Supplement for:
Identifying Mechanisms that Structure Ecological Communities by Snapping Model Parameters to Empirically-Observed Tradeoffs

Authors:
Adam Clark*, Clarence Lehman, and David Tilman
**Correspondence to: adam.tclark@gmail.com*

Contents:

Materials and Methods:

I. Data Collection

As described in the main text, we used data from monocultures in seven experiments: Three competition experiments (E026, E055, E070), three diversity experiments (E120, E123, E249), and one monoculture garden (E111). We also compared regression and mechanism-based model predictions to multi-species mixtures in E120. Full methods for E026 are described in Wedin and Tilman (1993); E055 and E070 (which was a subset of E055) are described in Dybzinski and Tilman (2007); E111 is described in Craine (2002); E123 is described in Tilman *et al.* (1996); E120 is described in Tilman *et al.* (1997); and E249 (which is nested within E120) is described in Whittington *et al.* (2013) and Cowles *et al.* (2016).

Describing each experiment briefly:

E026: Established in 1986. Intended as a test of soil nutrient gradients on competitive hierarchies: we used only data from monocultures grown in nitrogen poor soils. Plots were 0.75m by 0.75m. Vegetation was sampled using 0.1m by 0.4m clip strips. We include data from sampling done in 1988, 1990, and 1992.

E055: Established in 1988. Intended as a test of soil nutrient gradients on competitive hierarchies: we used only data from monocultures grown in nitrogen poor soils. Plots were 1.1m by 1.1m. Vegetation was sampled using 0.1m by 0.5m or 0.1m by 0.1m clip strips. We include data sampling done in 1992, 1996, and 1999. Annual burning treatments to control litter build-up (conducted in the spring before biomass greening) began in 1991.

E070: Same establishment and sampling data as for E055 (though clip strips were always 0.5m long). We use data collected in 1998.

E111: Established in 1992. Included only monocultures grown in un-amended soils. Plots were either 1.5m by 2.4m, or 1.5m by 1.2m. Vegetation was sampled using 0.1m by 2.3m clip strips in the larger plots, and two 0.1m by 1.2m clip strips in smaller plots. We use data collected in 1997.

E120: Established in 1994, but still ongoing. Experiment includes plots planted with 1, 2, 4, 8, and 16 species. We used data from both monocultures and multi-species mixtures in this experiment. Plots are 9m by 9m. Vegetation is sampled using four 0.1m by 1.5m clip strips per plot. We include sampling conducted annually between 1997 and 2014. Plots are burned annually in the spring to prevent litter build-up. Two grasses, *Agropyron smithii* and *Elymus canadensis*, never established in monocultures or multi-species mixtures in E120, and two oak species, *Quercus ellipsoidalis* and *Q. macrocarpa*, were removed or competitively excluded from most plots. For our analyses, we therefore treated all plots as though *A. smithii* and *E. canadensis* had not been planted, and excluded all plots where oaks were present from analyses. One species, *Amorpha canescens*, is present only in the 1 and 16-species plots in this experiment.

E123: Established in 1994. Includes plots planted with various numbers of species, but we use only data from monocultures. Plots were 3m by 3m. Vegetation was sampled using 0.1m by 3m clip strips. We use data collected in 1997.

E249: Built into a subset of the plots in E120. Established in 2008, but still ongoing. Includes experimental warming treatments, but we use data only from un-manipulated monocultures. Plots are divided by treatment into 2.5m by 3m subplots. Vegetation is sampled using two 0.1m by 1.5m clip strips. We use data collected in 2011 and 2012. Plots are burned annually in the spring to prevent litter build-up.

All data are available through the LTER data portal (lternet.edu/sites/cdr), or at the Cedar Creek website (cbs.umn.edu/explore/field-stations/cedarcreek). See Appendix *XII* for directions

on how to access the source code used to implement data cleaning and standardization. To assure comparability of data across these experiments we use only data:

1. *From plants grown on nitrogen-limited soils.* E026, E055, and E077 included soil nitrogen gradients, and E026 included a fertilization treatment. We excluded the fertilized treatment, and all soil nitrogen treatments that were more than one standard deviation above the mean 1994 total soil nitrogen concentration for plots in E120.
2. *With low litter.* E026 was not burned annually, which lead to litter accumulation. We removed all observations where litter was more than one standard deviation above the mean litter abundance observed in E120.
3. *For well-established monocultures.* Some monocultures did not establish well. We excluded any observations with aboveground biomass below 10 g m^{-2}. Similarly, *Poa pratensis* did not establish well in its planted monocultures in E120, but did do well in more diverse plots (likely because of limited water availability in monocultures). Because of this, we used only monoculture measurements from experiments other than E120 for this species.
4. *With complete sampling.* We excluded any years that did not include measurements of a majority of species planted in the experiment, to reduce bias from random variation across years.
5. *For herbaceous perennials.* We included only species that fell into one of four functional groups: C3 and C4 grasses, non-legume forbs (flowers), and legumes. We excluded woody species, because these were usually not allowed to mature over the course of experiments, and sedges, because we did not have sufficient species coverage for them.
6. *For mature, maintained plants.* We only included data collected three or more years after monocultures were planted to avoid establishment effects. Similarly, we excluded any measurements from plots that were taken after weeding treatments had ceased.

II. Measurements of characteristics

For each of the characteristics that we report, we estimated mean values and within-species characteristic variation using linear mixed-effects models implemented through the `lmer` function in the `lme4` package (Bates *et al.* 2015) in the `R` programming language (R. Development Core Team, 2016). To account for repeated measurements and spatial pseudo-replication, we fit each model with a random intercept with respect to subplot (if present), nested within plot, nested within experiment, following the form: characteristic ~ species + (1|experiment/plot/subplot). We used the fixed effect parameter estimates and standard errors from these fitted models for characteristic estimates. To meet distributional assumptions, we log-transformed R^* and B^*_{mono}, and logit-transformed q prior to all analyses.

In order to assure that mean estimates for monocultures from our fitted models matched observations from monocultures in E120, we centred estimates for $B_i^*{}_{mono}$ for all species in E120 to match the mean monoculture biomass observed in experimental plots in E120 (except for *P. pratensis* as outlined above). We did not centre means prior to fitting the tradeoff surface. For the model with snapped characteristics, we adjusted q_i after centring means in order to maintain the original $q_i B_i^*{}_{mono}$ (and therefore coexistence criteria) predicted from the tradeoff surface.

To account for variability in initial soil fertility (S) among plots in E120, we multiplied $B_i^*{}_{mono}$ by C_k/C, where C_k is the pre-treatment (in 1994) total carbon content of soil in plot k, and C is the mean pre-treatment total carbon content across all plots in E120. We use this metric for fertility because soil carbon concentration is more closely correlated with nitrogen mineralization rates at our site than is soil nitrogen concentration (or any other easily measured soil characteristics) (Fornara & Tilman 2008). This ratio was meant to represent S_k/S, which determines changes in biomass due to soil fertility following Eq. (2) in the main text: $B_{i,j}^*{}_{mono} =$

$(1 - m_i/c_i)S_{i,k}/q_i$, which implies $B_{i,j}{}^*_{mono}/B_i{}^*_{mono} = S_{i,k}/S$. Characteristics for all species are available in Table S1.

III. Observed pairwise associations in E120

To demonstrate competitive relationships among species observed in multi-species communities in E120, we fit linear regressions relating the biomass of inferior competitors (based on snapped R^* values) to the biomass of all superior competitors in each plot. For all regressions, we used log-transformed biomass + 0.01 gm^{-2} in order to account for zero observations, and included pre-treatment total soil carbon, year, and their interaction as covariates. Given an inferior competitor species *i* and a superior competitor species *1, 2, 3...*, this yielded a model of the form:

$\log_{10}(B_i) = \beta_0 + \beta_1\log_{10}(\%\text{ soil C}) + \beta_2\text{year} + \beta_3\log_{10}(\%\text{ soil C})*\text{year} + \beta_4\log_{10}(B_1) + \beta_5\log_{10}(B_2) + \beta_6\log_{10}(B_3) + ... + \varepsilon$

where β_i are fitted coefficients, and ε is unexplained residual error.

To estimate the statistical effect of each superior competitor on inferior competitors, we fit a model with all species but the focal superior competitor, and then used backwards selection based on regression AIC to remove terms from the model (e.g. to estimate the effect of species 1 on species 4, we fit a model with % soil C, year, and species 2 and 3, and then ran backwards selection). This produced a "reduced" model of covariates, including % soil C, year, and other potential competitors. Finally, we fit a regression comparing residual variance from this reduced model to residual variance comparing the abundance of the focal superior competitor to the full suite of covariates. This is analogous to building an added variable plot or conducting a type III sum of squares regression, and provided an estimate of the marginal statistical effect of a single

superior competitor on an inferior competitor, after controlling for the effects of potential covariates.

In most cases, superior competitors had either negative, or non-significant statistical effects on inferior competitors (Fig. S1). This included negative associations between legume abundance and the abundance of several grass species (with grasses predicted to be competitively superior to the legumes). The only consistent positive statistical effects were of other species on the abundance of *P. pratensis*, which is probably a symptom of failed establishment in low-diversity mixtures.

IV. Characteristic-based regressions

We fit linear regressions estimating species abundances in the multi-species mixtures using the three characteristics identified in the main text (i.e. R^*, q, and B^*_{mono}). Regressions included each characteristic individually as predictor variables, all three characteristics, or all two- or three-way interactions among these characteristics. As described in the main text, we used log transformations for biomass all all characteristics but q, for which we used a logit transformation. Models also included one or two covariates, which were interacted with all other terms in the model. First, all models included percent soil carbon measured at the start of the E120 experiment (i.e. in 1994), as described in Appendix *II*. We included this term in order to account for its role in mechanistic model predictions. Second, we fit all models with and without planted richness as a covariate. This was to test for changes in the association between characteristics and estimates of species abundances relating to mixture diversity.

All models were fit using the `lm` function in `R`. Regressions were fitted to mean species abundance in each plot (i.e. averaged across all observations between 2001 and 2014). As such,

regressions followed the form: `lm(abundance ~ (characteristics and interactions)*(covariates))`. We also used these regressions to estimate total plot biomass, by summing together predictions of species abundances in each plot. We did this, rather than fit a regression to total observed plot-level biomass, because this allowed us to use species-level characteristic data in a relatively simple regression framework (as opposed to having to include information about multiple species characteristics per plot). To assess model fit, we used the `lmodel2` function (Legendre 2014) to conduct ranged major axis regressions comparing observations to regression predictions. This provided estimates of model fits that were directly comparable to those we generated for the mechanism-based models. To account for the number of fitted parameters in each regression, we calculated adjusted R^2 for the regression models, as $R^2_{adj} = 1 - [(1 - R^2)(n - 1)/(n - k - 1)]$, where n was the number of observations, and k was the number of fitted parameters. However, this had minimal effect, and differences between R^2 and R^2_{adj} were never greater than 0.015.

For each regression, we also calculated the partial R^2 for each predictor variable, describing the relative contribution of each variable to total model fit. To calculate partial R^2, we refit each model after removing each term sequentially, and calculated the change in fit as partial $R^2 = (SSE_{reduced} - SSE_{full})/SSE_{reduced}$, where SSE_{full} and $SSE_{reduced}$ describe the summed square error of the full and reduced models, respectively.

V. Major axis regression

To fit tradeoff surfaces and compare observations and predictions, we used major axis regression (type II regression models), which minimizes the squared Euclidian distance between observed points and the fitted surface (i.e. the point where the tangent connecting the regression

line and the observation meet the regression plane, or formally, the projection of the observed points onto the regression surface). Given an n-dimensional regression fit $\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \beta_0 = 0$, for observed values $x_1, x_2, x_3, \ldots, x_n$, and fitted parameters $\beta_1, \beta_2, \beta_3, \ldots, \beta_n$, the "snapped" coordinates are found by solving $x_{i,snap} = x_i - \beta_i k$, where $k = (\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ...\beta_n x_n + \beta_0)/(\beta_1^2 + \beta_2^2 + \beta_3^2 + \ldots + \beta_n^2)$. We use this method because standard least squares regression assumes that there is no error in the predictor variables, whereas we have no a priori reason for assuming error is absent in any of the variables we consider. We implement this algorithm in the `nondirfit` function, available in the `get_filtered_estimate_functions.R` script (see Appendix *XII.* for instructions on accessing source code).

We fit the regression to the mean observed characteristic values for each species. To test whether the relationships among characteristics matched those predicted by our model (i.e. decreases in R^* must correspond to decreases in q or B^*_{mono}), we estimated confidence intervals and covariance among regression parameters using nonparametric bootstrapping. For each of 20,000 iterations, we resampled species characteristics, with replacement, from the pool of all species, and re-calculated the model parameters for each iteration. We then tested the strength and direction of relationships among characteristics to determine their significance. This also allowed us to test the significance of a three-way negative relationship among characteristics, as predicted by our resource competition model.

VI. Model equilibria and stability

The model we present in the main text can be written in terms that match the classical Lotka-Volterra competition equations, and a general proof of global stability for our system in this form is available in Theorem 3.3.1 of Takeuchi (1996), (p. 36). A briefer proof is available

in Takeuchi *et al.* (1978) in Theorem 4, though this 1978 journal article may prove easier to access than the 1996 book. Here, we *(i)* include a somewhat simpler and relatively robust proof of global stability for this model using linearized stability analysis (i.e. calculation of the dominant eigenvalue at equilibrium), and *(ii)* show how the same solution can be derived using Chesson's pairwise niche overlap and relative fitness framework (Chesson 1990, 2000, 2013; Letten *et al.* 2017).

VI.i. Linearized stability analysis

First, we begin with a classical Lotka-Volterra competition model of the form:

$$dN_i/dt = r_iN_i(1 - (N_i + \alpha_{ji}N_j)/K_i) \quad \text{(S}VI.i.1\text{)}$$

Starting with Eq. (1) from the main text, we can define biomass dynamics with species i competitively inferior to species j as:

$$dB_i/dt = c_iB_i(1 - (q_iB_i + q_jB_j)/S) - m_iB_i \quad \text{(S}VI.i.2\text{)}$$

Redistributing terms yields:

$$dB_i/dt = c_iB_i(1 - (q_iB_i + q_jB_j)/S) - m_iB_i \quad \text{(S}VI.i.3\text{)}$$

$$= (c_i - m_i)B_i - c_i(q_iB_i + q_jB_j)/S)(c_i - m_i)/(c_i - m_i)$$

$$= (c_i - m_i)B_i(1 - c_i(q_iB_i + q_jB_j)/((c_i - m_i)S))$$

$$= (c_i - m_i)B_i(1 - c_i(q_iB_i + q_jB_j)/((c_i - m_i)S))$$

$$= (c_i - m_i)B_i(1 - (q_iB_i + (q_i/q_i)q_jB_j)/((1 - m_{i.}/c_i)S))$$

$$= (c_i - m_i)B_i(1 - (B_i + (q_j/q_i)B_j)/((1 - m_{i.}/c_i)S/q_i))$$

Substituting $N_i \equiv B_i$, and $N_j \equiv B_j$, and $K_i \equiv (1 - m_i/c_i)S/q_i \equiv B_i{}^*_{mono}$ in (S6) yields:

$$dN_i/dt = (c_i - m_i)N_i(1 - (N_i + (q_j/q_i)N_j)/(K_i) \quad \text{(S}VI.i.4\text{)}$$

From equation (S*VI.i*.1), this implies that $r_i = (c_i - m_i)$, and $\alpha_{ij} = q_j/q_i$. Note that competition in the reverse direction (i.e. the effect of inferior competitors on superior competitors) is simply $\alpha_{ji} = 0$. Because of this, equilibrium conditions can be derived sequentially, starting with the best competitor (for which $B_{i'}^* = K_{i'}$), and proceeding through sequentially worse competitors based on α_{ij} and B_j^*.

Given $\alpha_{ij} < 0$ and $\alpha_{ji} = 0$, Eq. (S*VI.i*.1) shows that $\partial(dN_i/dt)/\partial N_j < 0$ for all populations where $N_i > 0$ and $N_j > 0$. To evaluate $\partial(dN_i/dt)/\partial N_i$ at equilibrium, we define the effect of all superior competitors on species i as $\Sigma_{j<i}\,(q_j/q_i\,N_j^*) \equiv \kappa_i$. From Eq. (3) in the main text, this implies that for the non-trivial equilibrium, $N_i^* = K_i - \kappa_i$. Substituting this into (S*VI.i*.4) at equilibrium yields:

$$dN_i/dt|[Ni = N_i^*] = r_i(K_i - \kappa_i)(1 - ((K_i - \kappa_i) + \kappa_i)/(K_i) = 0 \qquad (S\mathit{VI.i}.5)$$

Given a small change in population size of species 1, η_i, this can be re-written as:

$$dN_i/dt|[Ni = N_i^* + \eta_i] = r_i(K_i + \eta_i - \kappa_i)(1 - ((K_i + \eta_i - \kappa_i) + \kappa_i)/(K_i) \qquad (S\mathit{VI.i}.6)$$

Calculating the partial derivative of (S*VI.i*.6) with respect to η_i yields:

$$\partial(dN_i/dt)/\partial\eta_i = -r_i((K_i + \eta_i)/K_i - 1) - r_i(K_i - \kappa_i + \eta_i)/K_i \qquad (S\mathit{VI.i}.7)$$

Given $\eta_i > 0$, the first term is always negative, because $(K_i + \eta_i)/K_i > 1$. The second term is always negative for any $K_i < \kappa_i$ (i.e. $N_i^* > 0$). Because η_i describes changes in N_i around equilibrium, this implies that $\partial(dN_i/dt)/\partial N_i\,|[N_i = N_i^*] < 0$.

Following from this, the Jacobian matrix describing interspecific interactions for any number of competing species in this model at equilibrium is a triangular matrix with purely negative non-zero elements. Because the diagonal elements of a triangular matrix are also its eigenvalues, this implies that all eigenvalues for the system are also negative. Thus, the equilibrium is locally stable, given any mixture of species with positive abundances at

equilibrium. As described in Eq. (3) and Eq. (5b) in the main text, this requires $K_i > \Sigma_{j<i} (q_j/q_i)K_j$, for all species with $R_j^* < R_i^*$ (i.e. any feasible equilibrium is also stable).

This locally stable equilibrium can also be shown to be globally stable. Because dynamics of all superior competitors are independent from dynamics of inferior competitors, we can address dynamics of inferior competitor *i* given equilibrium abundances of all superior competitors in the community as:

$$dN_i/dt = r_iN_i(1 - (N_i + \kappa_i)/(K_i) \quad \text{(S}VI.i.8\text{)}$$

This system has only two possible equilibria: $N_i^* = K_i - \kappa_i$, and the trivial equilibrium $N_i^* = 0$. Following the procedure in (S*VI.i.*5)-(S*VI.i.*7) to solve for $\partial(dN_i/dt)/\partial N_i$ at $N_i = 0$ yields:

$$\partial(dN_i/dt)/\partial\eta_i = -r_i(\eta_i/K_i - 1) + r_i(\kappa_i + \eta_i)/K_i \quad \text{(S}VI.i.9\text{)}$$

Given $K_i > \kappa_i$, this is strictly positive for small deviations (i.e. $\eta_i < K_i$ and $\eta_i < \kappa_i$), implying that the diagonal elements of the (triangular) Jacobian matrix are positive, and that its eigenvalues are therefore also positive and that the trivial equilibrium is unstable. The equilibrium $N_i^* = K_i - \kappa_i$ is therefore the only stable equilibrium for species *i*. Finally, from (S*VI.i.*8), $dN_i/dt > 0$ for all $0 < N_i < K_i - \kappa_i$, and $0 > dN_i/dt$ for all $K_i - \kappa_i < N_i$. This implies that the system will approach the stable positive equilibrium from any feasible starting points, and is therefore globally stable at $N_i^* = K_i - \kappa_i$.

VI.ii. Chesson's pairwise method

Given a system with two competing species of the form:

$$dN_1/dt = r_1N_1(1 - a_{11}N_1 - a_{12}N_2) \quad \text{(S}VI.ii.1\text{a)}$$

$$dN_2/dt = r_2N_2(1 - a_{22}N_2 - a_{21}N_1) \quad \text{(S}VI.ii.1\text{b)}$$

Chesson defines niche overlap as:

$$\rho = [(a_{12}a_{21})/(a_{11}a_{22})]^{1/2} \quad \text{(S}VI.ii.2\text{a)}$$

and absolute fitness of species 2 relative to species 1 as:

$$f_2/f_1 = [(a_{11}a_{12})/(a_{22}a_{21})]^{1/2} \quad \text{(S}VI.ii.2\text{b)}$$

Derivations and discussions of these terms are available in several works by Chesson (1990, 2000, 2013), and in Letten *et al.* (2017). In this framework, stable coexistence is achieved when niche overlap is less than relative fitness, following:

$$\rho < f_2/f_1 < 1/\rho \quad \text{(S}VI.ii.3\text{)}$$

Following the re-parameterization of the model outlined in Eq. (*VI.i.*1), we can write the interaction terms in Eqs. (S*VI.ii.*1a-b) as:

$$a_{ii} = 1/K_i \quad \text{(S}VI.ii.4\text{a)}$$

$$a_{ij} = (q_j/q_i)(1/K_i) \quad \text{if } R_j^* < R_i^* \quad \text{(S}VI.ii.4\text{b)}$$

$$= 0 \quad \text{otherwise}$$

Because a_{ij} is always zero for any pair of species where $R_i^* < R_j^*$, pairwise interactions between any two species with $R_1^* < R_2^*$ results in $a_{21} > 0$ and $a_{12} = 0$ (i.e. species 2 is the inferior competitor). Substituting these into Eq. (S*VI.ii.*3) and taking the limit as a_{12} approaches zero, we find that the species can stably coexist provided that:

$$[(a_{21})/(a_{11}a_{22})]^{1/2} < [(a_{11})/(a_{22}a_{21})]^{1/2} \quad \text{(S}VI.ii.5\text{)}$$

Note that the upper limit described in the third term of the inequality can now be omitted, because $1/\rho$ necessarily tends towards infinity as a_{12} approaches zero, while the other two terms tend towards zero. Substituting Eqs. (S*VI.ii.*4a-b) into Eqs. (S*VI.ii.*5), we can rewrite the inequality as:

$$[\ (q_1/q_2)K_1]^{1/2} < [(K_2)^2/((q_1/q_2)K_1]^{1/2} \quad \text{(S}VI.ii.6\text{)}$$

From this, we can also rewrite the terms for niche overlap and relative fitness in Eqs. (*VI.ii.*2a-b) as:

$$\rho = [(q_1/q_2)K_1]^{1/2} \quad \text{(S}VI.ii.7\text{a)}$$

$$f_2/f_1 = [(K_2)^2/((q_1/q_2)K_1)]^{1/2} \quad \text{(S}VI.ii.7\text{b)}$$

Note that by multiplying both side of Eq. (*VI.ii*.6) by $[(q_1/q_2)K_1]^{1/2}$, we can further simplify the inequality to:

$$(q_1/q_2)K_1 < K_2 \quad \text{(S}VI.ii.8\text{)}$$

Lastly, recall that this model has a strict competitive hierarchy, in which species with higher values of R^* have no effect on the growth of species with lower R^*. As such, we can calculate equilibrium population abundances sequentially, starting with the best competitor, in order to determine the abundance of each species at equilibrium. This also means that we can abstract the pairwise solution in Eq. (S*VI.ii*.8) to systems with many competing species, by substituting species 1 with the summed effects of all superior competitors on species 2. This yields a general criterion for stable coexistence:

$$\Sigma_{j<i}\,(q_j/q_i)K_j < K_i \quad \text{(S}VI.ii.9\text{)}$$

for all species *j* that are superior competitors to species *i* (i.e. all species with $R_j^* < R_i^*$). Note that this inequality is identical to Eq. 5b in the main text, and to the solution derived in *Appendix IX.i*.

VII. Making model predictions

To predict species-level biomass for the polyculture mixtures planted in E120, we parameterized Eq. (3) in the main text using raw species characteristics or snapped characteristic estimates, and solved it sequentially from the best competitor to the worst competitor. This provided an analytically tractable prediction of mean species-level biomass at equilibrium.

Parameterizing the model with mean characteristic values (i.e. no intraspecific characteristic variation) under-predicted the number of coexisting species across all multi-species communities (Fig. S4A). To account for within-species characteristic variation, we calculated observed sample variance among monoculture replicates of each species, and included this in the model either as variability around raw characteristic values, or around snapped points on the tradeoff surface. This substantially improved predictions of coexistence for both types of models (Fig. S4B). We therefore used this method to account for intraspecific characteristic variability for all subsequent predictions.

To account for within-species variability in species characteristics, we sampled 20,000 separate estimates of species characteristics for each planted mixture, based on the standard error in characteristics as calculated and reported in Appendix *II*, and either the mean raw characteristics for each species, or the snapped characteristics from the tradeoff surface. We then used these 20,000 characteristic estimates to generate predictions of species-level aboveground biomass, plot-level biomass, and community stability, and calculated mean and standard deviation from the resulting distributions.

We compared our predictions to species-level aboveground biomass observations in E120 measured between 2001 and 2014. To calculate mean biomass across repeated observations or predictions, we log-transformed the values, and used a hurdle model to account for instances of zero biomass. Thus, the mean over "*n*" potential biomass values was calculated as $B_i = z_i \exp(\Sigma_{Bi^*>0} \log(B_i^*)/(n\ z_i))$, where z_i is the proportion of observations or predictions where $B_i > 0$. Variability was calculated solely from non-zero, log-transformed biomass values.

VIII. Assessing model fit

We used the two-sample Wilcoxon test to determine significance of differences between observed and estimated species richness in Fig. 3A-B, differences in error between models parameterized with raw and snapped characteristics in Fig. 3C-D, and differences in original and augmented model fit in Fig. 4. For all tests, we compared quantities paired by plot (i.e. observed vs. estimated richness in each plot, error from raw vs. snapped characteristics in each plot, etc.).

In most cases, we calculated R^2 from the difference between observed and predicted values as:

$$1 - \Sigma_i(observed_i - predicted_i)^2/\Sigma_i(observed_i - \text{mean}[observed])^2 \quad \text{(S1)}$$

The exceptions to this are the results in Table 1, Fig. 3A-D, and in Fig. S5, where we show R^2 for the fitted regression line comparing observed and predicted values (i.e. total fraction of variation in observed values explained by predicted values).

For regressions between pairs of variables (Fig. 3E-H; Fig 4), we used the `lmodel2` (Legendre 2014) function in `R` to fit ranged major axis regression, and to calculate p-values. Because this function only accepts bivariate data, we implemented our own function in `R` (`nondirfit`) to handle higher multivariate regressions (e.g. the tradeoff surface among three characteristics), with details described in Appendix *V*. This function uses a nonlinear optimizer to minimize the distance between standardized raw and snapped characteristic values (i.e. (x_i-mean(x))/sd(x)), and perfectly matches outputs from the `lmodel2` function for bivariate ranged major axis regression.

To estimate significance of estimates from the major axis regressions fit using the `nondirfit` function, we used a simple bootstrapping routine. Given *n* observations, we sampled with replacement from the total pool of observations *n* times, and fit the regression

using the sampled dataset. We then repeated this procedure for 20,000 iterations to generate a distribution of estimates for all terms in the regression. For the tradeoff surface, we used the multivariate distribution of all slope parameters to calculate the proportion of iterations for which there was a negative relationship among all three characteristics included in the tradeoff. See Table S3 for fitted values, intervals, and p-values.

IX. Simulating communities from the tradeoff surface

We sampled simulated communities from across the fitted tradeoff surface using the covariance relationship among $\log_{10}$-transformed R^* and B^*_{mono}, and logit-transformed q, using the `mvtnorm` package in `R` (Genz & Bretz 2009; Genz *et al.* 2015). For each of 20,000 iterations, we sampled species from the tradeoff surface, and randomly assembled them into mixtures of species matching the planted richness in plots in E120 to predict total community biomass, and changes in species-level biomass relative to monocultures as a function of other competitors in the community.

To characterize communities in E120, we constrained sampling from this surface, because the species used to assemble E120 were not sampled randomly from across characteristic space (and because there were not enough species in E120 to directly parameterize the multivariate normal distribution only from species in E120). Rather, the experiment included four species from each of the following functional groups: C3 grass, C4 grass, non-legume forbs, and legumes. These differ substantially in many characteristics.

We constrained our sampling routine in two ways. First, for each iteration, we constrained samples such that the total plant community matched the functional group distribution found in E120 – 2 C3 grasses (4 species, minus *A. smithii* and *E. canadensis*, which

did not germinate), 4 C4 grasses, 4 non-legume forbs, and 3 or 4 legumes (4 species, minus *A. canadensis* in 2, 4, and 8 species mixtures, where it was not planted). To do this, we only included random samples from the tradeoff surface where characteristics fell within the observed ranges of these functional groups based on the 35 species used to fit the surface. Second, within each iteration, we randomly assembled the simulated species into multi-species communities of varying richness, where probability of drawing a species from a particular functional group matched the average planted proportion of that functional group in E120. Thus, while our sampling routine ensured that simulated species characteristics would be similar to those of the functional groups planted in E120, it did not require any specific knowledge of the planted species or functional group composition of any individual plot.

X. Constraints from the empirical tradeoff surface

Because the empirical tradeoff that we discuss in the main text imposes strict relationships among R^*, q, and B^*_{mono}, it is possible that many of our findings, such as increased coexistence and transgressive over-yielding, are trivial outcomes resulting from the negative correlation among characteristics. However, these results turn out not to be trivial because they rely not only on the *shape* of the tradeoff surface, but also the *distribution of characteristics* across the surface. In this section, we demonstrate that neither coexistence nor transgressive over-yielding are assured for species with characteristics that fall along the empirical tradeoff surface (or any tradeoff surface that imposes simple linear constraints on transformed parameter values, as outlined in Appendix *V*).

X.i. Coexistence on the tradeoff surface

As noted in the text and in Appendix *VI*, coexistence between two species in our resource competition model with $R_i^* < R_j^*$ requires that $q_i B_i^*_{mono} < q_j B_j^*_{mono}$. This implies that $dqB^*_{mono}/d(-R^*) < 0$. Given the fitted tradeoff surface $\beta_1 \log_{10}(B^*_{mono}) + \beta_2 \text{logit}(q) + \beta_3(-\log_{10}(R^*)) + \beta_0 = 0$, we can locally approximate this coexistence criterion as:

$$d[\text{logit}(q)\ \log_{10}(B^*_{mono})]/d[-\log_{10}(R^*)] = -\beta_3(\beta_1 \log_{10}(B^*_{mono}) + \beta_2 \text{logit}(q))/(\beta_1 \beta_2) \quad (S\mathit{X}.\mathit{i}.1)$$

because $\log_{10}(R^*)$ scales monotonically with R^*, and $\text{logit}(q)\ \log_{10}(B^*_{mono})$ scales monotonically with qB^*_{mono}. To assure coexistence between all potential pairs of species on the tradeoff surface, this therefore requires:

$$\log_{10}(B^*_{mono}) > -\beta_2/\beta_1 * \text{logit}(q) \quad (S\mathit{X}.\mathit{i}.2)$$

However, the fitted surface need not satisfy the requirement in equation (S*X.i*.2), as $\log_{10}(B^*_{mono})$ and $\text{logit}(q)$ vary in an inverse, but otherwise unconstrained, relationship for any given value of R^*. This implies that any of the species that we observed in our system could potentially be competitively excluded by some other species, even if both fall somewhere along our fitted tradeoff surface. This is exemplified in Fig. 1E in the main text, where many species can be competitively excluded by other species even after snapping characteristics to the empirically observed tradeoff surface.

X.ii. Over-yielding on the tradeoff surface

In our model (which reduces to Lotka-Volterra competition), over-yielding (i.e. greater mean productivity in multi-species mixtures than in the average monoculture) occurs for any species that stably coexist (Loreau 2004). Transgressive over-yielding (i.e. greater mean productivity in multi-species mixtures than in any constituent monoculture) occurs among coexisting species when inferior competitors have higher tissue nitrogen concentration than do

superior competitors (i.e. $q_i > q_j$) (Loreau 2004). Under these circumstances, superior competitors produce more biomass per unit nitrogen uptake than inferior competitors do, but inferior competitors nevertheless increases total community biomass when present because they are able to access nitrogen that superior competitors cannot. We can demonstrate this mathematically following Eq. (3) in the main text, which shows that the equilibrium biomass of species i in mixture, B_i^*, is defined as:

$$B_i^* = B_i^*_{mono} - \Sigma_{j<i} (q_j/q_i) B_j^* \quad (S\mathit{X}.\mathit{ii}.1)$$

where species are arranged in terms of competitive ability, such that $j < i$ implies that j is a superior competitor. Note that when $q_i > q_j$, this implies that the negative effect of species j on the biomass of species i is less than the total biomass of species j, meaning that species i and j growing together in mixture will produce more biomass than either growing alone.

As with coexistence, transgressive over-yielding is not imposed by the shape of the empirical tradeoff or by the assumptions in our model, and depends on both the specific shape of the tradeoff surface, and the distribution of characteristics across it. Nevertheless, transgressive over-yielding (i.e. $q_j > q_i$) is realized for almost all communities that we tested (see Fig. S2B,E). This result also accords with experimental evidence from Cedar Creek that shows productivity is maximized in mixtures that include both species that are efficient nitrogen utilizers (e.g. C4 grasses) and effective at accessing multiple pools of nitrogen (e.g forbs and legumes) (Tilman *et al.* 1997).

XI. Augmenting model with additional factors

In the main text, we discuss three specific factors omitted from in our model which lead to significant improvements in model fit when included – changes in competitive hierarchy with

soil fertility, legume competition for resources other than nitrogen, and seasonal changes in species characteristics. Here, we discuss specific details for how we included these additional strategies. Results from these analyses are presented in Fig. S5.

To test for the effects of changes in competitive hierarchy with soil fertility, we utilized data from monocultures of the two species with the lowest R^* in E120 – *Andropogon gerardi* and *Schizachyrium scoparium* – grown along a soil fertility gradient in E055 (Dybzinski & Tilman 2007). Though mean characteristic estimates suggest that *S. scoparium* is a superior nitrogen competitor, results from E055 suggest that *A. gerardi* is a superior competitor in soils with greater than 0.1% total soil nitrogen concentration (Fig. S5A). While this is well above the initial soil nitrogen concentrations in E120, soil fertility has increased substantially over time because of feedbacks between plants and soils, particularly in diverse plots (Fornara & Tilman 2008). To account for this, we switched competitive hierarchy between *A. gerardi* and *S. scoparium* in 80% of simulations for planted richness treatments of 8 or more species. To better address these changes in competitive hierarchy in the model, we would need to know how soil fertility gradients influenced competitive hierarchies for all species in the multi-species mixtures, and would need methods for predicting total soil nitrogen dynamics through time.

To test for the effects of including other forms of resource competition among legumes, we compared legume species abundances as a function of the abundance of other legume species in E120. Based on these comparisons, it appeared that the legume *Lupinus perennis* most strongly suppressed the abundance of other legumes (Fig. S5D). Though *L. perennis* has the highest R^* of all species in E120, this is not necessarily surprising, as legumes can fix their own nitrogen, and soil nitrate concentration in monoculture may therefore be an indication of fixation rate, rather than competitive hierarchy. To account for this, we switched the R^*-based competitive hierarchy of *L. perennis* with that for *Amorpha canescens* (the legume species with

the lowest R^*) in 80% of simulations. To better address legume competition in the model, we would need more data on legume competition for other potential limiting resources, such as water, phosphorus, or light (Ritchie & Tilman 1995).

Finally, to test for the effects of seasonal changes in species characteristics, we augmented our model to allow *Poa pratensis* to access one third of its total nitrogen before it encountered competitive interactions with other species. This was meant to account for seasonal differences and rooting depth differences with may allow *P. pratensis* to avoid competitive interactions with many species that have lower R^* (McKane *et al.* 1990; Fargione & Tilman 2005). For example, *P. pratensis* grows most of its biomass early in the season, before more dominant competitors, such as the forb *Liatris aspera*, reach their peak biomass (Fig. S5G). To better address seasonality in the model, we would need to account for covariance in characteristics among species across space and time (i.e. to determine how competition among species changes by season and by soil depth), but this would require replicated data from monocultures growing in the same locations during the same years, which we lack for this study.

XII. Source code for replicating study

The full source code for replicating the analyses in this manuscript, including all data needed for fitting and testing models, is available in the file `tradeoff_model.zip` in the supplement, and at https://github.com/adamtclark/tradeoff_model. The file `main.R` is an automated script that will load and run all necessary functions, and output all tables and figures presented here.

References:

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.*, 67.

Chesson, P. (1990). MacArthur's consumer-resource model. *Theor. Popul. Biol.*, 37, 26–38.

Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annu. Rev. Ecol. Syst.*, 31, 343–366.

Chesson, P. (2013). Species Competition and Predation. In: *Ecological Systems* (ed. Leemans, R.). Springer New York, New York, NY, pp. 223–256.

Cowles, J.M., Wragg, P.D., Wright, A.J., Powers, J.S. & Tilman, D. (2016). Shifting grassland plant community structure drives positive interactive effects of warming and diversity on aboveground net primary productivity. *Glob. Change Biol.*, 22, 741–749.

Craine, J.M., Tilman, D., Wedin, D., Reich, P., Tjoelker, M. & Knops, J. (2002). Functional traits, productivity and effects on nitrogen cycling of 33 grassland species. *Funct. Ecol.*, 16.

Dybzinski, R. & Tilman, D. (2007). Resource use patterns predict long-term outcomes of plant competition for nutrients and light. *Am. Nat.*, 170.

Fargione, J. & Tilman, D. (2005). Niche differences in phenology and rooting depth promote coexistence with a dominant C-4 bunchgrass. *Oecologia*, 143, 598–606.

Fornara, D.A. & Tilman, D. (2008). Plant functional composition influences rates of soil carbon and nitrogen accumulation. *J. Ecol.*, 96, 314–322.

Genz, A. & Bretz, F. (2009). *Computation of multivariate normal and t probabilities*. Lecture notes on statistics. Springer, Dordrecht ; New York.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., *et al.* (2015). *mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-3.*

Legendre, P. (2014). *lmodel2: Model II Regression*.

Letten, A.D., Ke, P.-J. & Fukami, T. (2017). Linking modern coexistence theory and contemporary niche theory. *Ecol. Monogr.*, 87, 161–177.

Loreau, M. (2004). Does functional redundancy exist? *Oikos*, 104, 606–611.

McKane, R.B., Grigal, D.F. & Russelle, M.P. (1990). Spatiotemporal differences in N-15 uptake and the organization of an old-field plant community. *Ecology*, 71, 1126–1132.

R. Development Core Team,. (2016). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing.

Ritchie, M. & Tilman, D. (1995). Responses of legumes to herbivores and nutrients during succession on a nitrogen-poor soil. *Ecology*, 76, 2648–2655.

Takeuchi, Y. (1996). Global Stability. In: *Global dynamical properties of Lotka-Volterra systems*. pp. 21–56.

Takeuchi, Y., Adachi, N. & Tokumaru, H. (1978). Global stability of ecosystems of the generalized volterra type. *Math. Biosci.*, 42, 119–136.

Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M. & Siemann, E. (1997). The influence of functional diversity and composition on ecosystem processes. *Science*, 277, 1300–1302.

Tilman, D., Wedin, D. & Knops, J. (1996). Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature*, 379, 718–720.

Wedin, D. & Tilman, D. (1993). Competition Among Grasses Along a Nitrogen Gradient: Initial Conditions and Mechanisms of Competition. *Ecol. Monogr.*, 63, 199.

Whittington, H.R., Tilman, D. & Powers, J.S. (2013). Consequences of elevated temperatures on legume biomass and nitrogen cycling in a field warming and biodiversity experiment in a North American prairie. *Funct. Plant Biol.*, 40, 1147.

Supplementary Tables and Figures:

See repository `sup_tables.zip` in the supplement for .csv versions of each of the tables described below.

Table S1. Raw and snapped characteristics for species at Cedar Creek. Standard errors show within-species variation from fitted linear mixed effect models described in Appendix *II* (i.e. the standard error for the fixed effects in the model). For species included in E120, abbreviations correspond to labels in the main text.

Table S2. Number of plot replicates for monoculture species and E120 mixtures.

Table S3. Parameters for the fitted tradeoff surface. Columns show mean and 95% confidence interval for intercept and slope terms, following parameterization for a plane $0 = \beta_0 + \beta_1 \log_{10}(B^*_{mono}) + \beta_2(-\log_{10}(R^*)) + \beta_3 \,\mathrm{logit}(q)$. Negative value for R^* term is included because R^* is inversely related to competitive ability. "3-way" shows p-value for bootstrapped iterations where all three characteristics are negatively related to one another (i.e. increasing any one term must be compensated for by decreasing the other two, or, said a different way, β_1, β_2, and β_3 all share the same sign).

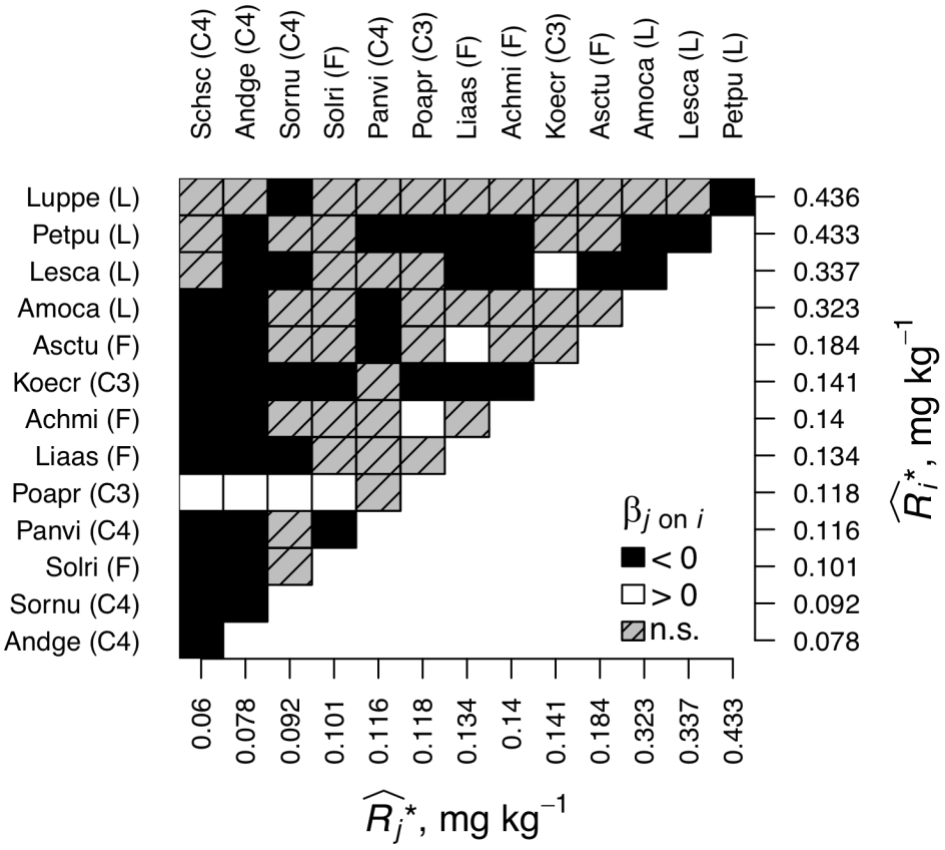

Fig. S1. Pairwise relationships between species observed in experimental multi-species communities in E120. $B_{j \text{ on } i}$ describes the statistical effect of superior competitor species *j* (columns) on inferior competitor species *i* (rows), with competitive hierarchy corresponding to $\hat{R}^*$. Species name abbreviations are described in Table S1. Letters in parentheses correspond to functional groups: C3/C4 grass, (F)orb, or (L)egume.

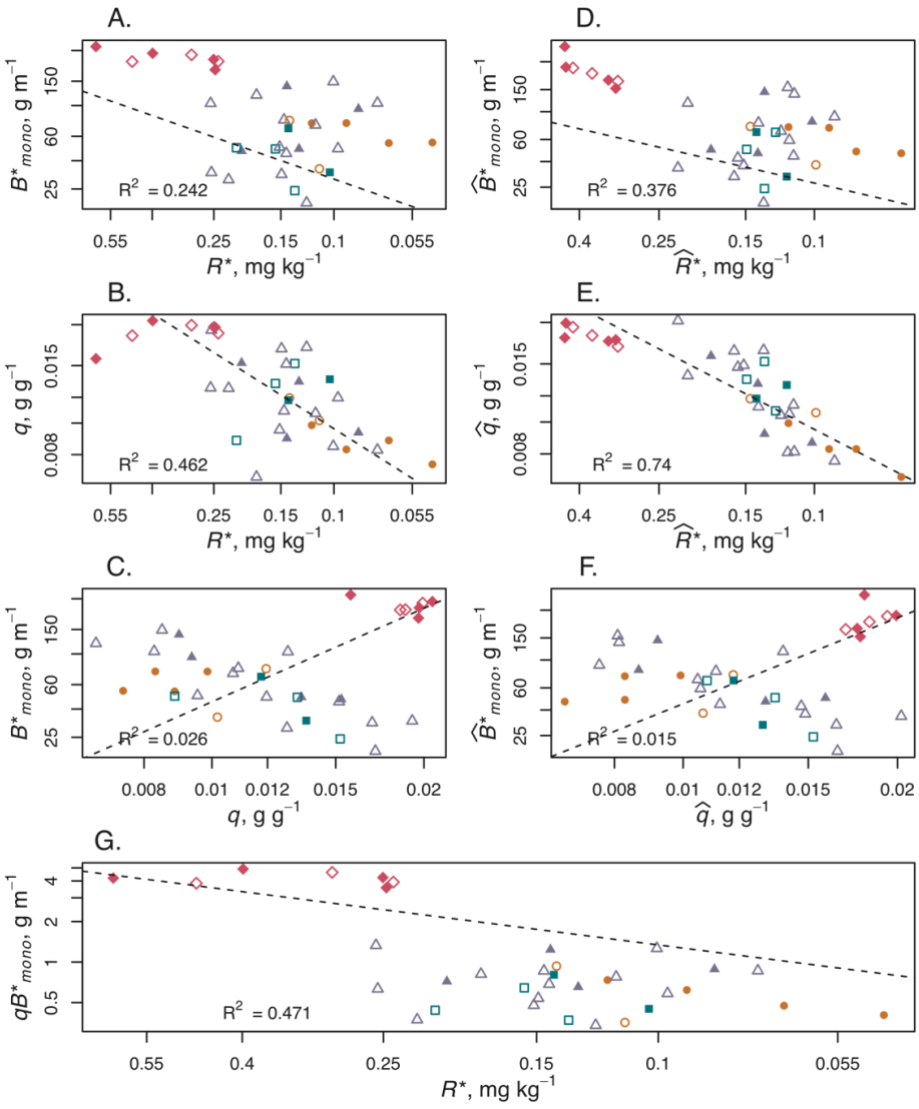

Fig. S2. Bivariate fits between all pairs of characteristics used in the analyses. **(A-C)** show pairwise comparisons for raw characteristic values, and **(D-F)** show snapped characteristics. **(G)** shows bivariate relationship between the two major groupings of plant characteristics that we use in our model, qB^*_{mono} and R^*. Dashed line and R^2 values correspond to ranged major axis regression. Note that simple pairwise relationships explain much less of the variation than the three-dimensional fitted plane, even for snapped characteristics. Symbols and colours show functional groups and identify species planted in the multi-species plots, as described in the legend for Fig. 1 in the main text.

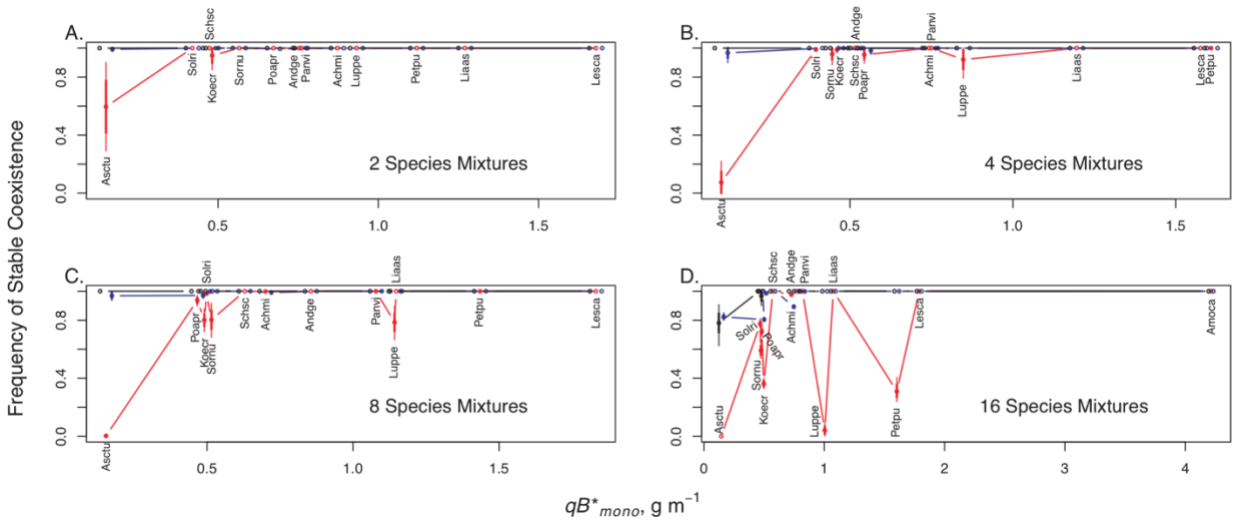

Fig. S3. Coexistence in observed communities and model predictions, as a function of species mean monoculture aboveground tissue nitrogen. Intervals show mean ± one standard deviation, and 95% confidence interval. Species abbreviations are described in Table S1. Black lines show fraction of multi-species plots in which planted species have persisted. Red and blue lines show results for mechanism-based models parameterized with raw and snapped characteristics, respectively. Stable coexistence in these models is identified using linearized stability analysis, as described in the methods section of the main text. Note that most species appear to stably coexist in the observed plots, as predicted by the model parameterized with snapped characteristics. For the model parameterized with raw characteristics, species that have both low qB^*_{mono} and low R^* (e.g. *Achilliea millefolium*) tend to be competitively excluded.

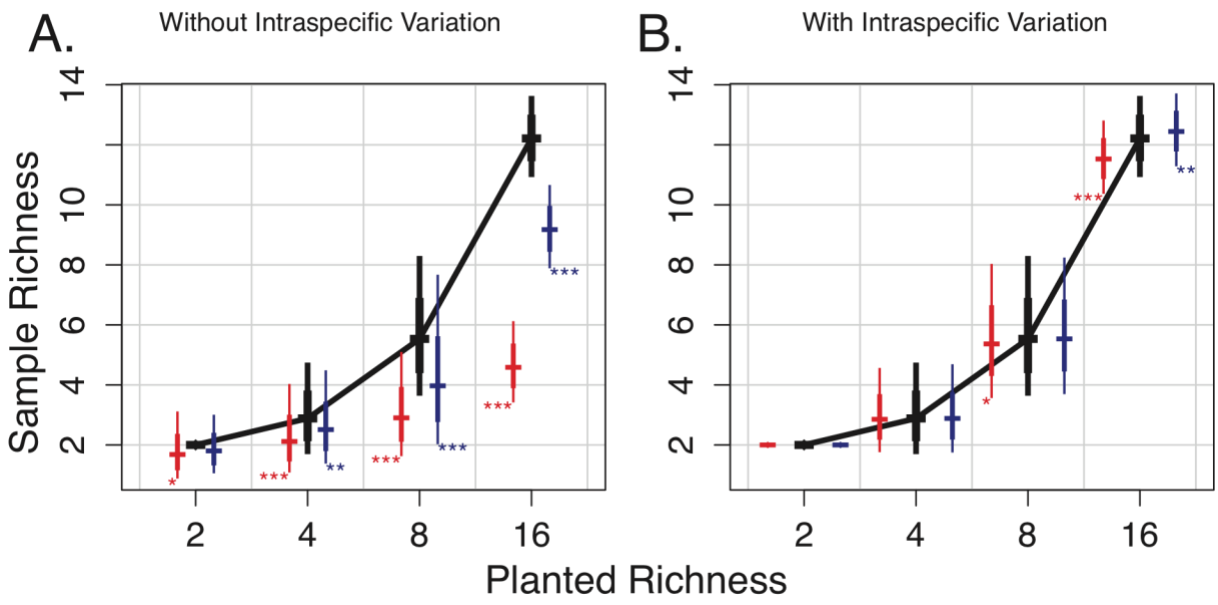

Fig. S4. Observed and predicted species richness across diversity treatments for models with and without intraspecific characteristic variation. Intervals show 95% confidence interval and mean ± standard deviation. Black, red, and blue intervals denotes observed richness, and fits for raw and snapped characteristics, respectively. Asterisks signify significant differences between observations and predictions (*p<0.05; **p<0.01; ***p<0.001); p-values are from two-sample Wilcoxon tests.

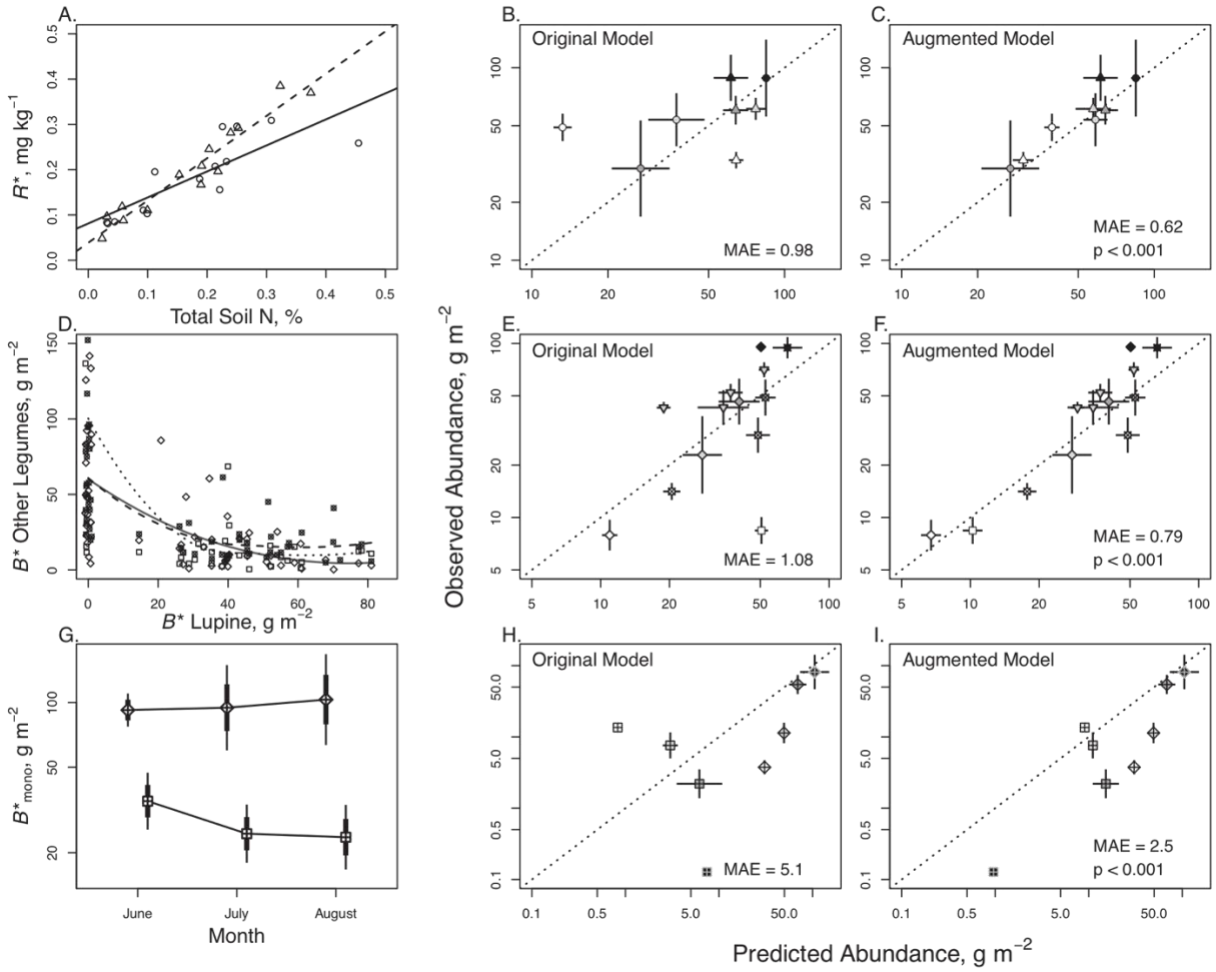

Fig. S5. Examples of model augmentations. Accounting for these factors significantly improves model fit. Intervals show mean ± standard error for observed and predicted biomass across all plots in each planted richness treatment. P-values show significant differences in MAE between original and augmented models from two-sample Wilcoxon tests. **(A-C)** *A. gerardi* (circles/solid line) and *S. scoparium* (triangles/dashed line) switch competitive hierarchy in rich soils. **(D-F)** Though it has a higher *R**, *L. perennis* (triangles) appears to suppress the legumes *A. canescens* (boxes/dotted line), *Lespedeza capitata* (crossed-out circles/dashed line), and *Petalostemum purpureum* (diamonds/solid line) (values of zero biomass for *L. perennis* are jittered for clarity). **(G-I)** Early-season species such as the cool-season grass *P. pratensis* (boxed triangles) grow the

majority of their biomass early in the year, while more dominant nitrogen competitors such as the forb *Liatris aspera* (gridded boxes) grow much of their biomass late in the year. Allowing *P. pratensis* to access resources before other species in one third of diverse plots leads to increases in its abundance, and decreases in warm season species abundances. Lines in **(G)** show mean ± standard error and 95% confidence interval.