**Methods**

*Study sites and experimental design*

The study sites are part of The Nutrient Network (NutNet) experiment (Table S2; http://nutnet.org/) (Borer et al. 2014, 2017). Plots at each site are 5 × 5 m separated by at least 1 m and grouped into spatial blocks typically spread over 320 m2. Treatments were randomly assigned to the 25-m2 plots and were replicated in three blocks at most sites, although the number of blocks ranged from one to six between sites. Sampling was done in 1 m2 subplots and followed a standardized protocol at all sites (Borer et al. 2014). All sites included in the analyses presented here included unmanipulated plots and fertilized plots with nitrogen (N), phosphorus (P) and potassium and micronutrients (K) added in combination (NPK). N, P and K were applied annually before the beginning of the growing season at relatively high rates: 10 gm-2 y-1. N was supplied as time-release urea ((NH2)2CO) or ammonium nitrate (NH4NO3). P was supplied as triple super phosphate (Ca(H2PO4)2), and K as potassium sulphate (K2SO4). In addition, a micronutrient mix (Fe, S, Mg, Mn, Cu, Zn, B and Mo) was applied at 100 gm-2 y-1 to the K-addition plots, once at the start of the experiment but not in subsequent years to avoid toxicity.

*Site selection*

Data were retrieved on 1 November 2019. To keep a constant number of communities per meta-community and treatment, we selected sites that had a minimum of three blocks per treatment and subsetted sites that included more than three blocks, usually by using the first three blocks (Table S2). Thus, for each site and treatment, a metacommunity was composed of three communities (three replicated blocks per treatment and site). Sites spanned a broad envelope of seasonal variation in precipitation and temperature (Fig. S1) and represent a wide range of grassland types including alpine, desert and semi-arid grasslands, prairies, old fields, pastures, savanna, tundra and shrub-steppe (Table S2). We selected sites that had a minimum of four years, and up to nine years of post-treatment data. Treatment application started at most sites in 2008 but some sites started later resulting in a lower number of sites with increasing duration of the study, from 48 sites with four years of duration to 23 sites with nine years of duration (Table S2). Longer time series currently exist but for a limited number of sites within our selection criteria (15 sites). The data set described above contains all the available sites per period of experimental duration. This means that the number of sites decreases with increasing period of experimental duration, making these variables partly confounded. We thus generated an additional data set that contained the same metacommunities (a total of 23 metacommunities for each treatment) for each period of experimental duration. Results of our analyses using a fixed number of metacommunities per period did not differ qualitatively from the results presented in the text using all the available sites per period (Fig. S2).

*Primary productivity and cover*

Primary productivity was estimated annually by clipping at ground level all aboveground live biomass from two 0.1 m2 (10 x 100 cm) quadrats per plot. For shrubs and subshrubs, leaves and current year’s woody growth were collected. Biomass was dried to constant mass at 60°C and weighed to the nearest 0.01 g. Areal percent cover of each species was measured concurrently with primary productivity in 1 x 1m subplots in which no destructive sampling occurred. Cover was visually estimated to the nearest percent independently for each species, so that total summed cover can exceed 100% for multilayer canopies. To better match theory and because direct measures of biomass for each individual species were unavailable, we re-run analyses with species-level biomass estimates based on percentage cover. Percentage cover was converted to biomass estimates for each species by assuming that the proportion of total cover for each species was equivalent to its proportion of total aboveground biomass. Results of our analyses using biomass estimates did not differ qualitatively from the results presented in the text using biomass estimates based on percentage cover (Figures S3).

*Diversity, asynchrony and stability across spatial scales*

Species richness was quantified as the average number of plant species in standard 1-m2 plots over the duration of post-treatment data. Species evenness was quantified in the same plots as the average Pielou’s evenness by taking the average of the ratio of the Shannon–Weaver index and the natural logarithm of species richness. Gamma diversity was quantified as the average number of plant species across the three replicated standard 1-m2 plots. Following theoretical models (Wang et al. 2016 EL), richness- and abundance-based beta diversity indices were quantified for each metacommunity and duration period as the multiplicative partitioning of gamma diversity: beta = gamma/alpha (Whittaker 1972). For richness-based beta diversity, gamma diversity is defined as the number of species over the whole metacommunity and alpha diversity as the average number of species within a metacommunity. For abundance-based beta diversity, gamma diversity is defined as the inverse Simpson index over the whole metacommunity and alpha diversity as the inverse of average Simpson indices within a metacommunity (Olszewski 2004). We used richness- and abundance-based beta diversity indices because they are directly linked to ecosystem stability in theoretical models and thus directly comparable to theories (Wang et al 2014, 2016). Results of our analyses using inverse Simpson and richness-based beta diversity did not differ qualitatively from the results presented in the text using species richness and abundance-based beta diversity (Figures S5). We used the R functions ‘diversity’, ‘specnumber’, and ‘vegdist’ to calculate Shannon-Weaver, Simpson and species richness indices within and across replicated plots (Ref to vegan). We used the R function ‘var.partition’ to calculate asynchrony and stability across spatial scales (Wang et al Ecography 2019).

*Climate data*

Precipitation and temperature seasonality were estimated for each site using the long-term coefficient of variation of precipitation (MAP\_VAR) and temperature (MAT\_VAR) respectively derived from the WorldClim Global Climate database (version 1.4; http://www.worldclim.org/) ().

*Analyses*

All analyses were conducted in R 3.5.1 (ref to cran). To improve normality, stability and asynchrony measures were logarithm transformed before analyses.

First, we used bivariate analyses and linear models to test the effect of fertilization and period of experimental duration on diversity-stability relationships at the two spatial scales investigated. Models including an autocorrelation structure with a first-order autoregressive model (AR(1)), where observations are expected to be correlated from one year to the next, gave substantial improvement in model fit when compared with models lacking autocorrelation structure (). Because plant diversity, asynchronous dynamics and temporal stability may be jointly controlled by inter-annual climate variability, we run similar analyses on the residuals of models that included the coefficient of variation among years for each of temperature and precipitation. Results of our analyses controlling for inter-annual climate variability did not differ qualitatively from the results presented in the text (Figures S4).

Second, we used structural equation modelling (SEM) (Grace et al. 2012) with linear models, to evaluate multiple hypothesis related to key predictions from theories (Table S1). SEM model shown in Fig. 1e was evaluated separately for each period of experimental duration and each treatment (Fig. S2) and a meta-analysis (Table 1) was performed to estimate effect sizes for SEM’s paths across time for each treatment (Fig. 3). We run similar models based on nutrient-induced changes in diversity, stability and asynchrony. Changes in diversity, stability and asynchrony at the two scales considered were calculated as the natural logarithm of the ratio between the variable in the fertilized and unmanipulated plots. For the meta-analysis, we used a first-order autoregressive model to account for autocorrelation over time (). We used the R functions ‘psem’ to fit separate piecewise SEM () and ‘rma.rv’ with time random effect and ‘AR’ structure to perform the meta-analysis ().