18 plots per site), at the moderate N fertilizer level. The average yield of the six-species equi-proportional mixture plots (three plots per site) is shown as a magenta triangle. The average yield of the monoculture plots from each functional group is highlighted with a circle (green for grass, orange for legume and light blue for herb; six monoculture plots per functional group per site). The average yield of the high N grass monoculture (five plots per site) is shown by a red horizontal line. Sites are arranged on the x-axis in order of decreasing median yield.

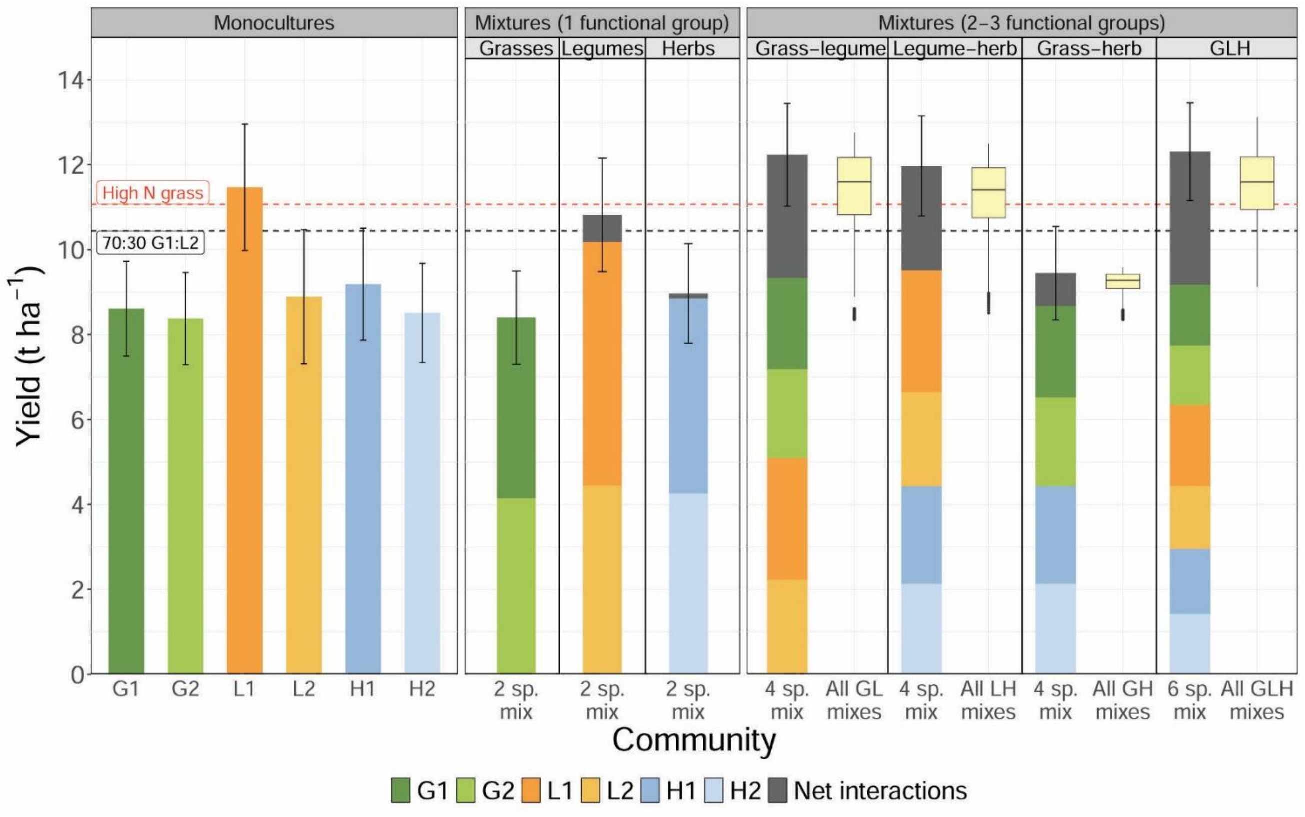


Fig. 2. Yields of monocultures and multispecies mixtures with varying functional group richness. Predicted yield per growing season (t ha^{-1}) for selected communities across sites (predicted from the model estimates in Table S3, n = 1,382). For each bar, yield predictions are broken down by the components of the model prediction (model species' identity effect estimates, each scaled by the respective sown proportion, plus scaled net interactions for mixtures; G1 = grass 1, G2 = grass 2, L1 = legume 1, L2 = legume 2, H1 = herb 1 and H2 = herb 2). The error bars show 95% confidence intervals for each prediction. The monoculture bars show predicted yield for the six monoculture communities. The 'Mixtures (1 functional group)' bars show predictions for the equi-proportional two-species mixtures where the two species are from the same functional group. For the G1:G2 equi-proportional mixture, the interaction term is negative and therefore not shown (effect = -0.09t ha^{-1}, SE = 0.17); the height of the bar represents the contribution of G1 and G2 to the yield, minus the negative interaction effect. The 'Mixtures (2–3 functional groups)' bars show predicted yields for the four-species equi-proportional GL, LH and GH mixtures, and the predicted yield of the six-species equi-proportional GLH mixture. The boxplots show the estimated distributions of predicted responses for all possible GL mixtures, LH mixtures, GH mixtures, and GLH mixtures. The dashed red line is the predicted yield of the high N grass monoculture. The dashed black line is the predicted yield of a 70:30 G1:L2 two-species community, where G1 is *L. perenne* and L2 is *T. repens* at the majority of sites.

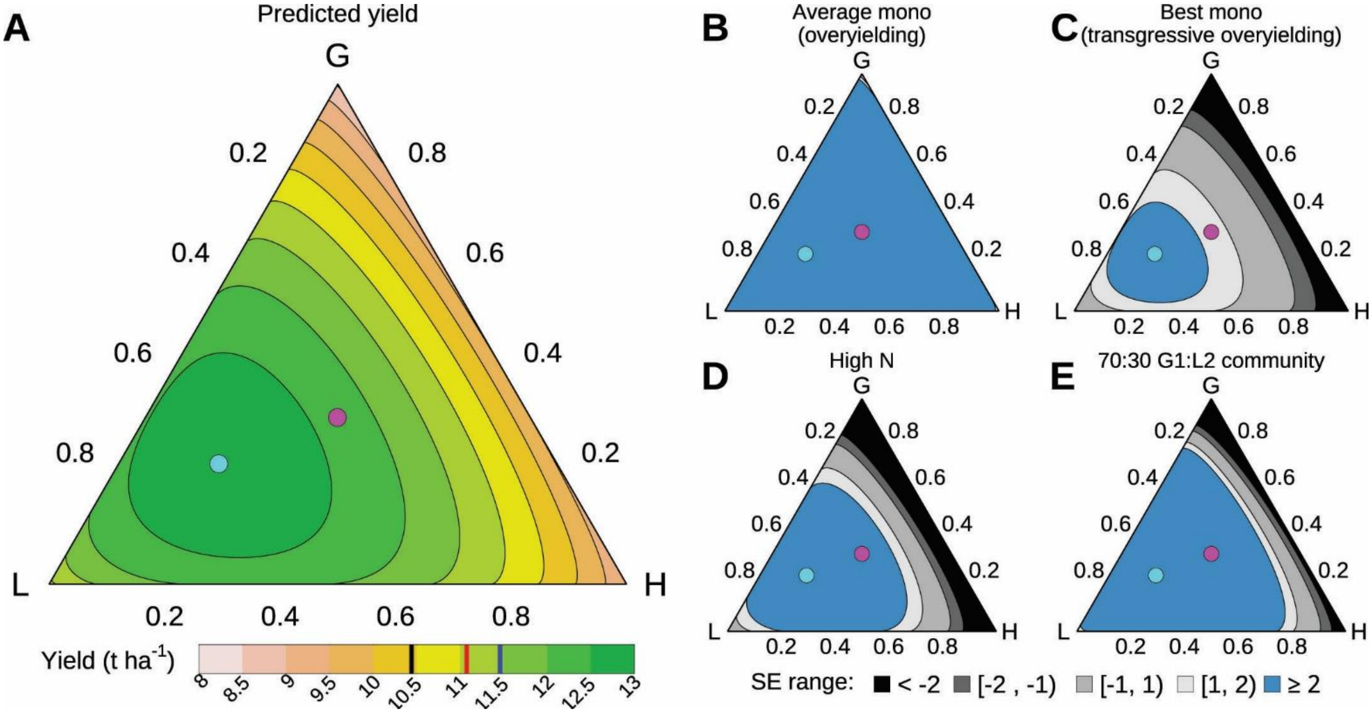


Fig. 3. Higher yields from increased plant diversity. (**A**) Ternary diagram illustrating predicted yield (t ha^{-1}) across all sites as the sown proportions of the three functional groups vary, and with the two species within each functional group sown in equal proportions to each other. Note that the vertices of the ternary are two-species communities of grasses (G), legumes (L), or herbs (H), not monocultures of individual species. The highest predicted yield is marked by a cyan circle (with sown functional group proportions: G:L:H = 24:59:17 and corresponding sown species proportions = 0.12: 0.12: 0.295: 0.295: 0.085: 0.085). In the yield legend, values are marked for the predicted responses of the 70:30 G1:L2 community (black), the high N grass monoculture (red), and the highest yielding monoculture (blue). The predicted yield for the equi-proportional six-species mixture is shown by a magenta circle. The remaining panels show the difference (measured in standard errors) of the predicted yield in **A** from: (**B**) the weighted average monoculture (for the component species at each point in the ternary; overyielding), (**C**) the highest yielding monoculture (i.e., L1, which was *T. pratense* at most sites; transgressive overyielding), (**D**) the high N grass monoculture, and (**E**) the 70:30 G1:L2 two-species community, where G1 is *L. perenne* and L2 is *T. repens* at most sites. For example, a point on panel **D** within the highest contour level (colored blue) represents a mixture whose predicted yield is more than two standard errors greater than the predicted yield of the high N grass monoculture. The six-species equi-proportional mixture and the optimal mixture are marked in **B** to **E** for comparison to panel **A**.

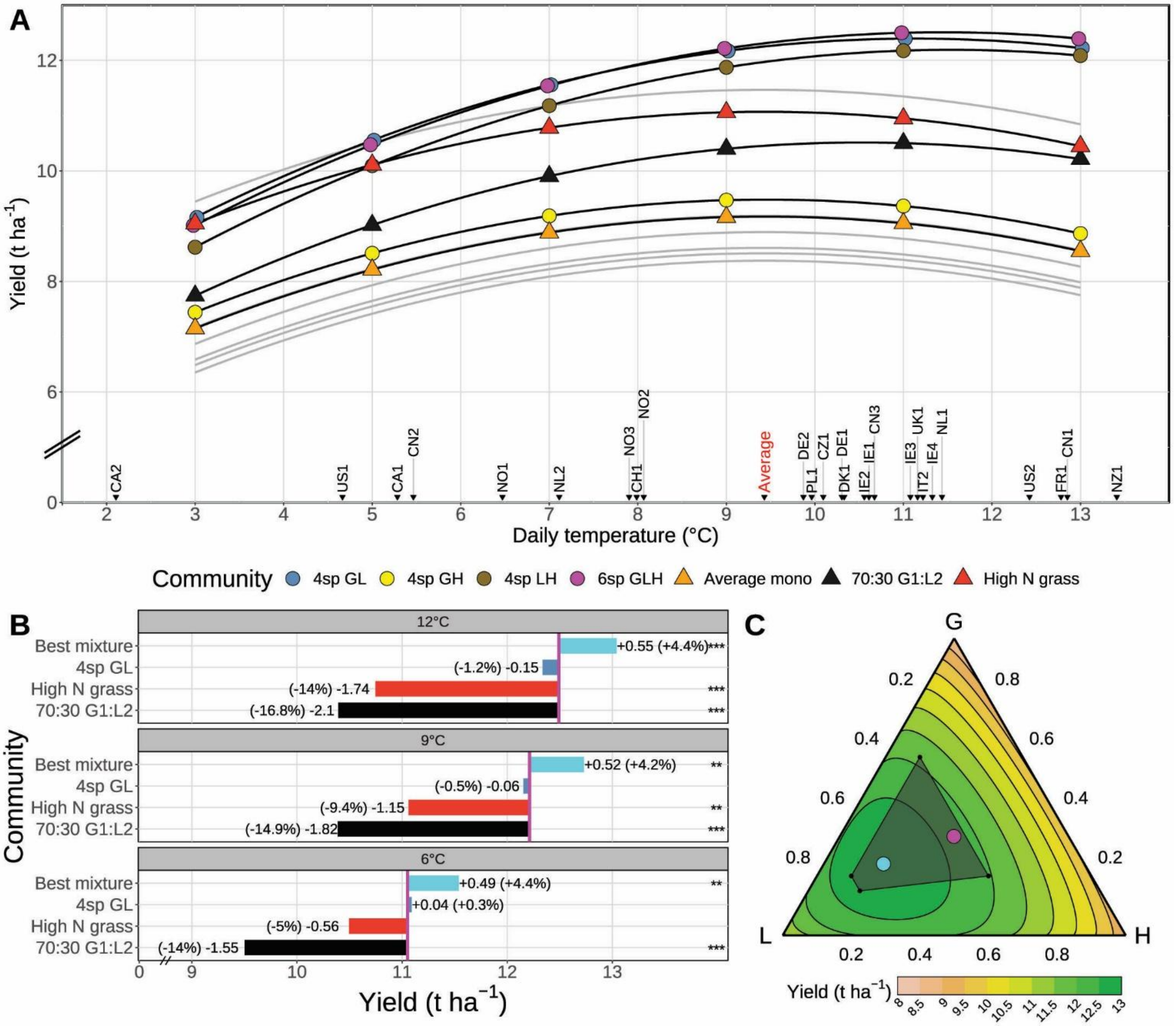


Fig. 4. As mean site temperatures increased, multispecies mixtures best maintained yield. (A) Predicted yields across the gradient of 3°C to 13°C average daily temperature for selected communities: the four-species equi-proportional mixtures of grass-legume, grass-herb and legume-herb (blue, yellow and brown circles, respectively), the six-species equi-proportional mixture of grass, legume and herb (magenta circle), the average of the six monocultures (orange triangle), the two-species 70:30 G1:L2 community (black triangle), and the high N grass monoculture (red triangle). Grey curves show the predicted yields of the six monoculture species (from highest- to lowest-yielding these are L1, H1, L2, G1, H2, G2). Circles on the top two curves are jittered. Site ID labels (Table S1) are included along the x-axis to indicate the average daily temperature at each site during the experimental period, the across-site average of daily temperatures is indicated as 'Average' (in red). **(B)** The predicted yield (t ha^{-1}) for the equi-proportional six-species mixture for three average daily temperatures 3°C, 9°C and 12°C (vertical magenta line in each panel). For each temperature, differences are shown between the six-species equi-proportional mixture and each of four selected communities (as labeled on the y-axis): 70:30 G1:L2

community (black), high N grass (red), four-species equi-proportional grass-legume mixture (blue), and the optimal 'best' mixture (cyan). Significant differences are highlighted on the right-hand side (*** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$). **(C)** Ternary diagram showing predicted yield (t ha^{-1}) across all sites. The shaded region identifies a range of sown proportions of grasses, legumes and herbs for multispecies communities that optimize yield.

This supplement contains:

Materials and methods
Tables S1 – S5
Figures S1 – S9
References (*51*-*60*)

Materials and methods

Experimental design of the LegacyNet common experiment

LegacyNet is a voluntary network of experiments that spans 26 sites across 15 countries. The aim of LegacyNet is to test the benefits of multispecies grassland leys within crop rotations. All sites within the network implemented a common two-stage experimental design that consisted of a grassland ley phase followed immediately by a follow-on crop phase. A range of plant communities in the grassland phase, from monocultures of each of the six component species to six-species mixtures, were tested with the objective of identifying optimal species mixtures that maximize forage production and legacy effects. Conducting a common multi-site experiment across a variety of environment conditions (climate, soil type, local management etc.) increases the ability to draw more general inference and conclusions about effect sizes. The focus of this paper is on the biomass yield response measured during the grassland ley phase.

At each site, six species were selected from three functional groups (two grasses, two legumes, and two herbs; denoted G1, G2, L1, L2, H1, H2) for use in the grassland phase. Species were selected for their high forage yields and quality, and complementary traits regarding the manner of nitrogen acquisition and growth form. Furthermore, the species were chosen to be representative of the most commonly used forage species at their location and are therefore best suited to the climatic, soil, and management conditions of the region; the species used at each site are listed in Table S1. The first grass species, G1, was chosen as the grass species that would be most commonly sown in the site's locality, this was *L. perenne* at 21 of 26 sites. As the second grass species, G2, *P. pratense* was used at 17 of 26 sites. The first legume species, L1, was *T. pratense* in 23 of 26 sites, while L2 was *T. repens* at 20 of 26 sites; both species are sown in productive temperate grasslands that include legumes. The first herb species, H1 was *C. intybus* at 25 sites which is commonly sown in leys that include herbs (e.g., (*42*)), and the second herb species, H2, was *P. lanceolata* at 24 of 26 sites.

The experimental design of the grassland phase was a simplex design comprising 33 unique communities of systematically varying sown proportions of each of six species. There were monocultures of each of the six species, and mixture plots with two, three, four, or six species (where species were sown in equal proportion). The sown proportions of each species for the design communities are shown in Table S2 (communities 1-33). Monocultures and the six-species mixture were replicated three times each, while remaining communities were each established in one experimental unit (Table S2; communities 1-33). Note that although some mixtures are not replicated beyond one experimental unit, replication is achieved from the spread of individual points across the sown species proportions (as is possible for any continuous design space). The application rate of synthetic nitrogen (N) fertilizer was also manipulated experimentally in the design; most of the plots across the simplex design (Table S2, communities 1-33) were managed at a moderate N level (with average application rate of inorganic nitrogen fertiliser across sites of

108.7 kg ha^{-1} yr^{-1}), while five (additional) replicated plots of the G1 grass monoculture were managed at a high N level (Table S2, community 34; with average application rate of inorganic nitrogen fertiliser across sites of 260.5 kg ha^{-1} yr^{-1}). Specific levels of N application were dictated by local practice (values shown in Table S1, two sites used zero N fertilizer as their moderate level). The high N level was typically at least twice the moderate N level (the difference between the two rates was on average 152 kg ha^{-1}). At each site, the seeding rate for monocultures was the locally recommended seeding rate for the establishment of monocultures of the respective species. To calculate the intended sown species proportion for each species in mixture communities, the sown proportion for each species (Table S2) was multiplied by its corresponding monoculture seeding rate, and seeds for all species were combined to create the seed mixture for sowing in the field plot. At each site, the 52 plant communities were allocated to field plots in a fully randomized design. All plots measured a minimum of 3 m × 5 m. Across all sites, there were a total of 1,382 plots (while most sites had just the 52 plots listed in Table S2, there were some small deviations at individual sites).

During the grassland phase, plots were regularly harvested via mechanical harvester at time intervals dictated by local practice (the total number of harvests in the grassland phase for each site is shown in Table S1). At each harvest, the dry biomass yield (t ha^{-1}) of each plot was measured by cutting and weighing the fresh mass by a plot harvester and by determining its dry matter (DM) content on a herbage sub-sample that was oven-dried to constant weight. At most harvests, a sample of the harvest biomass was taken for forage quality analysis. No weeding took place throughout the duration of the experiment, with the exception of a cleaning cut within the first six months of establishment at some sites.

Weather data recorded at each site included maximum, minimum, and mean daily temperatures in °C and daily precipitation in mm (including snow and irrigation applications); these variables were recorded monthly in the case of CN2 and CN3. A range of summary statistics were calculated across the grassland phase period (from the establishment date to the date of the final harvest) for each of these variables, these were: average daily temperature, average daily precipitation, averages of the ten highest and lowest daily temperatures, and antecedent precipitation index (API, an estimate of soil moisture values). This gave rise to one value per site for each of these variables.

Plots were maintained during the grassland phase for a minimum of 18 months, after which they were terminated and a follow-on crop of a pure grass ley, cereal, or maize was established. The duration of the grassland phase was 24 months on average across sites. The follow-on crop was grown and measured for one full growing season and retained the same plot structure as the grassland phase. Follow-on crop measurements are not part of the analyses presented here.

Calculation of the yield response variable

The total number of harvests over the grassland phase and the duration of the phase varied by site (Table S1). The typical length of a growing season also varied substantially across sites. To account for these site variations, a yield per growing season for each plot at each site was calculated by first calculating the average daily yield of each plot during the grassland phase and then multiplying it by an estimate of the average number of growing days in a single growing season for the site. We define the following notation used in the calculation of the yield response variable:

- H_k is the total number of harvests taken at site k over the grassland phase.

- $m_{k[q]h}$ is the DM yield of harvest *h* for plot *q* at site *k*. The DM yields were summed across all harvests in the grassland phase to give the total DM yield per plot.
- GDL_k (growing days in the grassland phase) is the total number of days during the grassland phase at site *k* where the plots were deemed to be growing. It was calculated at each site by summing up the number of days in each year where the plots were harvested, between the 1st of January to the final harvest where the ten-day rolling average temperature exceeded a threshold of 5°C. If the plots were harvested in the same year as establishment, the number of growing days for that year were counted from the date of establishment (rather than the 1st of January) to the final harvest. The number of growing days in each year at site *k* were then summed to give GDL_k.
- Thus, $\frac{\sum_{h=1}^{H_k} m_{k[q]h}}{GDL_k}$ estimates the average daily yield of each plot during the grassland phase at a given site.
- $GDSS_k$ (growing days in a single season) is an estimate of the number of days in a single growing season of site *k*. This variable was calculated for each full (calendar) year where the experiment was running at each site; the year of termination was only included if the final harvest in that year was deemed to be at the end of the growing season. Thus, for a single year, we computed the number of days where the ten-day rolling average temperature exceeded a threshold of 5°C between the 1st of January (or from the 1st of July in the case of New Zealand), and the date of the site's final harvest in that year. Where there was more than one full year of grassland phase at a site, the average over the available years was taken as the $GDSS_k$ value.

Finally, $y_{k[q]}$, the yield per growing season for plot *q* at site *k* was computed as:

$$y_{k[q]} = \frac{\left(\sum_{h=1}^{H_k} m_{k[q]h}\right)}{GDL_k} * (GDSS_k) \qquad \text{Eq. (1)}$$

We refer to *y* as 'yield per growing season' or 'yield' for short throughout the main text and its units are t ha^{-1}.

Statistical analysis using Diversity-Interactions modelling

Traditional approaches to modelling the biodiversity and ecosystem function relationship focus on the number of species (species richness) as the main determinant of ecosystem function. More recently, the Diversity-Interactions modelling approach extends this to assess how species composition, richness and initial proportions jointly affect ecosystem function (*28-30*).

We modelled the LegacyNet plot-level yield data from across all sites using the Diversity-Interactions (DI) modelling framework, fitted as a random coefficients linear mixed-effects model (*51*), where the response, $y_{k[q]}$, was yield per growing season (t ha^{-1}), calculated as described in Eq. (1). This allowed us to quantify the effects of manipulating species diversity on yield in the multispecies grasslands across our multi-site international-scale experiment. The DI model takes the general form:

$$y = [Identity\ terms] + [Interaction\ terms] + [High\ N\ term] + \varepsilon \qquad \text{Eq. (2)}$$

The model explicitly includes species' identity effects through sown species' proportion predictors (six species' identity parameters), species' pairwise interactions that can take varying forms (some versions more parsimonious than others), and a term for the high N grass monoculture (one parameter). The model implicitly includes species richness as an explanatory variable (*30*).

A possible model specification is:

$$y_{k[q]} = \sum_{i=1}^{6} (\beta_i + b_{ik}) p_{k[q]i} + \sum_{1 \le i < j \le 6} (\delta_{ij} + d_{ijk}) \left(p_{k[q]i} p_{k[q]j}\right)^{\theta} + (\alpha + a_k) X_{k[q]} + \varepsilon_{k[q]} \quad \text{Eq. (3)}$$

Where $p_{k[q]i}$ represents the sown proportion of species i in plot q at site k. $X_{k[q]}$ is coded as one for high N monoculture plots and zero otherwise. When $X_{k[q]} = 1$, each $p_{k[q]i}$ is set to zero, hence the overall expected response for the high N monocultures is α. The lowercase b, d, and a terms represent the random coefficients in the model. The random terms assumptions are:

$$(b_{1k}, \dots, b_{6k}, d_{12k}, \dots, d_{56k}, a_k) \sim MVN(\mathbf{0}, \boldsymbol{D})$$

$$\varepsilon_{k[q]} \sim N(\mathbf{0}, \sigma_k^2)$$

Hence, there are overall (across site) effects β, δ, and α, that vary from site to site according to their corresponding random effect, with the variance of each random coefficient capturing the nature of the spread across sites. The variance-covariance matrix **D** can be structured in many forms, for example, a parsimonious and biologically motivated structure related to the meaning of the various predictors in the model is:

$$\mathbf{D} = \begin{pmatrix} \sigma_{b_1}^2 & \cdots & \sigma_{SS} & \sigma_{SI} & \cdots & \sigma_{SI} & \sigma_{SN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{SS} & \cdots & \sigma_{b_6}^2 & \sigma_{SI} & \cdots & \sigma_{SI} & \sigma_{SN} \\ \sigma_{SI} & \cdots & \sigma_{SI} & \sigma_{d_{12}}^2 & \cdots & \sigma_{II} & \sigma_{IN} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{SI} & \cdots & \sigma_{SI} & \sigma_{II} & \cdots & \sigma_{d_{56}}^2 & \sigma_{IN} \\ \sigma_{SN} & \cdots & \sigma_{SN} & \sigma_{IN} & \cdots & \sigma_{IN} & \sigma_a^2 \end{pmatrix} \quad \text{Eq. (4)}$$

Where each random coefficient has a unique variance term, to allow it to vary from site to site: $\sigma_{b_i}^2$ for the random effect associated with each species identity term; $\sigma_{d_{ij}}^2$ for the random effect associated with each pairwise interaction between species *i* and *j*; and σ_a^2 for the random effect associated with the high N term. Each pair of species identity random effects (b terms) have a common covariance (σ_{SS}); each pair of interaction random effects (d terms) have a common covariance (σ_{II}); each pair of species identity-interaction random effects have a common covariance (σ_{SI}), while the high N random effect has a covariance with the species identity random effects (σ_{SN}) and with the interaction random effects (σ_{IN}).

Model selection process

Our model selection procedure was adapted from West, Welch, and Galecki (*52*) for a random coefficients Diversity-Interactions modelling framework. There were two main components to our model selection process: select the best structure for the fixed effect interaction terms, including whether the θ parameter differed from 1, and select the best random coefficients structure to capture variation from site to site. The following process was used.

1. We fitted a Diversity-Interactions model (*28-30*) that assumed the full pairwise fixed effect interaction structure (i.e., all pairs of species interacted uniquely giving 15 individual interaction terms). Site was included as a fixed blocking factor to account for site-to-site variation (site as a fixed effect is used only for this first step in the selection process, mixed-effects models will be used in subsequent steps to capture site-to-site variation). This model was fitted firstly with θ set to 1, and secondly with θ estimated freely. The model with θ = 1 was fitted using ordinary least squares estimation, while the model with θ estimated was fitted via profile likelihood (*30*), and the two models were compared using the sample-size corrected Akaike's Information Criterion AICc.

2. We fitted a random coefficients DI model, still assuming the full pairwise interaction structure, with all identity terms, species interaction terms and the high N term as random coefficients varying from site to site. Following (*32*), we used the value of θ selected in step 1; i.e., the θ parameter was not re-estimated in this step of the selection process. We tested the variations in the structure of the variance-covariance matrix for the random effects (i.e., matrix D), including:
    - Variance components (the same variances for each random effect and zero covariances between each pair; 1 parameter),
    - Compound symmetry with homogeneous variances (equal variances and equal covariances; 1 + 1 parameters),
    - A biologically motivated covariance structure related to the meaning of the predictors in the model: unique variances for all random effects, covariances for all pairs of identity effects, all pairs of identity-interaction effects, all pairs of interaction-interaction random effects, all high N-ID effects, all high N-interaction effects; 22 + 5 parameters, as specified in the matrix D example shown in Eq. (4).

    These models were fitted using Restricted Maximum Likelihood (REML) and compared using AICc. Note that the variance-covariance structure that would assume unique variances and unique covariances between each pair of random effects (22 + 231 = 253 parameters) was not fitted due to convergence issues related to computational scale.

3. We assessed the assumption that the within-site error variance was constant across all sites and compared it to a heterogeneous structure that allowed the within-site error term variance to differ by site. These two models were estimated with REML and compared using AICc.

4. We compared the model selected in step 3 to alternative ways to model the pairwise interactions (*28, 30*). We tested the assumptions that the fixed interaction effects were dictated by functional group membership, that all interactions were equal, and that all interaction terms were equal to 0. We followed a two-step procedure to test this, whereby random effects and fixed effects were changed one at a time; models that differed only in

random effects were fitted using REML for comparison, while models that differed only in fixed effects were fitted using Maximum Likelihood (ML) for comparison.

5. We re-estimated θ using the model selected in step 4 to confirm its final estimate.

6. Finally, we tested a range of weather variables one by one for inclusion as fixed effects in the model selected in step 5 to select our final model (Sattherthwaite denominator degrees of freedom were used during the tests). The variables investigated were average daily temperature, average daily precipitation, averages of the ten highest and lowest daily temperatures, antecedent precipitation index (API, an estimate of soil moisture values), each calculated per site over the grassland phase. Interactions between the weather variables with species identity and species interaction terms and with the high N term, and quadratic terms of the variables were also tested.

As part of the model selection process, model diagnostics were checked via a range of measures. These included plots of residuals and random effect estimates, and Cookes Distance and the PRESS statistic to assess the influence of individual sites on parameter estimates and model predictions respectively.

The final model selected

The final model was fitted using REML and had six species identity effect terms (one each for G1, G2, L1, L2, H1, H2), six functional group species pairwise interaction terms (grass-legume, grass-herb, legume-herb, grass-grass, legume-legume, herb-herb) that each included the non-linear θ parameter that differed from 1, and a high N term. All of these were fitted as random coefficients, i.e., an overall effect across sites was estimated and an associated variance term captured its variation from site to site. The structure of the variance covariance matrix for the random effects was similar to the matrix D example provided in Eq. (4). We found evidence of a quadratic effect of average daily temperature, and a positive interaction between average daily temperature and the GL and HL interactions. Although the p-value for the quadratic temperature effect in Table S3 is non-significant, model diagnostics and tests revealed that the quadratic effect was important, but more variable at the extreme lower end than the middle of the temperature gradient; this was investigated using 'leave-one-out' statistics where sites were omitted one by one. The predictive power of the model was best when a quadratic form of the temperature effect was included. We did not find evidence for the inclusion of other tested weather variables.

The estimates of the fixed effects from the final model are in Table S3, and random effect variance component estimates are in Table S4. Note that the average daily temperature variable was centered across sites to aid in the interpretation of the fixed effects estimates in Table S3, where the average across sites of the average daily temperature variable was 9.43°C. Thus, all parameters in Table S3 (excluding the three weather-related variables) can be interpreted for a temperature of 9.43°C.

Deriving inference from the final Diversity-Interactions model

Community comparison tests such as overyielding (communities outperforming the weighted average of their component monoculture performances), transgressive overyielding (communities outperforming the best monoculture), and tests of comparison against the high N grass monoculture and the two-species 70:30 G1:L2 community, were conducted using predictions from the fitted final DI model. We describe the general principle of these comparisons.

Assume it is of interest to test if a community M is significantly different from another community B. Let x_M be a vector containing the predictor variables for community M, and x_B be the vector of predictor variables for community B. We calculate the contrast comparing the predicted yields of communities M and B as:

$$\hat{C} = \hat{y}_M - \hat{y}_B$$

Where $\hat{y}_M$ is the predicted yield of M, $\hat{y}_B$ is the predicted yield of B.
We calculate *t*, as

$$t = \frac{\hat{C} - 0}{SE(\hat{C})}$$

Assuming a null hypothesis that the contrast is equal to 0 and where the standard error of $\hat{C}$ can be calculated using the variance-covariance matrix of our fixed model parameters, denoted **V**.

$$SE(\hat{C}) = \sqrt{(x_M - x_B)\boldsymbol{V}(x_M - x_B)'}$$

We infer significance of the comparison between mixtures M and B, if *t* is greater than 2, or less than minus 2.

For tests performed across sites, we used the fixed effects estimates and fixed effects covariance matrices from our fitted mixed model (Tables S3 and S4). In the case of tests at the individual site level, we used the model fitted to each site individually (estimates not shown).

Some of our selected communities for comparison are in the experimental design (e.g., the high N grass monoculture), while some are not (e.g., the predicted two-species 70:30 G1:L2 community) but can be predicted since they are part of the continuous design space in our experiment that is modelled via the Diversity-Interactions modelling approach.

Testing for functional redundancy

Functional redundancy refers to the situation where two (or more) species fulfil identical roles in an ecosystem and thus substitution of one by another does not impact the ecosystem function of interest (*28, 53*). A pair of species is said to be functionally redundant, if they have identical monoculture performances (identity effects), do not interact with one another, and have the same interaction strengths with each of the other species in the ecosystem (*28*). Under the DI framework, the presence/absence of functional redundancy can be assessed by fitting a DI model where these conditions are respected and testing if it fits the data better than the DI model containing separate identity and interaction effects for each species. For each site separately, the functional redundancy between the two species within each functional group was tested (separately for each functional group). For example, to assess functional redundancy between the two grasses, G1 and G2 at a particular site, the model in Eq. (5) was compared to that in Eq. (6). Similar comparisons were also performed to assess functional redundancy between the two legume species and between the two herb species.

$$\hat{y} = \hat{\beta}_G(P_{G1} + P_{G2}) + \hat{\beta}_{L1}P_{L1} + \hat{\beta}_{L2}P_{L2} + \hat{\beta}_{H1}P_{H1} + \hat{\beta}_{H2}P_{H2} + \hat{\alpha}X +$$
$$\hat{\delta}_{GL1}(P_{G1} + P_{G2})^{\hat{\theta}}P_{L1}^{\hat{\theta}} + \hat{\delta}_{GL2}(P_{G1} + P_{G2})^{\hat{\theta}}P_{L2}^{\hat{\theta}} +$$

$$\begin{aligned}
&\hat{\delta}_{GH1}(P_{G1}+P_{G2})^{\hat{\theta}}P_{H1}^{\hat{\theta}} + \hat{\delta}_{GH2}(P_{G1}+P_{G2})^{\hat{\theta}}P_{H2}^{\hat{\theta}} + \\
&\hat{\delta}_{L1L2}(P_{L1}P_{L2})^{\hat{\theta}} + \hat{\delta}_{L1H1}(P_{L1}P_{H1})^{\hat{\theta}} + \hat{\delta}_{L1H2}(P_{L1}P_{H2})^{\hat{\theta}} + \\
&\hat{\delta}_{L2H1}(P_{L2}P_{H1})^{\hat{\theta}} + \hat{\delta}_{L2H2}(P_{L2}P_{H2})^{\hat{\theta}} + \hat{\delta}_{H1H2}(P_{H1}P_{H2})^{\hat{\theta}}
\end{aligned} \tag*{Eq. (5)}$$

$$\begin{aligned}
\hat{y} = {} & \hat{\beta}_{G1}P_{G1} + \hat{\beta}_{G2}P_{G2} + \hat{\beta}_{L1}P_{L1} + \hat{\beta}_{L2}P_{L2} + \hat{\beta}_{H1}P_{H1} + \hat{\beta}_{H2}P_{H2} + \widehat{\alpha}X + \\
& \hat{\delta}_{G1G2}(P_{G1}P_{G2})^{\hat{\theta}} + \hat{\delta}_{L1L2}(P_{L1}P_{L2})^{\hat{\theta}} + \hat{\delta}_{H1H2}(P_{H1}P_{H2})^{\hat{\theta}} \\
& \hat{\delta}_{G1L1}(P_{G1}P_{L1})^{\hat{\theta}} + \hat{\delta}_{G1L2}(P_{G1}P_{L2})^{\hat{\theta}} + \hat{\delta}_{G1H1}(P_{G1}P_{H1})^{\hat{\theta}} + \hat{\delta}_{G1H2}(P_{G1}P_{H2})^{\hat{\theta}} + \\
& \hat{\delta}_{G2L1}(P_{G2}P_{L1})^{\hat{\theta}} + \hat{\delta}_{G2L2}(P_{G2}P_{L2})^{\hat{\theta}} + \hat{\delta}_{G2H1}(P_{G2}P_{H1})^{\hat{\theta}} + \hat{\delta}_{G2H2}(P_{G2}P_{H2})^{\hat{\theta}} + \\
& \hat{\delta}_{L1H1}(P_{L1}P_{H1})^{\hat{\theta}} + \hat{\delta}_{L1H2}(P_{L1}P_{H2})^{\hat{\theta}} + \hat{\delta}_{L2H1}(P_{L2}P_{H1})^{\hat{\theta}} + \hat{\delta}_{L2H2}(P_{L2}P_{H2})^{\hat{\theta}}
\end{aligned} \tag*{Eq. (6)}$$

If $\theta = 1$, the model in equation (5) is nested within the model in equation (6) and they can be compared using an F-test. However, if the estimate of θ differs between the two models, they are only partially nested and are compared using Vuong's test (*54*). If the model in equation (5) fits the data as well as or better than the model in equation (6), then there was evidence of functional redundancy for that pair of species at the site.

Software

Step 1 in the model selection process was carried out using the `DImodels` package (*30*) in R version 4.4.2 (*55*), while remaining statistical analyses were carried out using SAS software (*56*) (copyright © 2020 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA). The figures presented in the main text and supplementary materials were generated using the `ggplot2` (*57*), `ggmap` (*58*), `DImodelsVis` (*59*) and `PieGlyph` (*60*) R packages.

Site ID	Country	Lat.	Long.	Grasses: G1, G2	Legumes: L1, L2	Herbs: H1, H2	Sowing date	Duration (months)	Total number harvests	N moderate (kg ha^{-1} annum^{-1})	N high (kg ha^{-1} annum^{-1})	Average daily temp (°C)	Average daily precip (mm)
CA1	Canada	46.72	-71.49	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	29/05/2019	27	5	100	300	5.28	3.08
CA2	Canada	48.83	-72.53	*B. inermis,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	14/05/2021	27	4	100	300	2.11	2.21
CH1	Switzerland	47.26	8.32	*L. perenne,* *F. arundinacea*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	27/08/2020	19	6	150	350	7.99	2.45
CN1	China	37.92	114.70	*L. perenne,* *E. dahuricus*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	15/09/2020	21	5	100	300	12.85	1.66
CN2	China	46.54	125.98	*B. inermis,* *E. dahuricus*	*O. viciifolia,* *M. sativa*	*S. divaricata,* *P. asiatica*	19/05/2021	24	4	150	300	5.47	2.02
CN3	China	37.37	117.30	*B. inermis,* *E. dahuricus*	*O. viciifolia,* *M. sativa*	*C. intybus,* *S. divaricata*	20/09/2020	20	4	150	300	10.67	0.34
CZ1	Czech Republic	49.80	15.50	*F. krasanii,* *D. glomerata*	*L. corniculatus,* *M. sativa*	*C. intybus,* *P. lanceolata*	07/04/2020	42	6	80	180	10.09	1.47
DE1	Germany	54.30	10.10	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	14/09/2019	18	5	140	400	10.31	2.51
DE2	Germany	48.10	11.60	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	22/05/2020	22	6	100	250	9.87	2.39
DK1	Denmark	56.30	9.50	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	01/04/2020	24	7	150	300	10.33	2.15
FR1	France	46.20	2.20	*L. perenne,* *F. arundinacea*	*T. pratense,* *M. sativa*	*C. intybus,* *P. lanceolata*	25/08/2020	27	6	50	140	12.78	1.87
IE1	Ireland	52.16	-6.30	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	13/04/2017	35	14	150	300	10.61	3.05
IE2	Ireland	52.16	-6.30	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	05/09/2018	25	12	150	300	10.56	3.06
IE3	Ireland	52.26	-7.11	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	04/09/2020	20	6	100	360	11.08	2.27
IE4	Ireland	52.16	-6.30	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	23/08/2021	25	8	50	100	11.33	2.91
IT2	Italy	46.40	11.30	*L. perenne,* *D. glomerata*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	10/09/2021	25	7	106.5	213	11.22	2.74
NL1	The Netherlands	52.13	6.29	*L. perenne,* *F. arundinacea*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	03/10/2019	42	14	150	300	11.44	2.12
NL2	The Netherlands	52.13	5.29	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	15/06/2020	21	8	150	300	7.11	2.38
NO1	Norway	63.91	11.44	*L. perenne,* *P. pratense*	*T. pratense,* *T. repens*	*C. intybus,* *P. lanceolata*	16/06/2020	23	3	150	300	6.47	2.22

Site ID	Country	Lat.	Long.	Grasses: G1, G2	Legumes: L1, L2	Herbs: H1, H2	Sowing date	Duration (months)	Total number harvests	N moderate (kg ha^{-1} annum^{-1})	N high (kg ha^{-1} annum^{-1})	Average daily temp (°C)	Average daily precip (mm)
NO2	Norway	59.40	10.51	*P. pratense, L. perenne*	*T. pratense, L. corniculatus*	*C. intybus, P. lanceolata*	20/05/2022	24	4	100	200	8.07	1.95
NO3	Norway	63.91	11.44	*L. perenne, P. pratense*	*T. pratense, T. repens*	*C. intybus, P. lanceolata*	28/05/2021	23	3	150	300	7.90	3.27
NZ1	New Zealand	-43.40	172.30	*L. perenne, P. pratense*	*T. pratense, T. repens*	*C. intybus, P. lanceolata*	26/10/2021	18	6	0	190	13.41	3.02
PL1	Poland	51.51	17.09	*L. perenne, P. pratense*	*T. pratense, T. repens*	*C. intybus, P. lanceolata*	28/08/2020	24	8	120	240	9.96	1.18
UK1	United Kingdom	51.50	0.98	*L. perenne, P. pratense*	*T. pratense, T. repens*	*C. intybus, P. lanceolata*	26/03/2019	24	4	100	250	11.16	2.45
US1	USA	44.95	-93.09	*L. perenne, P. pratense*	*T. pratense, M. sativa*	*C. intybus, P. lanceolata*	21/08/2020	22	4	0	100	4.66	1.78
US2	USA	41.20	-77.19	*L. perenne, P. pratense*	*T. pratense, T. repens*	*C. intybus, P. lanceolata*	08/04/2021	18	7	80	200	12.42	3.17

Table S1. **Site information table.** For each site the following information is provided: unique site identification code (site ID), country, latitude, longitude, the names of the six species sown at the site (G1, G2, L1, L2, H1, H2), the date of sowing of the experiment, the grassland phase duration, the rates of moderate and high nitrogen fertilizer, the average daily temperature and precipitation at the site over the experimental period. Short-hand notation for the sown species: *L. perenne - Lolium perenne, P. pratense - Phleum pratense, B. inermis - Bromus inermis, F. arundinacea - Festuca arundinacea, E. dahuricus - Elymus dahuricus, F. krasanii - Festulolium krasanii, D. glomerata - Dactylis glomerata, T. pratense - Trifolium pratense, T. repens - Trifolium repens, O. viciifolia - Onobrychis viciifolia, M. sativa - Medicago sativa, L. corniculatus - Lotus corniculatus, C. intybus - Cichorium intybus, P. lanceolata - Plantago lanceolata, S. divaricata - Saposhnikovia divaricata, P. asiatica – Plantago asiatica.*

Comm	Reps	N	G	L	H	G1	G2	L1	L2	H1	H2
1	3	Moderate	1	0	0	1	0	0	0	0	0
2	3	Moderate	1	0	0	0	1	0	0	0	0
3	3	Moderate	0	1	0	0	0	1	0	0	0
4	3	Moderate	0	1	0	0	0	0	1	0	0
5	3	Moderate	0	0	1	0	0	0	0	1	0
6	3	Moderate	0	0	1	0	0	0	0	0	1
7	1	Moderate	1	0	0	0.5	0.5	0	0	0	0
8	1	Moderate	0	1	0	0	0	0.5	0.5	0	0
9	1	Moderate	0	0	1	0	0	0	0	0.5	0.5
10	1	Moderate	0.5	0.5	0	0.5	0	0.5	0	0	0
11	1	Moderate	0.5	0.5	0	0.5	0	0	0.5	0	0
12	1	Moderate	0.5	0	0.5	0.5	0	0	0	0.5	0
13	1	Moderate	0.5	0	0.5	0.5	0	0	0	0	0.5
14	1	Moderate	0.5	0.5	0	0	0.5	0.5	0	0	0
15	1	Moderate	0.5	0.5	0	0	0.5	0	0.5	0	0
16	1	Moderate	0.5	0	0.5	0	0.5	0	0	0.5	0
17	1	Moderate	0.5	0	0.5	0	0.5	0	0	0	0.5
18	1	Moderate	0	0.5	0.5	0	0	0.5	0	0.5	0
19	1	Moderate	0	0.5	0.5	0	0	0.5	0	0	0.5
20	1	Moderate	0	0.5	0.5	0	0	0	0.5	0.5	0
21	1	Moderate	0	0.5	0.5	0	0	0	0.5	0	0.5
22	1	Moderate	0.33	0.33	0.33	0.33	0	0.33	0	0.33	0
23	1	Moderate	0.33	0.33	0.33	0.33	0	0.33	0	0	0.33
24	1	Moderate	0.33	0.33	0.33	0.33	0	0	0.33	0.33	0
25	1	Moderate	0.33	0.33	0.33	0.33	0	0	0.33	0	0.33
26	1	Moderate	0.33	0.33	0.33	0	0.33	0.33	0	0.33	0
27	1	Moderate	0.33	0.33	0.33	0	0.33	0.33	0	0	0.33
28	1	Moderate	0.33	0.33	0.33	0	0.33	0	0.33	0.33	0
29	1	Moderate	0.33	0.33	0.33	0	0.33	0	0.33	0	0.33
30	1	Moderate	0.5	0.5	0	0.25	0.25	0.25	0.25	0	0
31	1	Moderate	0.5	0	0.5	0.25	0.25	0	0	0.25	0.25
32	1	Moderate	0	0.5	0.5	0	0	0.25	0.25	0.25	0.25
33	3	Moderate	0.33	0.33	0.33	0.17	0.17	0.17	0.17	0.17	0.17
34	5	High	1	0	0	1	0	0	0	0	0

Table S2. **Experimental design of the grassland communities established in the LegacyNet grassland phase.** The variables are: a community identifier for each set of unique sown species proportions by N fertilizer level (comm), the number of replicate plots sown per community (reps); amount of N fertilizer applied (N: moderate or high); the sown proportions of each functional group (G: grasses, L: legumes, and H: herbs); and the sown proportions of each species (G1 to H2).

	Parameter	Variable	Fixed effect	Std. Error	p-value
Identity effects	β_1	Grass 1	8.61	0.568	
	β_2	Grass 2	8.38	0.554	
	β_3	Legume 1	11.47	0.759	
	β_4	Legume 2	8.89	0.807	
	β_5	Herb 1	9.19	0.672	
	β_6	Herb 2	8.51	0.596	
Interaction effects	δ_{GL}	Grass-legume	5.92	0.563	**< 0.0001**
	δ_{GH}	Grass-herb	1.67	0.343	**0.0002**
	δ_{HL}	Herb-legume	4.80	0.410	**< 0.0001**
	δ_{GG}	Grass-grass	-0.27	0.498	0.5895
	δ_{LL}	Legume-legume	1.88	0.510	**0.0002**
	δ_{HH}	Herb-herb	0.35	0.500	0.4798
	θ	Non-linear parameter	0.7816041	(Estimated using profile likelihood during the model selection process)	
Climate effects	γ	Centered mean daily temperature squared	-0.049	0.0395	**0.2283**
	τ_{GL}	Grasses-legumes × centered mean daily temperature	0.36	0.188	**0.0486**
	τ_{HL}	Herbs-legumes × centered mean daily temperature	0.45	0.128	**0.0045**
	α	High nitrogen grass	11.07	0.620	

Table S3. Diversity-Interactions model fitted to yield across sites. Fixed effect estimates, standard errors, and p-values from the final model. The model selection process showed that the fixed interaction effects were dictated by functional group membership, and that the theta parameter, θ, was significantly different from one (and thus incorporated a non-linear form of the species interactions, see Fig. S1). The parameters are grouped by species identity effects (beta values, β), species interaction effects (delta values, δ, plus theta, θ), climate effects (gamma, γ, and tau, τ), and the high N grass (alpha, α).

A

Effect	G1	G2	L1	L2	H1	H2	GH	GL	HL	N High
G1	5.6273	4.6230	4.6230	4.6230	4.6230	4.6230	-0.0139	-0.0139	-0.0139	4.9099
G2	4.6230	5.1997	4.6230	4.6230	4.6230	4.6230	-0.0139	-0.0139	-0.0139	4.9099
L1	4.6230	4.6230	12.1812	4.6230	4.6230	4.6230	-0.0139	-0.0139	-0.0139	4.9099
L2	4.6230	4.6230	4.6230	14.1401	4.6230	4.6230	-0.0139	-0.0139	-0.0139	4.9099
H1	4.6230	4.6230	4.6230	4.6230	8.9513	4.6230	-0.0139	-0.0139	-0.0139	4.9099
H2	4.6230	4.6230	4.6230	4.6230	4.6230	6.4500	-0.0139	-0.0139	-0.0139	4.9099
GH	-0.0139	-0.0139	-0.0139	-0.0139	-0.0139	-0.0139	1.5450	1.3532	1.3532	0.9957
GL	-0.0139	-0.0139	-0.0139	-0.0139	-0.0139	-0.0139	1.3532	6.3869	1.3532	0.9957
HL	-0.0139	-0.0139	-0.0139	-0.0139	-0.0139	-0.0139	1.3532	1.3532	2.6589	0.9957
N High	4.9099	4.9099	4.9099	4.9099	4.9099	4.9099	0.9957	0.9957	0.9957	7.2578

B

Site ID	Residual variance	Site ID	Residual variance
CA1	0.7202	IE3	1.3209
CA2	0.5067	IE4	2.9323
CH1	3.0575	IT2	1.2997
CN1	4.4206	NL1	0.5132
CN2	1.4423	NL2	0.3704
CN3	4.4031	NO1	0.7990
CZ1	1.8650	NO2	1.9568
DE1	0.7898	NO3	0.4991
DE2	3.1946	NZ1	1.6897
DK1	0.5641	PL1	0.6404
FR1	1.2467	UK1	3.5194
IE1	0.3669	US1	0.4218
IE2	0.6038	US2	0.3689

Table S4. **Variance parameter estimates for (A) the random coefficients and (B) the within-site error terms.** (**A**) The variance-covariance parameter estimates for the random site-to-site effects. (**B**) The estimated within-site residual error variance for each site.

	Conventional practices		Mixtures			
Site	**High N**	**70:30 G1:L2**	**GH**	**LH**	**GL**	**GLH**
Across sites	11.07	10.44	9.45	11.97	12.23	12.31
CA1	7.41	6.70	5.08	7.23	8.41	7.73
CA2	9.73	9.82	8.27	7.34	10.10	9.00
CH1	12.01	12.36	8.56	13.3	15.51	14.14
CN1	14.04	12.43	15.93	18.75	14.97	17.78
CN2	4.71	5.23	4.26	6.00	5.63	5.35
CN3	10.11	15.26	10.36	19.03	18.02	17.44
CZ1	13.49	10.45	12.79	9.33	12.40	12.01
DE1	14.14	10.18	9.88	9.82	10.39	11.03
DE2	13.24	11.82	10.93	14.73	14.69	14.86
DK1	10.86	9.92	8.75	10.62	12.13	11.52
FR1	3.90	4.47	4.15	8.31	7.38	7.33
IE1	10.54	10.12	9.83	12.29	11.39	11.98
IE2	9.10	10.12	8.02	12.49	13.13	13.01
IE3	10.47	7.97	7.93	10.64	10.33	10.48
IE4	9.83	10.55	12.19	13.16	12.66	13.89
IT2	11.24	11.66	10.74	13.90	14.30	13.86
NL1	12.22	10.07	9.13	11.02	10.92	11.20
NL2	12.54	10.56	9.79	11.62	11.44	11.63
NO1	12.54	10.36	9.67	10.45	10.32	11.10
NO2	9.93	10.77	4.71	12.60	12.98	12.09
NO3	12.70	12.20	11.49	11.69	14.01	13.40
NZ1	13.88	11.19	7.48	13.57	14.56	13.74
PL1	12.60	12.31	12.42	14.38	13.32	14.92
UK1	11.54	10.55	12.57	13.17	11.17	14.69
US1	7.88	7.27	7.48	10.64	9.21	8.85
US2	5.87	6.63	4.52	9.75	10.63	9.86

Table S5. **Average and site-level predicted yields (t ha^{-1}) for selected communities.** The high N grass and the two-species 70:30 G1:L2 communities reflect management practices that are widespread in temperate sown grasslands. The predictions across sites (first row) are calculated using the fixed effects estimates from the final model (Table S3), for example, the predicted yield of the four-species equi-proportional grass-legume mixture (GL) is:

$$12.23 = 8.61\left(\frac{1}{4}\right) + 8.38\left(\frac{1}{4}\right) + 11.47\left(\frac{1}{4}\right) + 8.89\left(\frac{1}{4}\right) + 5.92\left[4 \times \left(\frac{1}{4} \times \frac{1}{4}\right)^{0.7816041}\right] - 0.27\left(\frac{1}{4} \times \frac{1}{4}\right)^{0.7816041} + 1.88\left(\frac{1}{4} \times \frac{1}{4}\right)^{0.7816041}.$$

Similarly, we can predict from the model for any combination of sown proportions of G1, G2, L1, L2, H1 and H2, with moderate N fertilizer management. The predicted yield of the high N grass across sites is 11.07; note that we cannot predict yield for any other monoculture or mixture at high N, however, we can construct contrasts using our model to compare any monoculture or mixture at moderate N to the high N grass. Predictions at the site level are calculated by fitting the final model (excluding climate variables) to the data from each individual site.

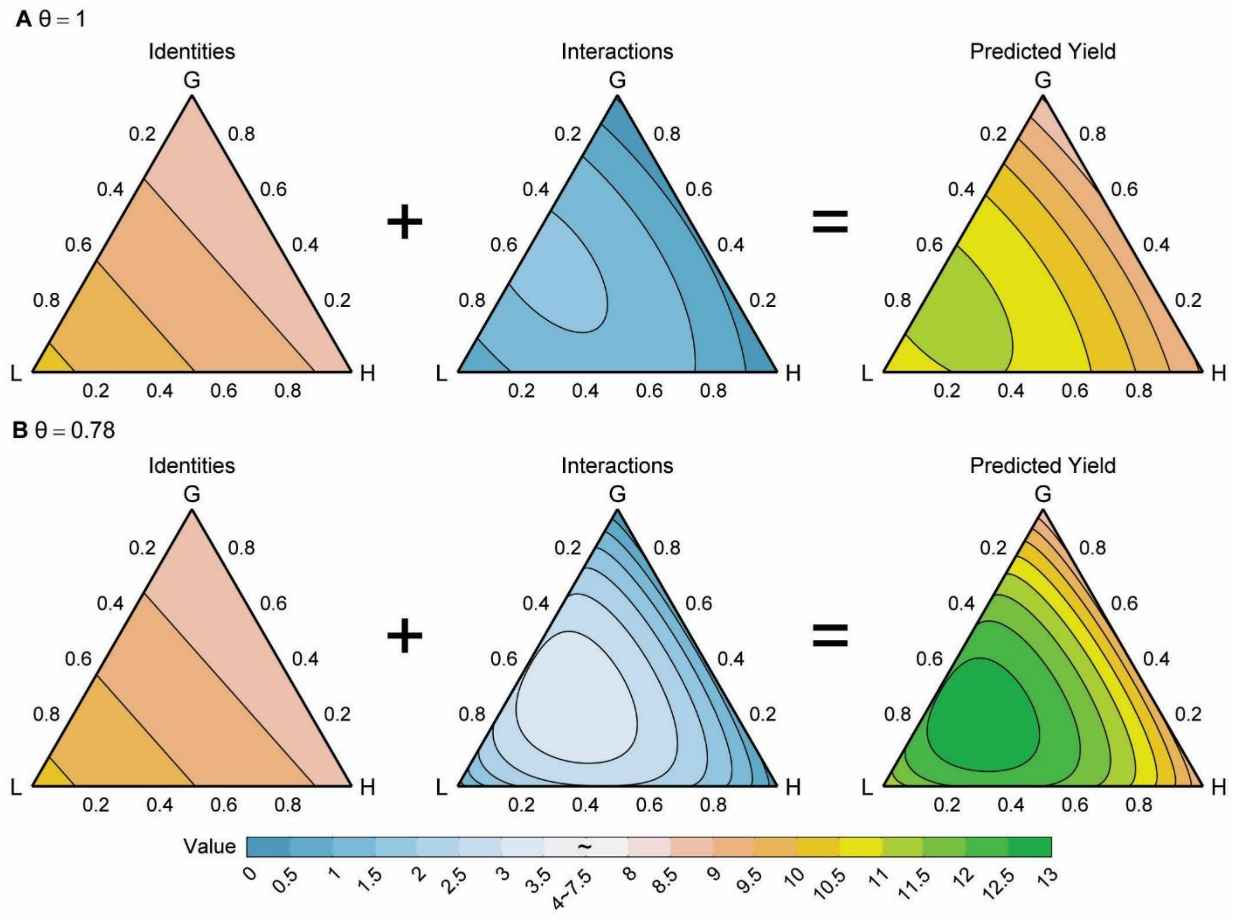


Fig. S1. Effect of the non-linear theta (θ) parameter in the final fitted Diversity-Interactions model. In each row, predictions from the model are split into a component for the net identity effects (first ternary, 'Identities'), a component for the net interactions (second ternary, 'Interactions') and the combined predictions (third ternary, 'Predicted Yield') as the sown proportions of grass (G), legume (L) and herb (H) vary. Only the value of θ differs between the two rows: **(A)** θ is equal to 1, and **(B)** θ is equal to 0.7816041, the estimate of θ in the final model. The legend is truncated between 4 and 7.5 for ease of readability.

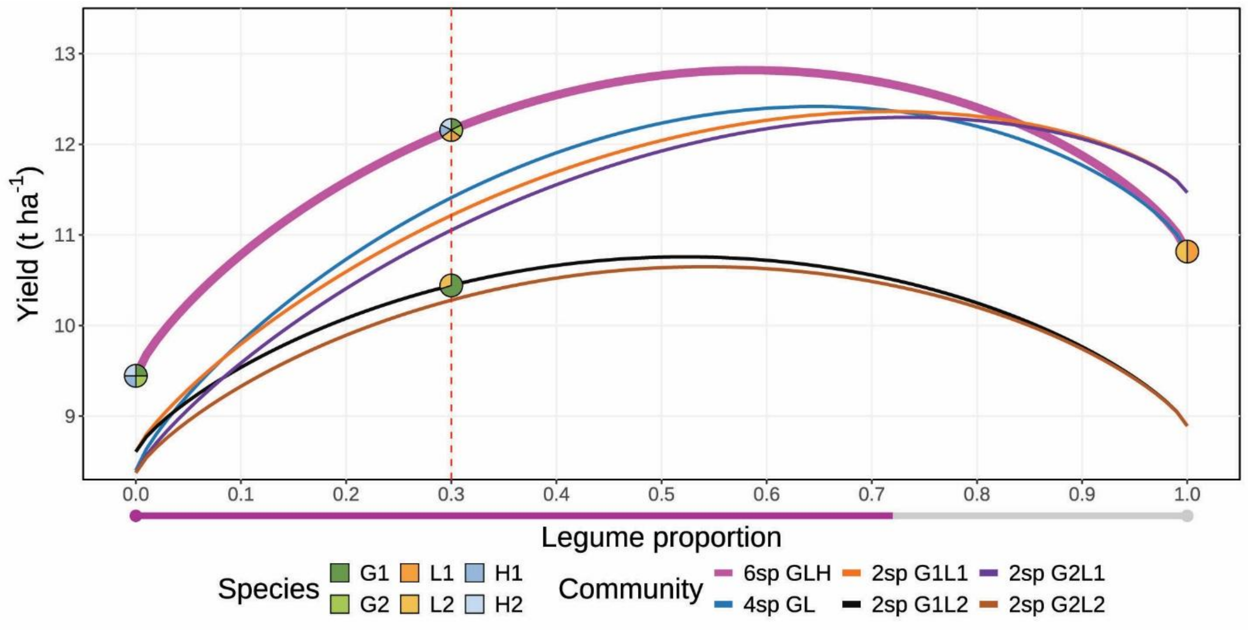


Fig. S2 At the same proportion of legumes, six-species mixtures of grasses, legumes, and herbs performed better than two- and four-species grass-legume communities with sown legume proportions < 0.7. Curves show the predicted yield of the six-species grass-legume-herb communities (bold magenta curve) and two- and four-species grass-legume communities versus sown legume functional group proportions (varying from 0 to 1). For the four- and six-species communities, species within each functional group have equal sown proportions, e.g., for sown legume proportion of 0.4, these communities would be: 0.3: 0.3: 0.2: 0.2: 0: 0 and 0.15: 0.15: 0.2: 0.2: 0.15: 0.15 respectively. The dashed red vertical line indicates the sown legume proportion = 0.3, with pie-glyphs highlighting the six-species mixture 0.175: 0.175: 0.15: 0.15: 0.175: 0.175 and the two-species 0.7:0.3 G1:L2 mixture. The pie-glyphs on the magenta line illustrate how the sown species proportions vary across the curve (noting that the lower end has legume = 0 and is a grass-herb equi-proportional four-species mixture, while the upper end has legume = 1 and is a 50:50 L1:L2 mix, while all other points on the curve are six-species mixtures). The magenta line underneath the x-axis indicates the range of sown legume proportion values for which the six-species curve performed better than all other curves (tested at each point vertically along the sown legume proportion gradient). While not shown, the six-species mixtures curve is significantly higher than the G1L2 and G2L2 curves at all sown legume proportions.

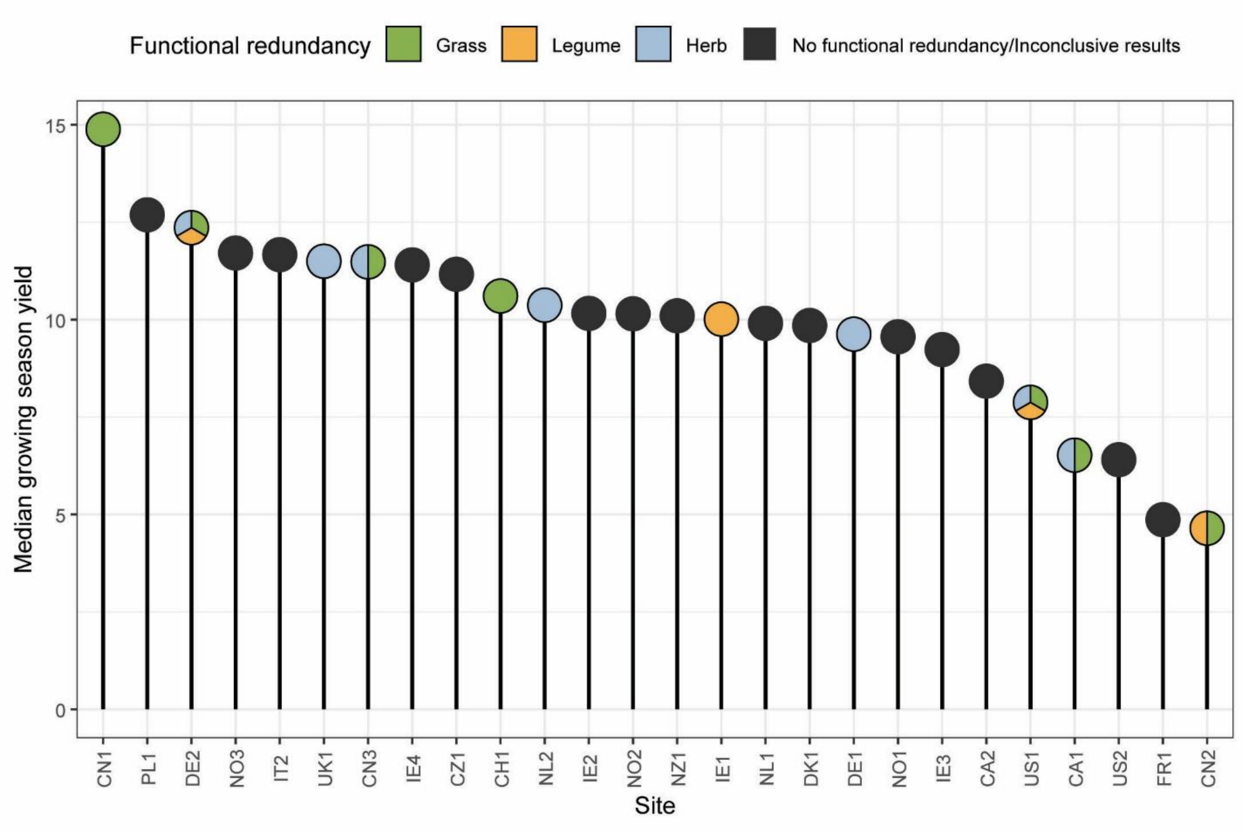
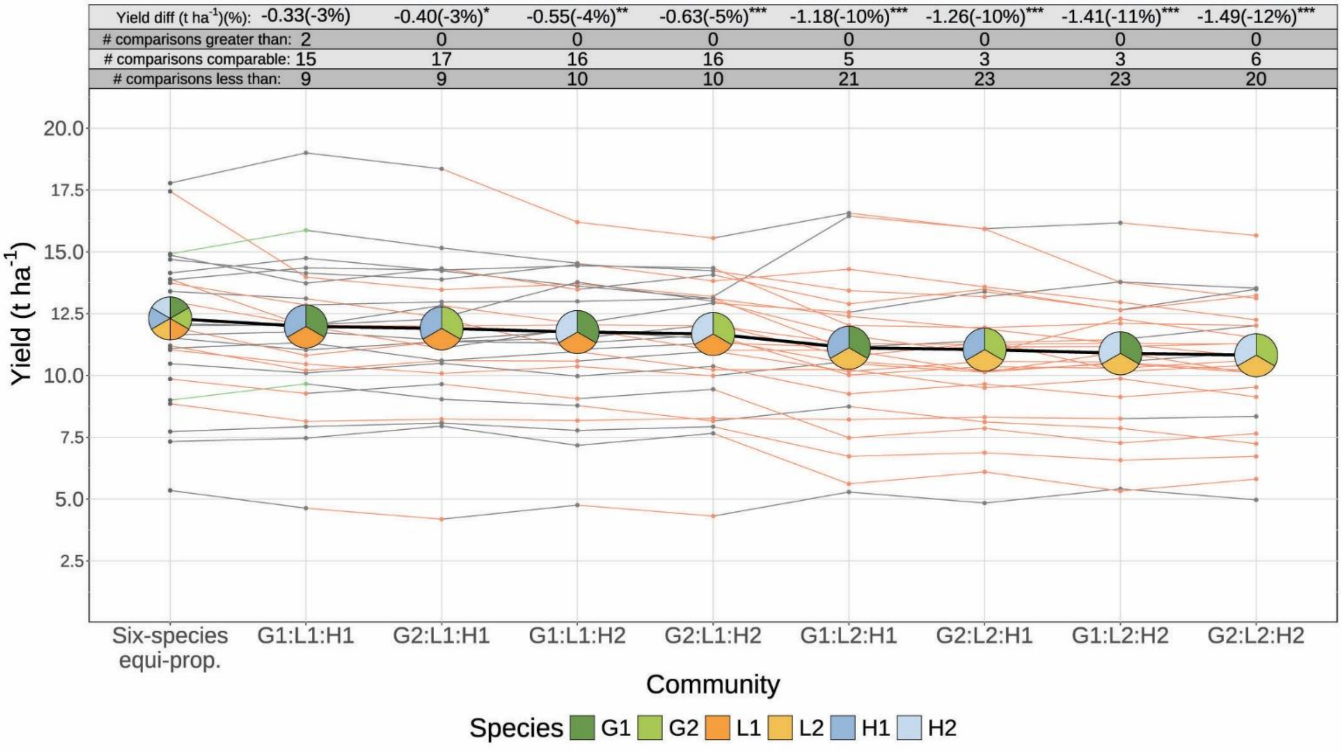


Fig. S3. Functional redundancy within functional groups. The presence/absence of functional redundancy between the two grass (green), the two legume (orange), and the two herb (blue) species at each site is shown using pie-glyphs. The pie-glyphs with green, orange and/or blue in them indicate the presence of functional redundancy between the species from the respective functional groups, while those colored in dark grey represent sites with no redundancy (or inconclusive results) between the pair of species in each functional group (details in (*27*)). The sites along the x-axis are arranged in decreasing order of their median yields.



Yield diff (t ha^{-1})(%):	-0.33(-3%)	-0.40(-3%)*	-0.55(-4%)**	-0.63(-5%)***	-1.18(-10%)***	-1.26(-10%)***	-1.41(-11%)***	-1.49(-12%)***
# comparisons greater than:	2	0	0	0	0	0	0	0
# comparisons comparable:	15	17	16	16	5	3	3	6
# comparisons less than:	9	9	10	10	21	23	23	20

Fig. S4. Predicted yields of equi-proportional three-species mixtures were all (except one) lower than the yield of the six-species equi-proportional mixture. Predicted yields of the six-species equi-proportional mixture and each possible equi-proportional three-species three-functional-group mixture, across sites (black line and pie-glyphs that illustrate the sown species proportions) and at each site (colored lines and smaller points). There are eight possible equi-proportional three-species three-functional-group combinations (G1:L1:H1 to G2:L2:H2 labels across the x-axis) and they are arranged in order of decreasing predicted yield across sites. Predictions at site level are calculated by fitting the final model (excluding climate variables) to the data from each individual site. The predictions from each individual site are joined by lines; both the color of a site prediction and the line to the left of it for three-species mixtures are green if the prediction was significantly greater (+2SE) than the six-species mixture, grey if it was comparable (within ± 2SE), and orange if significantly less (-2SE). Across sites, the difference in yield between each three-species mixture and the six-species mixture is shown at the top of the figure (with significance indicated by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). The number of sites where the three-species mixture combination was greater than, comparable to, and less than the six-species mixture at the site is also listed at the top of the graph. A total of 208 site-level comparisons were made (26 sites * 8 three-species communities).

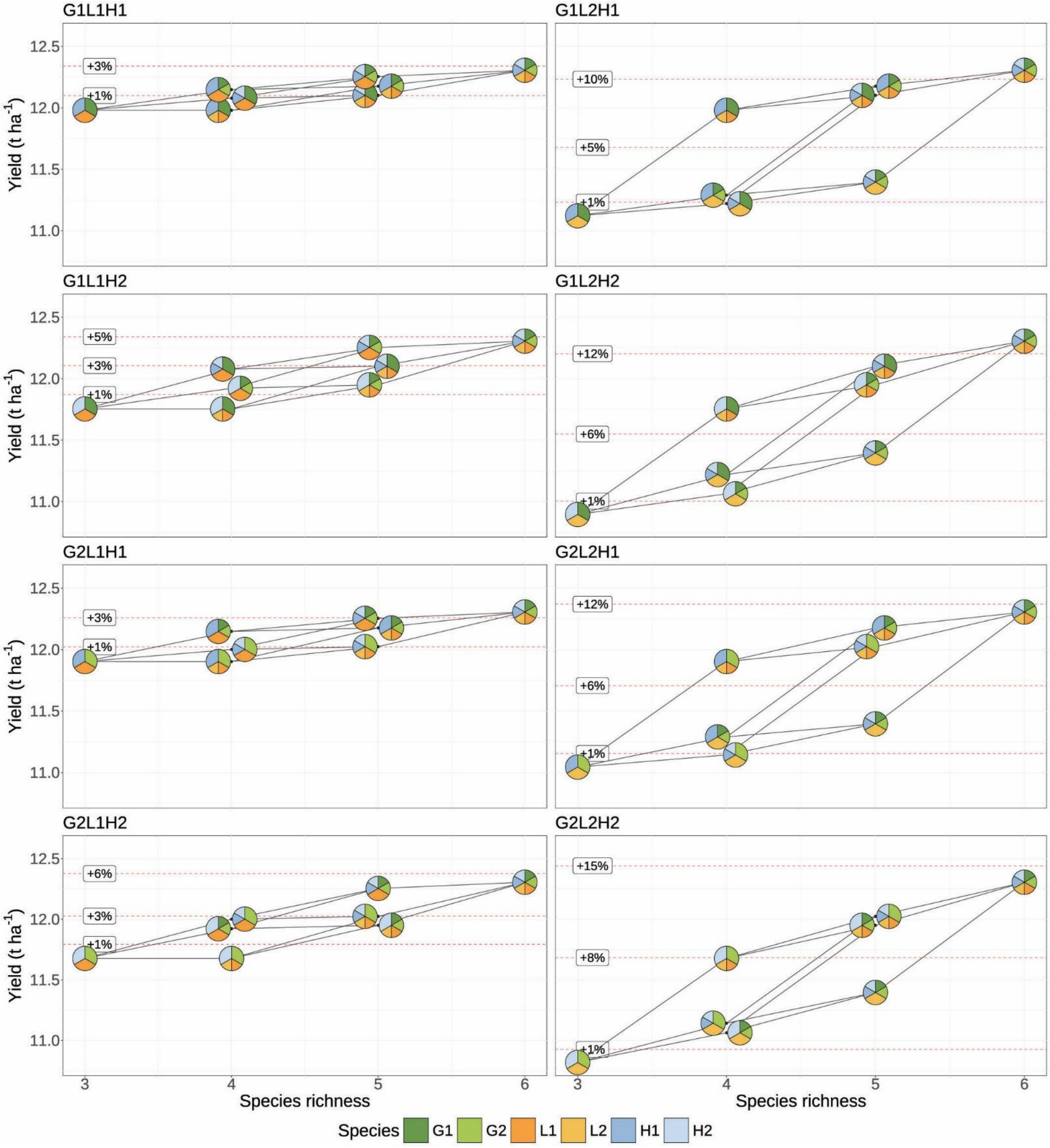


Fig S5 Adding one species at a time generally increased yield across the richness gradient for mixtures that each contained at least one grass, one legume, and one herb, with the magnitude depending on species identity. Each panel shows the predicted yield of three-, four-, five-, and six-species mixtures that each contained at least one grass, one legume, and one herb versus richness, with pie-glyphs illustrating the sown species proportions. For each mixture, the three sown functional group proportions are always equal to 1/3 each, and when there are two species within a functional group, their proportions are both equal to 1/6. The eight panels are distinguished by the combination of species sown in the three-species mixture on the left-hand side (see panel title). Dotted red lines in each panel indicate percentage level increases relative to the three-species mixture on the left-hand side. Moving from left to right within each panel, solid black lines trace the change in predicted yield when species richness is increased by adding one extra species to each functional group at a time. Moving from right to left illustrates the effects of species removals on yields (i.e., indicative extinction pathways). A jitter has been introduced on the pie-glyphs to prevent overplotting.

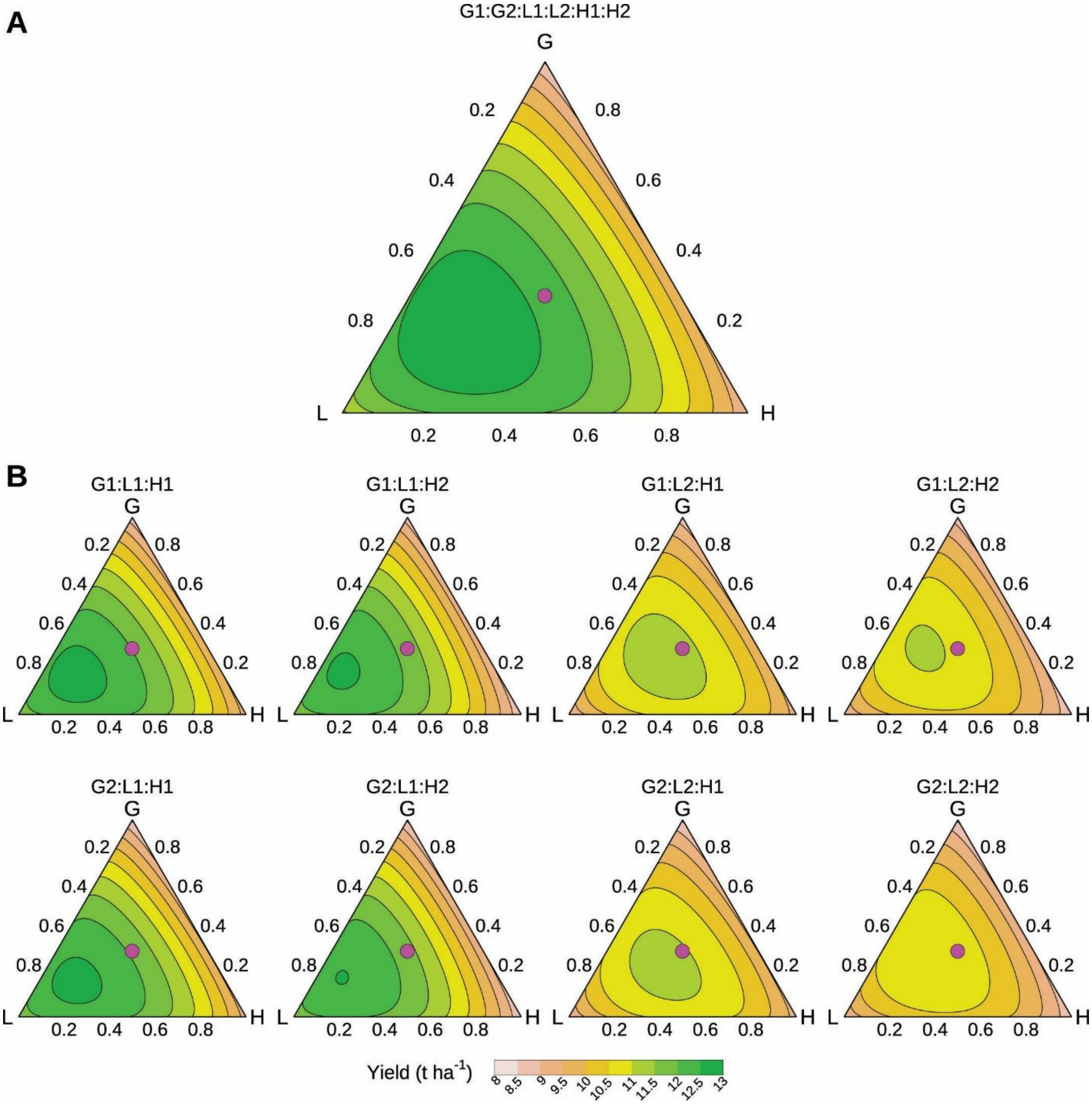


Fig. S6. Sowing two species, rather than one, within each functional group leads to higher or comparable yields. Each ternary diagram shows the predicted yield of communities as the sown proportion of the three functional groups varies. In **(A)** two species are sown in equal proportion to each other within the three functional groups, G = grasses, L = legumes, H = herbs. In **(B),** only one species is sown in each of the three functional groups. The title of each ternary diagram indicates the three species that were sown (G1 to H2). In each ternary, the equi-proportional mixture is shown as a magenta circle (this is a six-species mixture in **A** and a three-species mixture in each ternary in **B**).

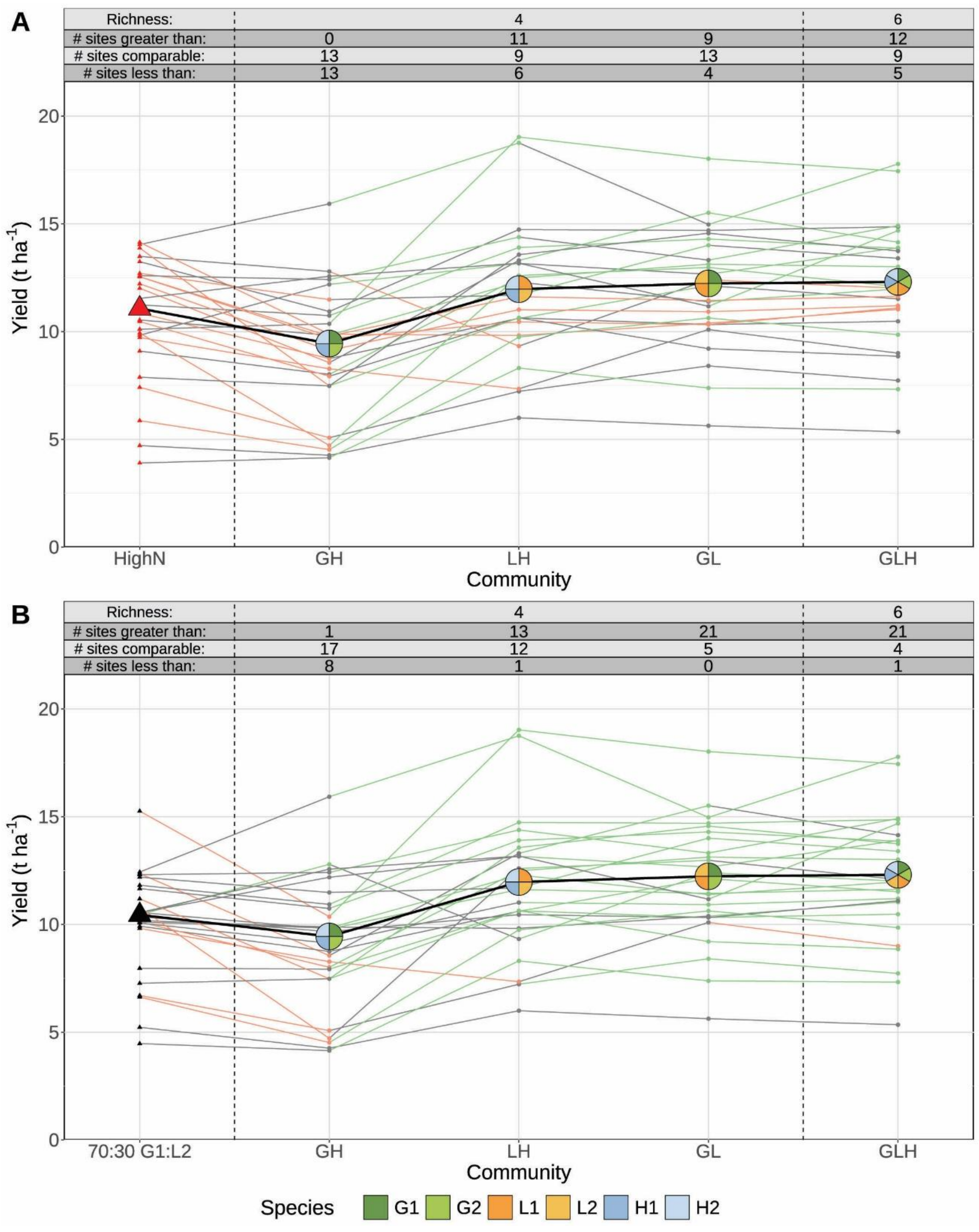


Fig. S7. Yields of four- and six-species equi-proportional mixtures were generally greater than, or comparable to, the high N grass and the 70:30 G1:L2 community at individual sites, despite high site-to-site variability. Predicted yields of selected communities across sites (black line and pie-glyphs) and at each site (colored lines and smaller points): the high N grass monoculture (red triangle), the 70:30 G1:L2 community (black triangle); the four-species equi-proportional grass-herb (GH), legume-herb (LH) and grass-legume (GL) mixtures; and the six-species equi-proportional GLH mixture (1/6:1/6: 1/6:1/6: 1/6:1/6 for G1:G2: L1:L2: H1:H2). The four species communities are arranged in order of increasing predicted yield across all sites (calculated from estimates in Table S3). The communities are joined by lines to show the predictions specific to each site (values in Table S5); both the color of a site prediction and the line to the left of it are green if the prediction was significantly greater (+2SE) than the reference community (left-hand side of each panel), grey if it was comparable (within ± 2SE), and orange if significantly less (-2SE). The predicted comparison community is represented by a triangle: (**A**) red for the high N grass monoculture, and (**B**) black for the 70:30 G1:L2 two-species community over all sites (large) and for each individual site (small). The number of sites where the comparison is greater than, comparable to, or less than the reference community is listed along the top row for each mixture community. Site-level predictions were made using the DI model fitted to data from each site separately.

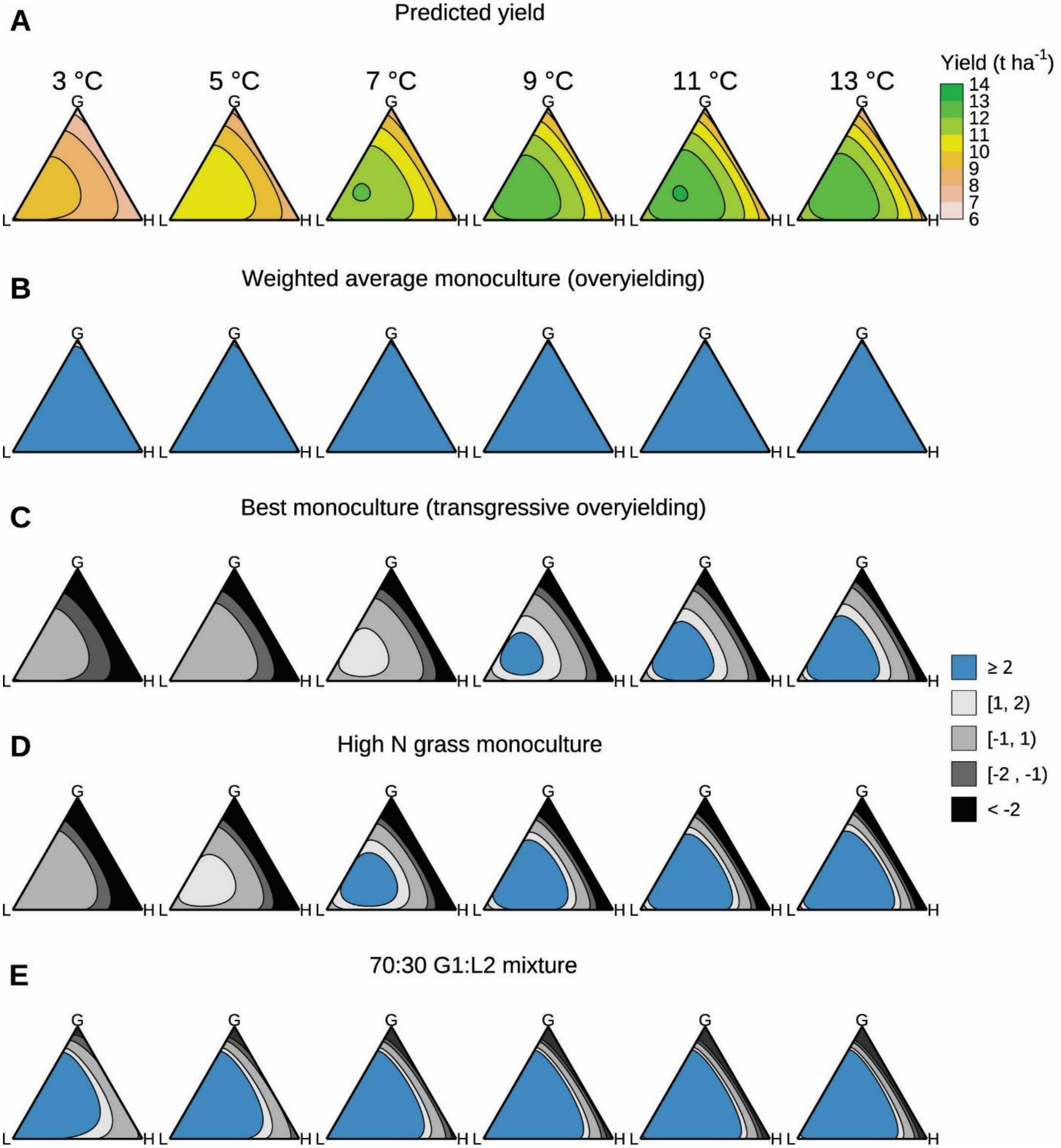


Fig. S8. Across the temperature gradient, mixture communities performed strongly in comparison to selected communities. Across the daily temperature gradient from 3°C to 13°C, the ternary diagrams show the predicted yield (row A) and the regions where mixtures performed significantly better than the weighted average monoculture (row B, overyielding), the best-performing monoculture (row C, transgressive overyielding), the high N grass monoculture (row D), and the 70:30 G1:L2 community (row E).

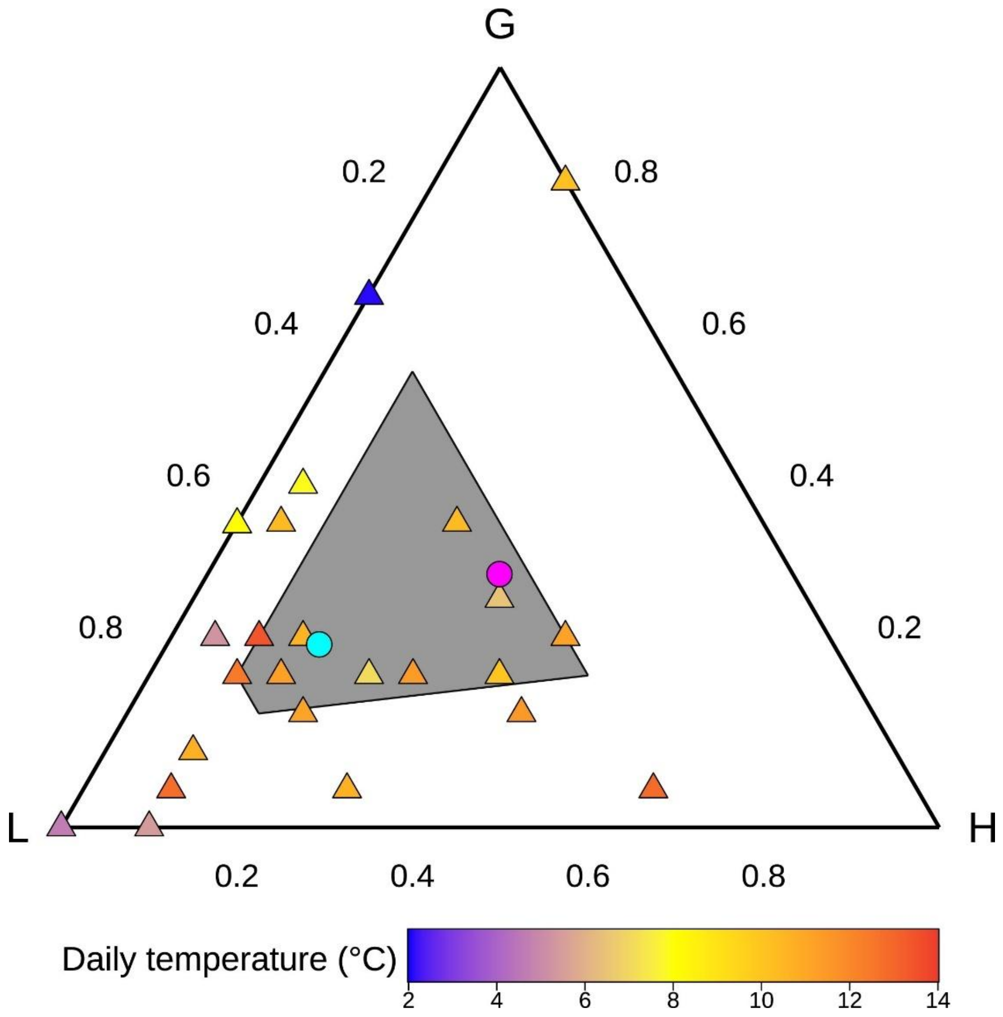


Fig S9. The best-performing mixtures at each site lie within, or close to, the optimal region. The sown functional group proportions for the 'best mixture' from each site are represented by the location of triangle symbols, with each triangle colored according to the average daily temperature of the site. The cyan and magenta circles highlight the sown functional group proportions (G:L:H = 0.24:0.59:17) that gave the highest predicted yield (12.83 t ha^{-1}) and the six-species equi-proportional mixture (12.31 t ha^{-1}), respectively.