**Appendix S1. Supplemental Methods**

*Site preparation and experiments design*

Starting from an initial tallgrass prairie species list from Prairie Moon Nursery (Winona, MN), we filtered to prairie species that are (1) native to the broader Chicago region (Wilhelm and Rericha 2017), (2) not exceptionally aggressive (based on our own observations and consultation with colleagues), and (3) not limited to the driest or wettest prairies (based on reported categorical habitat moisture; Lichvar et al. 2014) Given local seed and plant availability, our final design resulted in 127 species.

Each mix was installed in the experiment in two replicates, for a total of 72 treatment or mixture plots. In addition, each species was installed in two replicates as a monoculture, for a total of 254 monoculture plots. The experiment was initiated in summer of 2015 in a field dominated primarily by *Bromus inermis*, *Dipsacus* sp*.*, and *Pastinaca sativa*, along with other common Eurasian agricultural weeds (see full list of common weeds in Hipp et al. 2018, supplement). The field was treated with 5% glyphosphate herbicide and disced, then planted with a cover crop and disced once again before planting in August 2016. In mixture plots installed by plugs, all 15 species were assigned random positions within each 1m2 quarter of the 4m2 plot, so that each quarter of the plot contained a full set of 15 species. Plots were laid out in a fully blocked design with three blocks in each of two super blocks. Blocks and superblocks were delineated based on pre-planting measurements of soil A-horizon depth (Hipp et al. 2018). All plots were initially planted/sown in fall of 2016 with supplemental planting in spring/summer 2017 based on survival surveys (Supplemental File S1). The final experiment is approximately 1 acre in size consisting of a total of 398 4m2 experimental plots. Throughout the experiment, monocultures and treatment plots have been weeded to maintain only the original species planted. Every plot is burned annually in early spring.

*Sampling cover*

The initial survey (August of 2017) was conducted on only the southwest square meter of each plot while subsequent surveys added the northwest and southeast square meter of each plot. For this reason, analyses in this study that are based on all surveys use only the southwest quarter plot, which was sampled in all surveys. Due to phenological variation among species, analyses presented here treat early season surveys separately from late season surveys.

*Phylogenetic ordinations*

Community phylogenetic ordinations, wherein dissimilarity was based on phylogenetic dissimilarity rather than species dissimilarity, used the functions comdist (for MPD-based calculations) and comdistnt (for MNTD) in the R package *picante* v. 1.8 (Kembel et al. 2010). To evaluate whether there was a turnover of species with greater phylogenetic distinctiveness (long independent branches are preferentially lost), we permuted lost taxa per plot and calculated the mean phylogenetic distinctiveness of species lost for each plot class.

*Permutation tests of taxonomic ordinations*

Because pairwise dissimilarities given ordinations are not independent, we conducted a randomization of taxonomic ordinations to estimate Type-I error (p-values) for shifts or convergence in trait space and used delete-one jackknife on plots to recalculate mean dissimilarities for confidence intervals. We tested for significant movement in ordination space over the sampling intervals by simulating data that allows plots to shift any direction in ordination space but constraining the distance of shifts to be identical to observed distances. This tests the null hypothesis that plots can shift the same distance but in any direction (thus change is not significantly driven by a few specific species). We simulated 1000 positions using the Euclidean distance between individual plots starting distance (at planting) to final position (last survey); fixing this distance, but randomizing the direction. To test for significance, we used a two-tailed p-test given the distribution of simulated centroid distances (vector of change).

*Generalized linear mixed models*

Quantitative fixed covariates were scaled to a mean of zero and unit variance for assessing effect sizes. Models were evaluated using the default priors, which assume a posterior distribution with zero mean and high variance for the fixed effects with flat or weakly informative priors. Models were compared based on minimization of residual variance and the deviance information criterion (DIC; Spiegelhalter et al. 2002). While DIC has been critiqued in part on the grounds that it may tend to favor overfitted models, it represents a computationally tractable model-comparison method for MCMC applications that allows us to evaluate the relative support for less-than-global models (Spiegelhalter et al. 2014). Models were generally run for up to 200,000 iterations with a 10% burn-in, modified as needed based on effective sample sizes. Model convergence was assessed for both fixed and random effects separately by viewing the trace and intercept plots, and by Gelman and Rubin’s (1992) convergence diagnostic implemented in the R package *coda* (Plummer et al. 2006).

Literature Cited

Lichvar, R.W., M. Butterwick, N.C. Melvin, and W.N. Kirchner. 2014. The National Wetland Plant List: 2014 update of wetland ratings. *Phytoneuron* 2014-41: 1–42.

Plummer M., Best N., Cowles K.,Vines K. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R news* 6: 7-11.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:583–639.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2014. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76: 485–493.