

To remain modern the coexistence program requires modern statistical rigour

<https://doi.org/10.1038/s41586-023-06919-3>

David W. Armitage^{1✉}

Received: 20 December 2022

Accepted: 30 November 2023

Published online: 28 August 2024

 Check for updates

ARISING FROM: M. N. Van Dyke et al. *Nature* <https://doi.org/10.1038/s41586-022-05391-9> (2022).

A 2022 study by Van Dyke et al.¹ paired experimental drought manipulations with demographic models and trait data to project major shifts in coexistence among a number of annual plant taxa. A reanalysis of these data that includes comparisons of alternative competition models reveals that the authors' original conclusions are strongly sensitive to model choice. Furthermore, propagating error in model parameters into coexistence predictions results in relatively weak support for the majority of coexistence shifts that were predicted by the authors' original model. These results suggest that there is a need for increased statistical rigour when treating binary predictions of species coexistence as observed experimental outcomes.

Forecasting the effects of precipitation change on plant communities is a major challenge. The theoretical framework of modern coexistence theory has been used to predict the joint contributions of niche and fitness differences (ND and FD, respectively) to competitive outcomes and species coexistence under future precipitation projections^{1,2}. By tracking the demographic rates of plants in experimental communities receiving either reduced or ambient precipitation over a single growing season, Van Dyke et al.¹ use this framework to argue that moderate decreases in water availability will substantially change the predicted coexistence outcomes of 10 out of the 15 pairs of annual plant species under study, and that these shifts are more likely in functionally diverse communities. Such a finding is noteworthy because functional diversity is anticipated to contribute to the maintenance of ecosystem services and is therefore often a desired outcome of restoration and conservation projects³.

However, the authors' results depend heavily on the key assumption that species pairs that satisfy the inequality $\rho < k_j/k_i < 1/\rho$ (where ρ denotes niche overlap and k_j/k_i fitness differences) will stably coexist⁴. In the absence of independent data to benchmark the empirical accuracy of this inequality, it is imperative that the estimates of ND and FD are statistically robust. To this end, Van Dyke et al.¹ omit some important statistical analyses, such as model selection and error propagation, and this affects their conclusions of substantial drought-mediated shifts in coexistence and relationships between trait and fitness differences.

The first issue is that of model specification. There are many ways to write phenomenological competition models that are nearly equivalent in both assumptions and complexity, but which assume slightly different functional forms of density dependence⁵. Following the authors' previous work^{6,7}, Van Dyke et al. assume that a simple form of the Beverton–Holt (BH) competition model best describes the dynamics of their system. Given that the output of the analysis is a theoretically motivated prediction (coexistence or competitive exclusion), and there is no a priori basis to strongly favour the BH model over similar alternatives⁵, then it follows that the model with the best

predictive accuracy on withheld data should be the one that is most trusted to generate the parameter estimates used in subsequent predictions and analyses.

To investigate the sensitivity of model choice on the results, I used a Bayesian approach to sample the posterior distributions of competition (α_{ij}), growth rate (λ_i) and treatment effect parameters for seven different alternative competition models of similar complexity. For each focal species, and using weakly informative priors with the same constraints as those used by the authors (with model 7's $\theta_i \sim \mathcal{N}_{\text{trunc}}(\mu = 1, \sigma = 0.5, a = 0.6, b = 1.4)$ where a and b specify the truncation interval), I ran eight Markov chains of length 10,000, discarding the first 50% as warm-up samples. After confirming Markov chain Monte Carlo convergence and that the posteriors and resulting ND and FD estimates of the BH model matched those from Van Dyke et al.¹, I fitted six alternative model forms, which have been advocated elsewhere^{5,6,8}, to the same data. Comparing models using the Watanabe–Akaike information criterion (WAIC)—a complexity-penalized measure of a model's out-of-sample predictive performance⁹—I identified three models that predicted withheld data better than does the BH model of Van Dyke et al.¹ (Table 1). An exponentiated BH model (no. 7) offered the best improvement in predictive ability and stability compared with other high-ranking models, which either underperformed in leave-one-out prediction or resulted in inflated, unrealistic population equilibria. This model is a generalized form of the standard BH model with an additional parameter, θ , that accommodates a more flexible response-surface shape⁵.

Using 1,000 posterior draws of λ_i , α_{ii} , α_{ij} and θ_i from model 7, I calculated FD and ND (replacing λ_i with λ_i^{1/θ_i} in the new model's ND formula) for each species pair, and assessed whether these draws satisfied the aforementioned coexistence inequality (Fig. 1). Performed over the set of posterior draws for each species pair, this process generates a distribution of coexistence probabilities conditioned on the model, priors and data. It is important to note, however, that the Bayesian posteriors of model 7 can generate zero-net-growth equilibria (that is, seed carrying capacities) up to six times higher than those of the standard BH model (Extended Data Table 1). Because coexistence is predicted using invasion analysis at these equilibria, it is important to acknowledge the potential trade-off between models' predictive performance on observed data (affecting estimates of λ and α), and realism when these are used to extrapolate carrying capacities. As Van Dyke et al.¹ did not conduct competition experiments at abundances near either model's predicted equilibria, it is currently not known which equilibria are more realistic, nor how well either model would perform when fitted to data collected at the relevant densities.

I then calculated the probability that a switch in coexistence outcomes had occurred between treatments. This probability, $p(\text{switch})$,

¹Integrative Community Ecology Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Japan. ✉e-mail: david.armitage@oist.jp

Matters arising

Table 1 | Comparison of various competition models of density-dependent fecundity, F_i , using the WAIC

Model	$\langle \text{WAIC} \rangle$	$\langle \text{SE.WAIC} \rangle$	$\langle \Delta \text{WAIC} \rangle$
(1) $F_i = \lambda_i$	488 ± 117	16.6 ± 4.9	108.0 ± 63.9
(2) $F_i = \lambda_i - \alpha_{ii} N_i - \alpha_{ij} N_j$	423 ± 81	19.3 ± 3.7	42.4 ± 21.2
(3) $F_i = \lambda_i e^{-\alpha_{ii} N_i - \alpha_{ij} N_j}$	423 ± 81	19.3 ± 3.7	42.5 ± 21.3
(4) $F_i = \lambda_i / (1 + \alpha_{ii} N_i + \alpha_{ij} N_j)$ (BH model)	392 ± 68	21.0 ± 5.4	12.0 ± 10.1
(5) $F_i = \lambda_i e^{-\alpha_{ii} \log(N_i + 1) - \alpha_{ij} \log(N_j + 1)}$	383 ± 69	21.5 ± 5.3	2.7 ± 2.1
(6) $F_i = \lambda_i / (1 + N_i^{\alpha_{ii}} + N_j^{\alpha_{ij}})$	382 ± 70	21.5 ± 5.2	1.6 ± 1.9
(7) $F_i = \lambda_i / (1 + \alpha_{ii} N_i + \alpha_{ij} N_j)^{\beta_i}$	382 ± 70	21.4 ± 5.4	2.4 ± 2.6

Values shown are mean ± s.d. over all species × treatment combinations. Models 5, 6 and 7 have a better predictive ability than the authors' BH model (no. 4) does. Comparisons with the corrected Akaike information criterion (AIC_c) and the Bayesian information criterion (BIC) on maximum likelihood fits return quantitatively similar results. SE.WAIC values are the within-treatment standard error of each WAIC score averaged over species × treatment combinations. Δ WAIC denotes the model's average WAIC difference from the top model for a species × treatment combination. For all three metrics, lower values indicate better model performance and stability across treatments.

is defined as $p(C_i \cap E_j)$, $i \neq j$, where $p(C_i)$ is the probability of coexistence in the precipitation treatment i with the highest coexistence probability, and $p(E_j)$ is the probability of exclusion ($= 1 - p(C_j)$) of the other treatment, j . Two key findings emerged. First, that coexistence predictions for most species pairs are highly sensitive to slight variations in model form, which presents a danger when choosing among phenomenological models with equivalent fit statistics. Second, of the original ten species pairs that were predicted to have switched coexistence outcomes between treatments, only four such switches are now predicted at probabilities greater than 0.5 (Fig. 1), including for two species pairs that were scored as not having switched in the original analysis.

Carrying the posterior means of model 7's niche and fitness differences forward through the remaining analyses results in the loss of statistically significant differences between competition and demographic differences between treatments (Extended Data Fig. 1). Furthermore, and perhaps most importantly, changes in FD between treatments are no longer significantly positively associated with the functional trait differences between species pairs (Extended Data Fig. 1). We are left

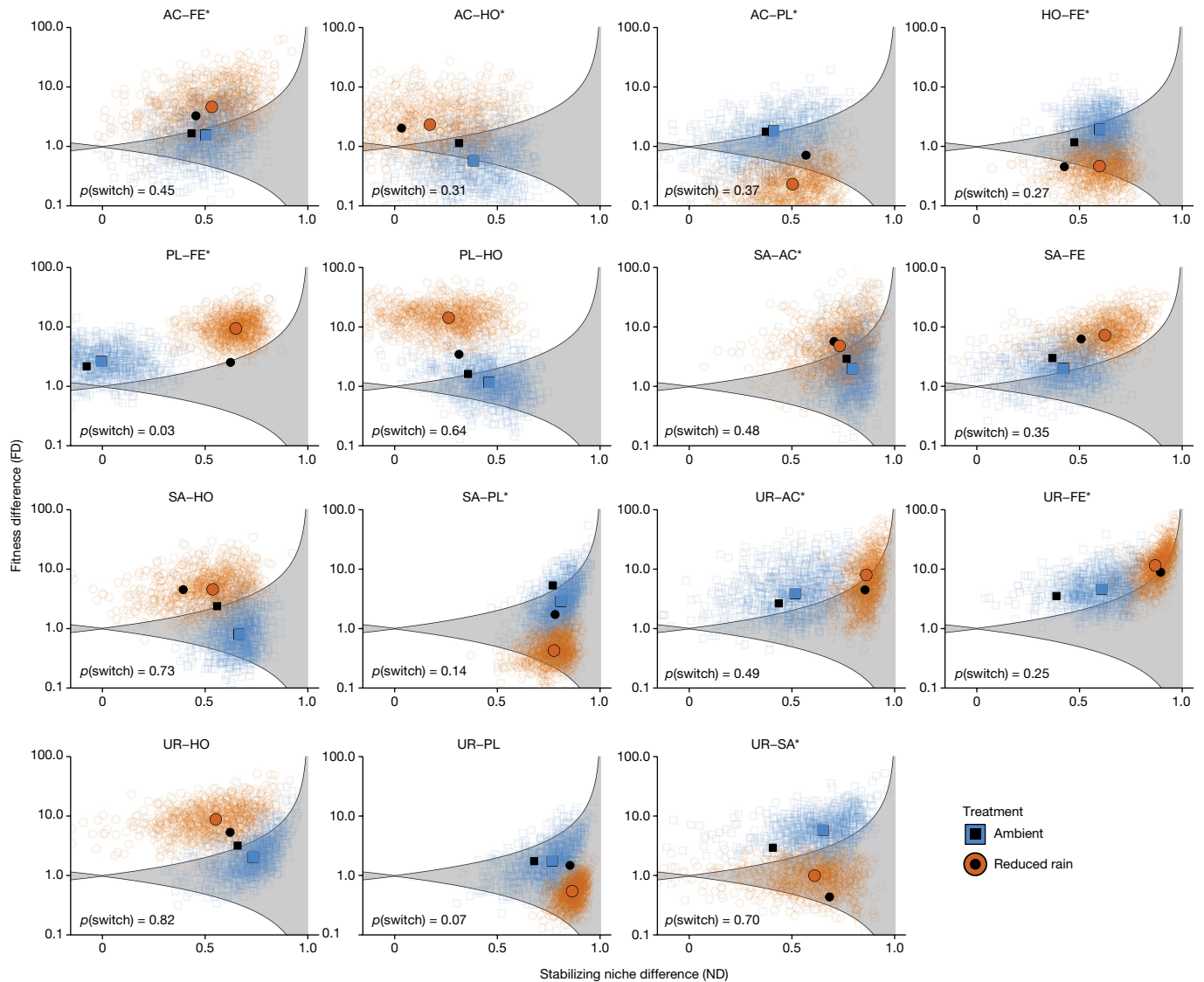


Fig. 1 | Error-aware predictions of coexistence outcomes for 15 species pairs under the best-fit alternative model. Points falling inside of the shaded region are those that satisfy the coexistence inequality. Coloured points represent posterior draws from model 7 ($n = 1000$). Solid coloured shapes are median posterior estimates from this model and black shapes are the authors' estimates

from the BH model. For each panel, the probability that a switch between coexistence and exclusion has occurred is also shown. Asterisks denote species pairs that were predicted to have experienced coexistence shifts in the original analysis. AC, *Acmispon wrangelianus*; FE, *Festuca microstachys*; HO, *Hordeum murinum*; PL, *Plantago erecta*; SA, *Salvia columbariae*; UR, *Uropappus lindleyi*.

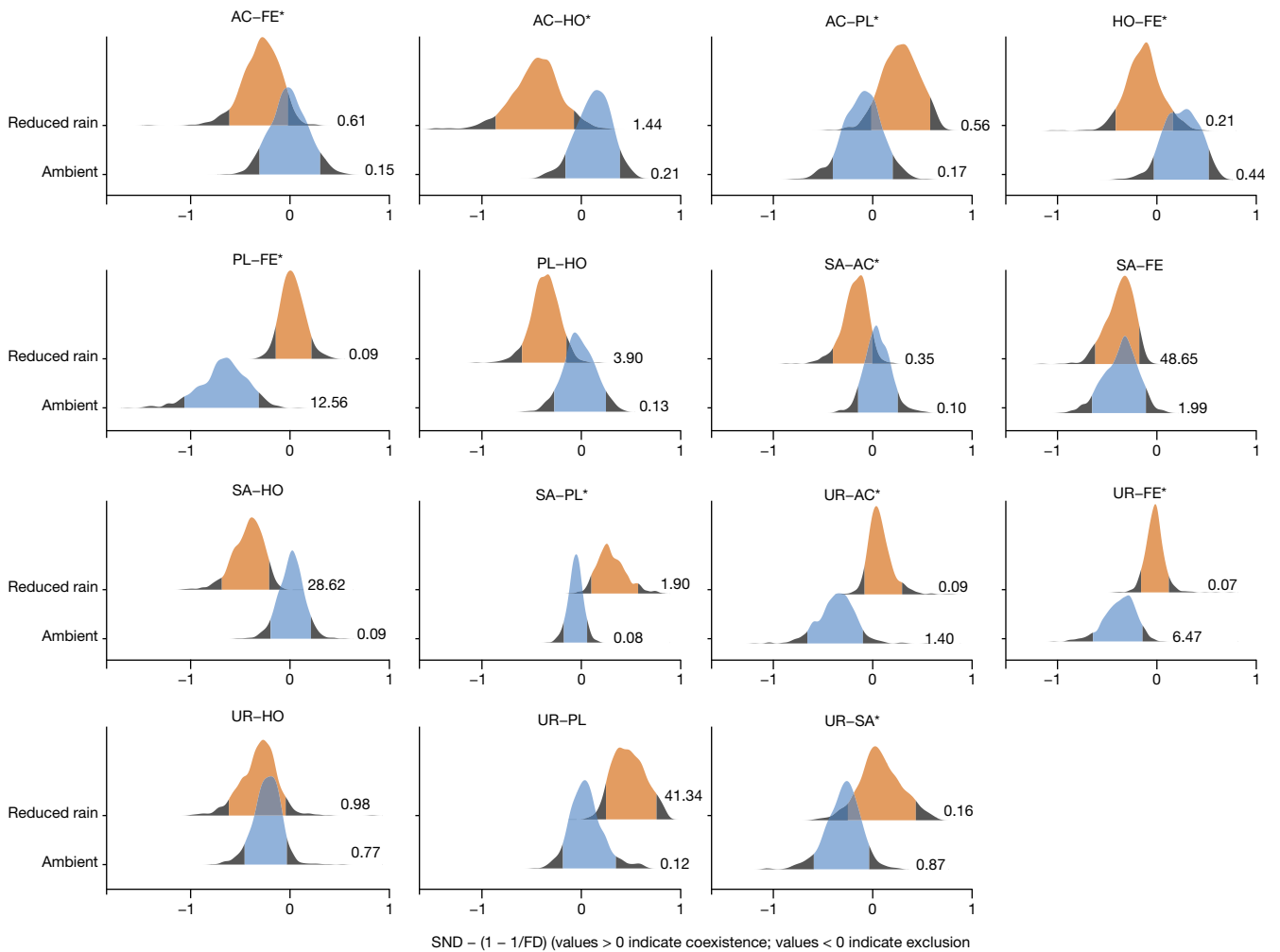


Fig. 2 | Posterior distributions of coexistence predictions for 15 species pairs under the authors' original model. Values greater than zero predict coexistence; values below zero predict exclusion. Grey regions of each distribution fall outside the 89% highest density interval. Bayes factors comparing the observed data to the point null of 0—indicating an undetermined

coexistence or exclusion classification—are shown next to each plot. Values less than one indicate odds in favour of the point null. For example, the value of 0.15 in the first panel indicates the data are $1/0.15 = 6.66$ times more likely under the null hypothesis of undetermined outcome over the alternative of coexistence or exclusion.

to conclude that under a competition model with a better fit to the observed data than the standard BH, many of the major conclusions concerning drought-mediated shifts in coexistence disappear.

It could be argued that despite support for better-performing models, the standard BH model has been so widely used, that it should be considered the preferred model for standardized comparisons across studies. However, even in the rare cases in which these studies present error estimates for ND and FD, decisions concerning predicted coexistence outcomes are rarely evaluated with the same standards of statistical confidence as are used elsewhere in the analyses. This inconsistency is found throughout the analyses of Van Dyke et al.¹. Most of the authors' analyses present statistical evidence in the form of null hypothesis tests with a type I error tolerance of 5%. However, this is abandoned in one key area—decisions about whether or not a species pair is scored as coexisting. Instead, the authors use median values of ND and FD from a nonparametric bootstrap to assign binary outcomes to the predictions with an error tolerance of 50%. Although error bars are provided in a supplementary figure, many clearly transect the coexistence boundary defined by ND and FD, yet this massive uncertainty is ignored by the authors in assigning outcomes to their treatments and concluding that 'substantial shifts' in coexistence have occurred. Although there are no agreed-upon methods for what a null hypothesis test of coexistence predictions should entail, I suggest that

propagating error either through the nonparametric bootstrap samples or through posterior draws of parameters can quantify the degree of support for these competitive outcomes without the need for setting an arbitrary accept or reject criterion.

I illustrate this by using posterior draws from the original BH model to propagate error through to ND and FD estimates. Median values of these draws closely matched the authors' maximum likelihood estimates. I then used the authors' coexistence criterion to generate a posterior distribution of predicted competitive outcomes for each species pair × treatment combination. Bayes factors were used to assess the relative statistical evidence for the point null hypothesis $f(ND, FD) = 0$, where $f(ND, FD) = ND - (1 - FD^{-1})$ versus the alternative of $f(ND, FD) \neq 0$. Support for this null value indicates that a particular scenario could not reliably be classified as coexistence or exclusion, but rather somewhere in-between¹⁰. To quantify this support, Bayes factors were calculated using the Savage–Dickey density ratio with a uniform prior bounded from -2 to 1. Using the standard evidence-based interpretation of Bayes factors¹¹, the alternative hypothesis of $f(ND, FD) \neq 0$ is only supported by the data for 9 of the authors' 30 original predictions (Fig. 2). Instead, most of the data support a boundary scenario in which neither coexistence nor exclusion of either species can be assigned with confidence, but rather both are plausible at non-trivial probabilities. In other words, the predictions of coexistence or

Matters arising

exclusion being made under an experimentally parameterized BH model are not particularly strong, and, as such, many of the authors' predicted shifts between coexistence and exclusion could—under an error-aware reading—be more fairly characterized as shifts between one uncertain and one higher-confidence outcome (for example, the PL-FE and SA-HO species pairs) or as two slightly offset distributions both occupying a region of high uncertainty (for example, the AC-FE and SA-AC species pairs).

Looking forward, researchers are encouraged to move from binary, all-or-nothing predictions of species coexistence to probabilistic, error-inclusive metrics more transparent in their predictions^{10,12}. Crucially, since coexistence predictions made from pairwise experiments have yet to be sufficiently vetted with independent data^{13,14}, practitioners should exercise caution when using the approach employed herein to forecast the effects of climate change on communities. Spatial and temporal replication of experiments beyond a single site and single generation would help resolve some of this uncertainty. Likewise, moving beyond phenomenological competition models of species interactions to more mechanistic formulations¹⁵ will reduce the need for bias-prone model selection and permit an explicit accounting of the various limiting factors that give rise to niche and fitness differences between competitors.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-023-06919-3>.

Code availability

Code to replicate this analysis is available at <https://doi.org/10.5281/zenodo.7460881>

1. Van Dyke, M. N., Levine, J. M. & Kraft, N. J. B. Small rainfall changes drive substantial changes in plant coexistence. *Nature* **611**, 501–511 (2022).
2. Matias, L., Godoy, O., Gómez-Aparicio, L. & Pérez-Ramos, I. M. An experimental extreme drought reduces the likelihood of species to coexist despite increasing intransitivity in competitive networks. *J. Ecol.* **106**, 826–837 (2018).
3. Díaz, S. & Cabido, M. Vive la différence: plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.* **16**, 646–655 (2001).
4. Chesson, P. in *Ecological Systems: Selected Entries from the Encyclopedia of Sustainability Science and Technology* (ed. Leemans, R.) 223–256 (Springer, 2013).
5. Law, R. & Watkinson, A. R. Response-surface analysis of two-species competition: an experiment on *Phleum arenarium* and *Vulpia fasciculata*. *J. Ecol.* **75**, 871–886 (1987).
6. Levine, J. M. & HilleRisLambers, J. The importance of niches for the maintenance of species diversity. *Nature* **461**, 254–257 (2009).
7. Kraft, N. J. B., Godoy, O. & Levine, J. M. Plant functional traits and the multidimensional nature of species coexistence. *Proc. Natl Acad. Sci. USA* **112**, 797–802 (2015).
8. Hart, S. P., Freckleton, R. P. & Levine, J. M. How to quantify competitive ability. *J. Ecol.* **106**, 1902–1909 (2018).
9. Watanabe, S. A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **14**, 867–897 (2013).
10. Bowler, C. H., Weiss-Lehman, C., Towers, I. R., Mayfield, M. M. & Shoemaker, L. G. Accounting for demographic uncertainty increases predictions for species coexistence: a case study with annual plants. *Ecol. Lett.* **25**, 1618–1628 (2022).
11. Jeffreys, H. *Theory of Probability* 3rd edn (Oxford Univ. Press, 1961).
12. Terry, J. C. D., Chen, J. & Lewis, O. T. Natural enemies have inconsistent impacts on the coexistence of competing species. *J. Anim. Ecol.* **90**, 2277–2288 (2021).
13. Godoy, O. & Levine, J. M. Phenology effects on invasion success: insights from coupling field experiments to coexistence theory. *Ecology* **95**, 726–736 (2014).
14. Armitage, D. W. & Jones, S. E. Coexistence barriers confine the poleward range of a globally distributed plant. *Ecol. Lett.* **23**, 1838–1848 (2020).
15. Stouffer, D. B. A critical examination of models of annual-plant population dynamics and density-dependent fecundity. *Methods Ecol. Evol.* **13**, 2516–2530 (2022).

Acknowledgements I thank M. Van Dyke, C. Terry and J. Spaak for constructive discussion and debate.

Author contributions D.W.A. performed all writing and analyses.

Competing interests The author declares no competing interests.

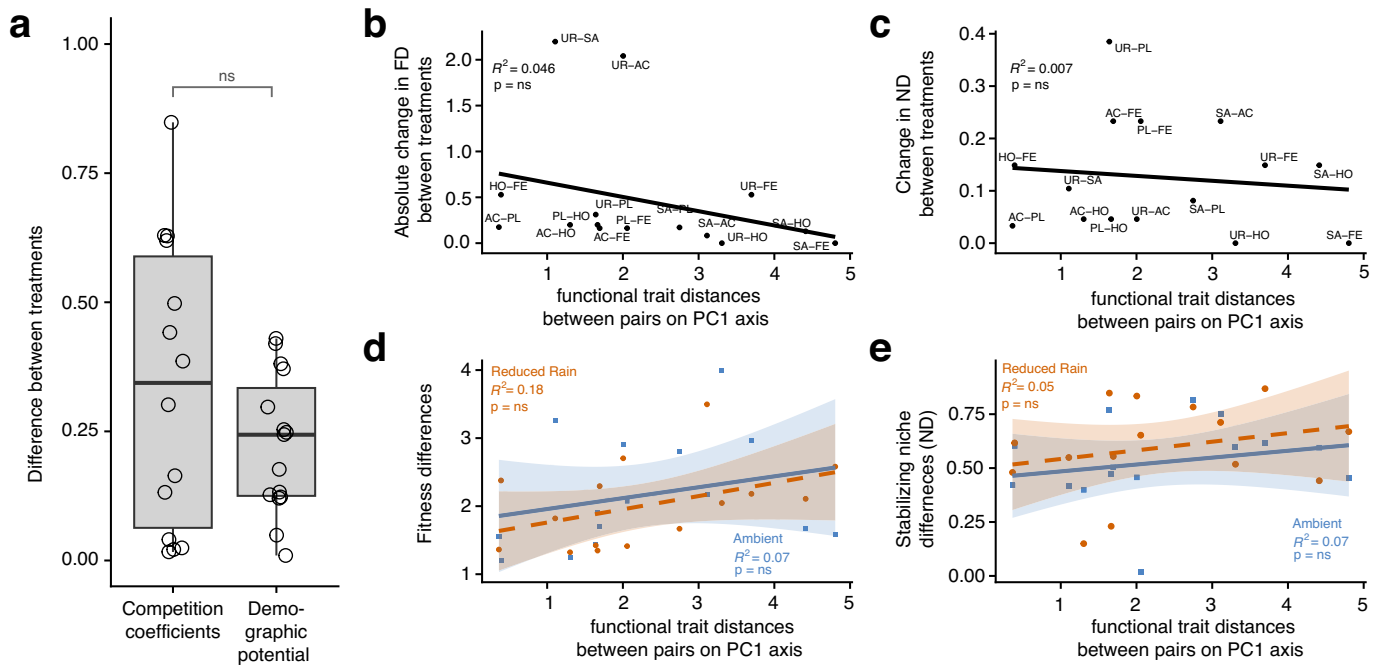
Additional information

Correspondence and requests for materials should be addressed to David W. Armitage.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2024



Extended Data Fig. 1 | An alternative competition model removes the predicted effects of trait differences on coexistence mechanisms. a–e. Using coefficients of model 7 has the effect of removing statistically significant trends in the relative magnitudes of treatment differences in demographic potential and competition coefficients ($t = -1.68$, $p = 0.10$) (a). This also has the

consequence of removing the positive associations between functional trait distances and absolute changes in FD between treatments (b), absolute changes in ND between treatments (c), and overall fitness differences (d) and overall niche differences (e). ND results remain unchanged from the original analysis.

Matters arising

Extended Data Table 1 | Comparisons of predicted equilibria

Species	Ambient rain			Reduced rain		
	M.4	M.7	M.7/M.4	M.4	M.7	M.7/M.4
ACWR	8,055	28,132	3.5	4,766	18,579	3.9
FEMI	3,619	8,921	2.5	2,665	10,144	3.8
HOMU	776	1,746	2.3	1,481	8,473	5.7
PLER	1,530	3,136	2.1	645	719	1.1
SACO	6,722	31,617	4.7	3,962	18,360	4.6
URLI	6,166	12,402	2.0	2,988	8,076	2.7

Comparison of equilibrium abundance predictions by model 4 (BH) and model 7.