Uncertain competition coefficients undermine inferences about coexistence

https://doi.org/10.1038/s41586-023-06859-y

Received: 25 November 2022

Accepted: 8 November 2023

J. Christopher D. Terry[1]✉

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

Improving our understanding of how climate and competitive interactions act together to determine species ranges is crucial to address the challenges of global change[1]. In their 2022 paper, Van Dyke et al.[2] make two principal claims: (1) that a drought treatment causes qualitative changes in the predicted coexistence outcome in 10/15 pairwise competition trials; and (2) that this effect on coexistence is principally due to drought-induced changes in the impact of competition. Here I highlight the substantial uncertainties in the estimates of their model parameters, identify the low support for treatment effects on competition coefficients and show that differential uncertainties in key quantities could generate artefacts that are capable of explaining the observations.

The authors fitted a frequently used model, $F_i = \frac{\lambda_i}{1 + \alpha_{ii}N_i + \alpha_{ij}N_j}$, to observed seed production rates from a controlled trial, where F_i is the per-germinant fecundity of focal species i, λ_i is the fecundity without any competition, α_{ii} and α_{ij} are the intra- and interspecific competition coefficients and N_i and N_j are the number of intra- and interspecific competitors. They separately fitted models for each focal species for each of the two levels of their water-level treatment. Accurately estimating interactions between species is a well-known and long-standing challenge in ecology. The α terms require assessment over a range of competitor densities, and their values will almost always be more uncertain than those of the λ terms. This is because λ can be measured directly by trials without additional noise introduced by competition, and can be informed by all experiments involving that species.

Although the number of trials used was large (1,677), so too was the number of ecological parameters to be estimated $((6\lambda + 6^2\alpha) \times 2$ treatments $= 84)$. In principle, this leaves just under 20 data points per parameter. However, because the sampling design depended on variable germination rates and there were more observations of monoculture trials, the number of observations for some species pairs was much lower, with a low of just nine observations in one case (UR focal, FE background, ambient water treatment; species codes are defined in the footnote for Table 1). The finding by the authors that changes between treatments in the α coefficients dominate the reported effect of reduced rainfall treatment would be a surprising and important advance, but invites a certain degree of scepticism—especially with the large degree of noise in the original data (Extended Data Fig. 1).

The authors did not report any statistical assessment to support the foundational claim that drought significantly affected the α terms. Using the authors' original data and models, I used standard model selection approaches to test whether the differentiation of competition treatments by drought treatment is supported (Supplementary Methods 1). I fitted four sets of models: (1) not including any drought treatment differentiation; (2) allowing drought to influence the λ terms; (3) allowing drought to influence the α terms; and (4) allowing drought to influence both the λ and the α terms (that is, the original model). Following the original approach as closely as possible, models were separately fitted for each of the six focal species, allowing all available data to be used to contribute to the estimation of the λ terms.

Akaike information criterion (AIC) comparisons are shown in Table 1. In no case is the original model uniquely well supported by the standard threshold of at least a 2 AIC unit improvement. For only one species is a treatment effect on α terms confidently supported (*Plantago erecta*; PL). Furthermore, using the Bayesian information criterion (BIC) (that is, a $\log(n)k$ penalty for introducing additional parameters)—which is generally favoured when seeking to make inferences rather than predictions[3]—the original model is never well supported (Extended Data Table 1). Although this analysis cannot exclude the possibility that individual α terms are affected by the drought, to mitigate the threat posed by such multiple comparisons in such an analysis it would be necessary to raise the statistical threshold considerably. Because small changes to individual α terms can have major effects on predicted coexistence, either all uncertainty should be propagated through to the end conclusions, or careful statistical support should be developed for treatment effects.

Demonstrating that competition is indeed affected by the treatment is an essential precursor to subsequent analyses and inferences. There is a particular risk that comparisons between the contribution of competition coefficients relative to demographic potential (Fig. 3 in the original paper) will give erroneous results by simply capturing differences in uncertainty. In the authors' original bootstrap samples, the average relative standard deviation (σ/μ) of the α terms is more than double that of the λ terms (0.74 compared with 0.32). To quantify the potential contribution of these effects on the authors' results, I generated 1,000 artificial datasets with the same size and underlying variation as that of the original data, but without any between-treatment differences (Supplementary Methods 2). I did this by taking draws from the prediction posterior of a seed production model with the same structure as that of the original model, but without any treatment terms fitted to the raw data, using the R package brms (ref. 4). In this simulated dataset, there therefore exists no 'true' difference between the treatments. To assess the potential for differential uncertainty to generate artefacts, I then applied the original analysis pipeline: first identifying best-fit parameters for each treatment group; then, for each species pair, comparing whether the ratios between the competition coefficients or demographic potential was higher; and finally conducting a t-test across all species pairs. In 48% of cases, the analysis pipeline found that the 'change' in the competition coefficients was significantly ($P < 0.05$) larger than the apparent change in the demographic potential ratio.

The simulated 'no treatment' data can also be used as a reference to assess the extent to which the observations show larger changes in the

---

[1]Department of Biology, University of Oxford, Oxford, UK. ✉e-mail: christopher.terry@biology.ox.ac.uk

Table 1 | Model comparison to assess the support for identifying an impact of the water treatment on the coefficients of the seed production model

		Model			
		No treatment effects	Treatment affects only λ	Treatment affects only α terms	Treatment affects α and λ (original)
Number of parameters per focal species		7	8	13	14
Focal species	***n***	**AIC**	**AIC**	**AIC**	**AIC**
AC	294	890.6	889.2	890.5	892.1
FE	305	848.8	843.3	853.6	842.4
HO	214	551.5	553.2	553.8	555.4
PL	325	894.9	881.6	868.0	870.0
SA	274	829.1	810.5	810.0	810.3
UR	266	679.2	680.0	678.5	679.8

The best (lowest) AIC and those values that are within 2 AIC of the best supported model are underlined. Species codes follow the original paper (AC, *Acmispon wrangelianus*; FE, *Festuca microstachys*; HO, *Hordeum murinum*; PL, *Plantago erecta*; SA, *Salvia columbariae*; UR, *Uropappus lindleyi*). See Extended Data Table 1 for equivalent results using the BIC.

Table 2 | Fraction of bootstrap samples that support the coexistence scenarios for the fifteen species pairs

	Most likely scenario			Second-most likely scenario		
Species pair	**Ambient**	**Drought**	**Bootstrap fraction**	**Ambient**	**Drought**	**Bootstrap fraction**
Identified as changing						
SA–PL	SA wins	Coexist	0.714	Coexist	Coexist	0.284
UR–SA	UR wins	Coexist	0.702	UR wins	SA wins	0.158
HO–FE	Coexist	FE wins	0.698	Coexist	Coexist	0.203
AC–HO	Coexist	AC wins	0.698	AC wins	AC wins	0.217
SA–AC	Coexist	SA wins	0.659	SA wins	SA wins	0.245
UR–AC	UR wins	Coexist	0.621	Coexist	Coexist	0.159
PL–FE	PL wins	Coexist	0.548	PL wins	PL wins	0.447
AC–PL	AC wins	Coexist	0.546	Coexist	Coexist	0.309
AC–FE	Coexist	AC wins	0.515	AC wins	AC wins	0.415
UR–FE	UR wins	Coexist	0.493	UR wins	UR wins	0.453
Identified as not changing						
SA–FE	SA wins	SA wins	0.984	Coexist	SA wins	0.014
UR–PL	Coexist	Coexist	0.841	UR wins	Coexist	0.091
SA–HO	SA wins	SA wins	0.568	Coexist	SA wins	0.425
PL–HO	PL wins	PL wins	0.557	Coexist	PL wins	0.440
UR–HO	UR wins	UR wins	0.550	Coexist	UR wins	0.356

The first ten species were identified as showing a qualitative change in coexistence outcome owing to the drought treatment in the original paper, and are sorted by the bootstrap fraction that supports the most likely outcome. Species codes follow the original paper and are defined in the footnote for Table 1.

log-ratio of the competition coefficients or the demographic potentials after accounting for the different uncertainties. Subtracting the average change observed in the null model for each log-ratio estimates the bias-corrected change. Repeating the paired *t*-test presented in the original Fig. 3 with these values gives a non-significant result ($t = -1.061$, degrees of freedom = 14, $P = 0.306$; Supplementary Methods 2). Furthermore, comparing the overall cross-species average difference in these key ratios with the differences observed in 10,000 draws of the null model distribution suggests that the difference is not significantly different from the null expectation ($P = 0.116$, Supplementary Methods 2, section 7). Note that because the direction of changes is uncertain, care needs to be taken when handling the uncertainty in fitted estimates of changes in these key ratios (Supplementary Methods 2, section 7). Hence, the original result identifying the dominant role of changes in species interactions over changes to growth rates is likely to be an artefact of greater uncertainty in the competition coefficients.

The authors conduct bootstrapping of their original data to assess uncertainty in individual parameters, and they present confidence intervals (central 66%) around the fitness differences and niche differences from this bootstrapping in their Extended Data Fig. 1. The largely non-overlapping error bars presented in their Extended Data Fig. 1 might suggest that concerns about the statistical significance of the treatment effect are unwarranted. However, classic vertical and horizontal error bars can be misleading because the shape of the error distribution (whether it is a posterior, bootstrap sample or likelihood surface) on a coexistence plane is rarely round, owing in part to the inherent inseparability of λ and α terms. For this reason, it is best practice to calculate and present the proportion of the final result that falls into each qualitative coexistence outcome[5–7].

To quantify the consequences for the uncertainty in the final assessment of change in coexistence outcomes, I propagated forward the whole of the authors' bootstrap multivariate sample through to the final prediction of long-term coexistence outcome (Supplementary Methods 1, plotted here in Extended Data Fig. 2). For each species pair, I calculated the inferred probability from the original bootstrap sample of each outcome (coexistence or exclusion of either species) in each treatment, and then calculated the inferred likelihood of each possible scenario of treatment effect (for example, coexistence under the ambient treatment but a win for species A under the drought treatment, or coexistence under both treatments). Across the ten cases identified as changing, the estimated chance of seeing the identified shift had a mean of 62% (Table 2) and in several cases was not much larger than the probability associated with a scenario in which the outcome was unchanged. Selecting a particular threshold to definitively identify a change in prediction is a challenge with multivariate results, and in most cases a 'narrative verdict' describing the probabilities associated with each predicted coexistence outcome is likely to be most informative. In cases in which the outcomes of interactions are inherently uncertain and near boundaries, there will inevitably be a sizeable identified probability of change, even when there is no true treatment effect. In the simulated 'no treatment' data, species pairs apparently changed in their best-fit coexistence outcome between treatments in, on average, 22.9% of draws (range 1.5% SA–FE, 40% UR–SA; Supplementary Methods 2, section 4).

Overall, these reanalyses highlight how the inherent difficulty in accurately characterizing competition parameters has cascading consequences for interpreting empirical analyses of coexistence. In particular, they show that there is not strong evidence in these data for the particular mechanistic route identified by Van Dyke et al.[2]; namely, that changes in coexistence are driven by the impact of the drought treatment on intra- and interspecific density dependence. These challenges are in addition to the questions posed by the fundamental uncertainty in the functional form of competition or the extent to which other processes, such as higher-order interactions, influence coexistence in 'real' systems.

Modern coexistence theory has provided bounteous mathematical insight[8–11], but key results and predictions can be highly dependent on accurate parameter values, representative model form and the validity of wider assumptions[6,7,12]. In this context, it is crucial to acknowledge and directly address uncertainties in data. Three general recommendations are worth reiterating: (1) it is essential to include plots of raw data and model fits to gauge the accuracy of model fits; (2) when making inferences (rather than just, say, pure predictions[13]) there is always a need for explicit statistical assessment of some kind to assess the evidence for the impact of any treatment variable; and (3) it can be highly informative to propagate uncertainty as far as possible through to the key results[5–7,14], whether in Bayesian or frequentist frameworks.

It is undisputed that empirical tests of coexistence theory are hugely valuable to enable a better understanding of how ecological communities might respond to future threats[15]. Nonetheless, the data requirements for building a reliable empirical foundation are high and cannot be easily circumvented.

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/s41586-023-06859-y.

Data availability

All reanalyses used only the publicly available data with the original publication.

Code availability

All R code scripts used are detailed in the Supplementary Methods and are also available at https://github.com/jcdterry/CoexistUncertain_public.

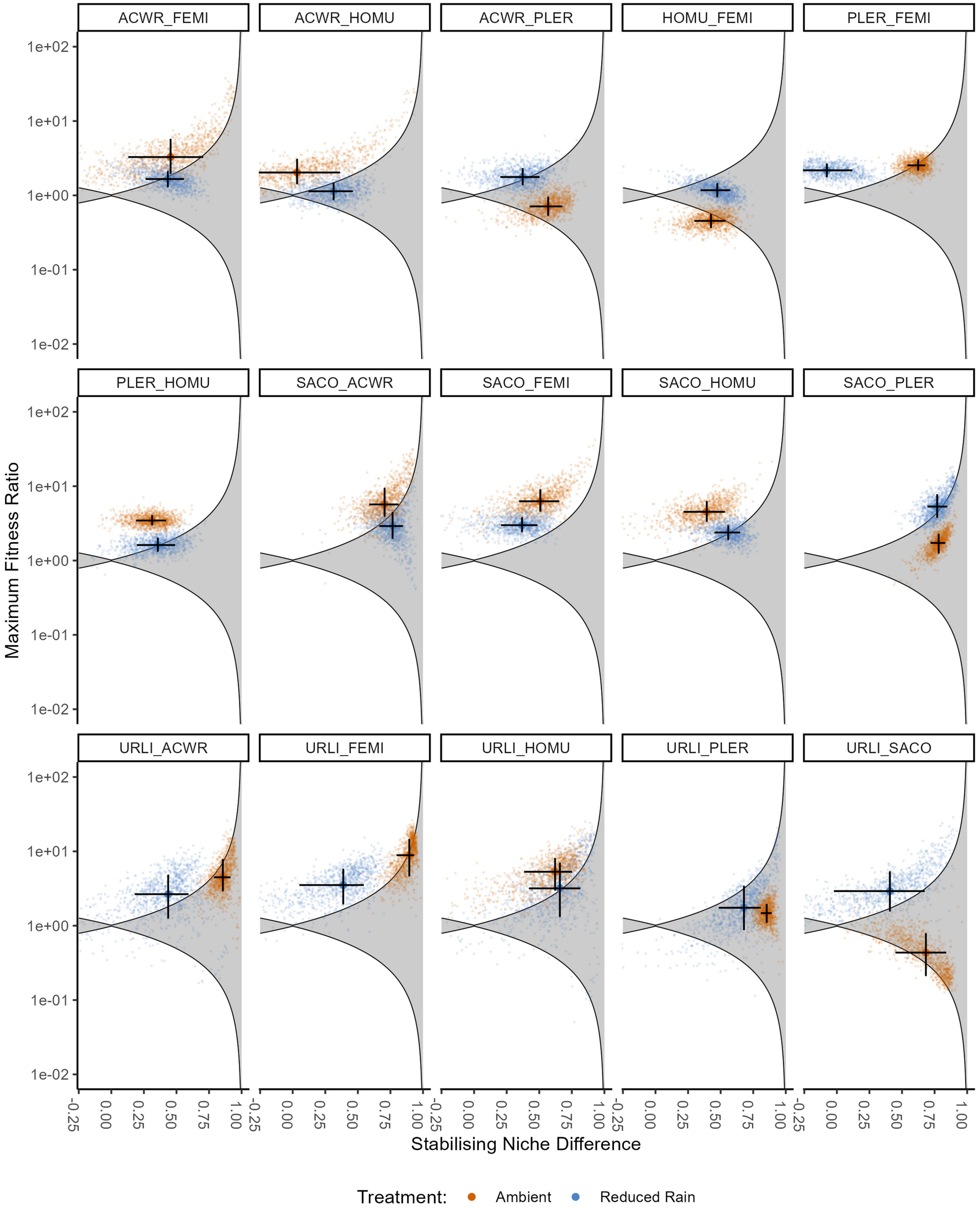
1. Urban, M. C. et al. Improving the forecast for biodiversity under climate change. *Science* **353**, aad8466 (2016).
2. Van Dyke, M. N., Levine, J. M. & Kraft, N. J. B. Small rainfall changes drive substantial changes in plant coexistence. *Nature* **611**, 507–511 (2022).
3. Aho, K., Derryberry, D. & Peterson, T. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* **95**, 631–636 (2014).
4. Bürkner, P.-C. Advanced Bayesian multilevel modeling with the R package brms. *R J.* **10**, 395–411 (2018).
5. 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).
6. 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).
7. Song, C., Uricchio, L. H., Mordecai, E. A. & Saavedra, S. Understanding the emergence of contingent and deterministic exclusion in multispecies communities. *Ecol. Lett.* **24**, 2155–2168 (2021).
8. Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Evol. Syst.* **31**, 343–366 (2000).
9. Adler, P. B., HilleRisLambers, J. & Levine, J. M. A niche for neutrality. *Ecol. Lett.* **10**, 95–104 (2007).
10. Barabás, G., D'Andrea, R. & Stump, S. M. Chesson's coexistence theory. *Ecol. Monogr.* **88**, 277–303 (2018).
11. Alexander, J. M., Diez, J. M., Usinowicz, J. & Hart, S. P. Species' distributions as a coexistence problem: a response to Godsoe et al. *Trends Ecol. Evol.* **33**, 144–145 (2018).
12. Terry, J. C. D., O'Sullivan, J. D. & Rossberg, A. G. Synthesising the multiple impacts of climatic variability on community responses to climate change. *Ecography* **2022**, e06123 (2022).
13. Tredennick, A. T., Hooker, G., Ellner, S. P. & Adler, P. B. A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology* **102**, e03336 (2021).
14. Uricchio, L. H., Daws, S. C., Spear, E. R. & Mordecai, E. A. Priority effects and nonhierarchical competition shape species composition in a complex grassland community. *Am. Nat.* **193**, 213–226 (2019).
15. Alexander, J. M., Diez, J. M., Hart, S. P. & Levine, J. M. When climate reshuffles competitors: a call for experimental macroecology. *Trends Ecol. Evol.* **31**, 831–841 (2016).

Author contributions J.C.D.T. performed all analyses and wrote the paper.

Competing interests The author declares no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-06859-y.
Correspondence and requests for materials should be addressed to J. Christopher D. Terry.
Reprints and permissions information is available at http://www.nature.com/reprints.

Extended Data Fig. 1 | Raw seed production data used to fit the underlying demographic model and best-fit model predictions in both treatment cases. Note that seed production is on a log scale, as used in the model fits, and that the *x* axis differs for each background species.



Extended Data Fig. 2 | Full distribution of 1,000 bootstrap samples of the predicted coexistence outcome. Error bars show the 'best fit' model of the original paper, calculated from the interval of the 16th to 84th quantile of the bootstrap sample of maximum fitness ratio or stabilizing niche difference. The right-hand grey region is the region of coexistence. Note the irregular shape of the distribution of samples.

Extended Data Table 1 | Model comparison as per Table 1, but using the Bayesian (or Schwarz) information criterion ($\log(n)k-2\log(L)$)

		Model			
		No Treatment effects	***Treatment affects only λ***	***Treatment affects only α terms***	***Treatment affects α and λ (Original)***
Number of parameters per focal species		7	8	13	14
Focal Species	**n**	**BIC**	**BIC**	**BIC**	**BIC**
AC	294	916.3	918.7	938.4	943.6
FE	305	874.9	873.1	902.0	894.4
HO	214	575.1	580.1	597.5	602.5
PL	325	921.4	911.8	917.2	923.0
SA	274	854.4	839.4	857.0	860.9
UR	266	704.3	708.6	725.1	730.0

Models within two units of the lowest are underlined. The models that include treatment effect on the alpha terms are consistently assigned considerably worse (higher) values.

Matters arising

To remain modern the coexistence program requires modern statistical rigour

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

Received: 20 December 2022

Accepted: 30 November 2023

David W. Armitage[1]✉

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 projects[3].

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 coexist[4]. 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 dependence[5]. 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 alternatives[5], 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 performance[9]—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 shape[5].

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(switch),

---

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

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

Model	⟨WAIC⟩	⟨SE.WAIC⟩	⟨ΔWAIC⟩
(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) $F_i=\lambda_i/(1+\alpha_{ii}N_i+\alpha_{ij}N_j)^{\theta_i}$	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. ΔWAIC 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

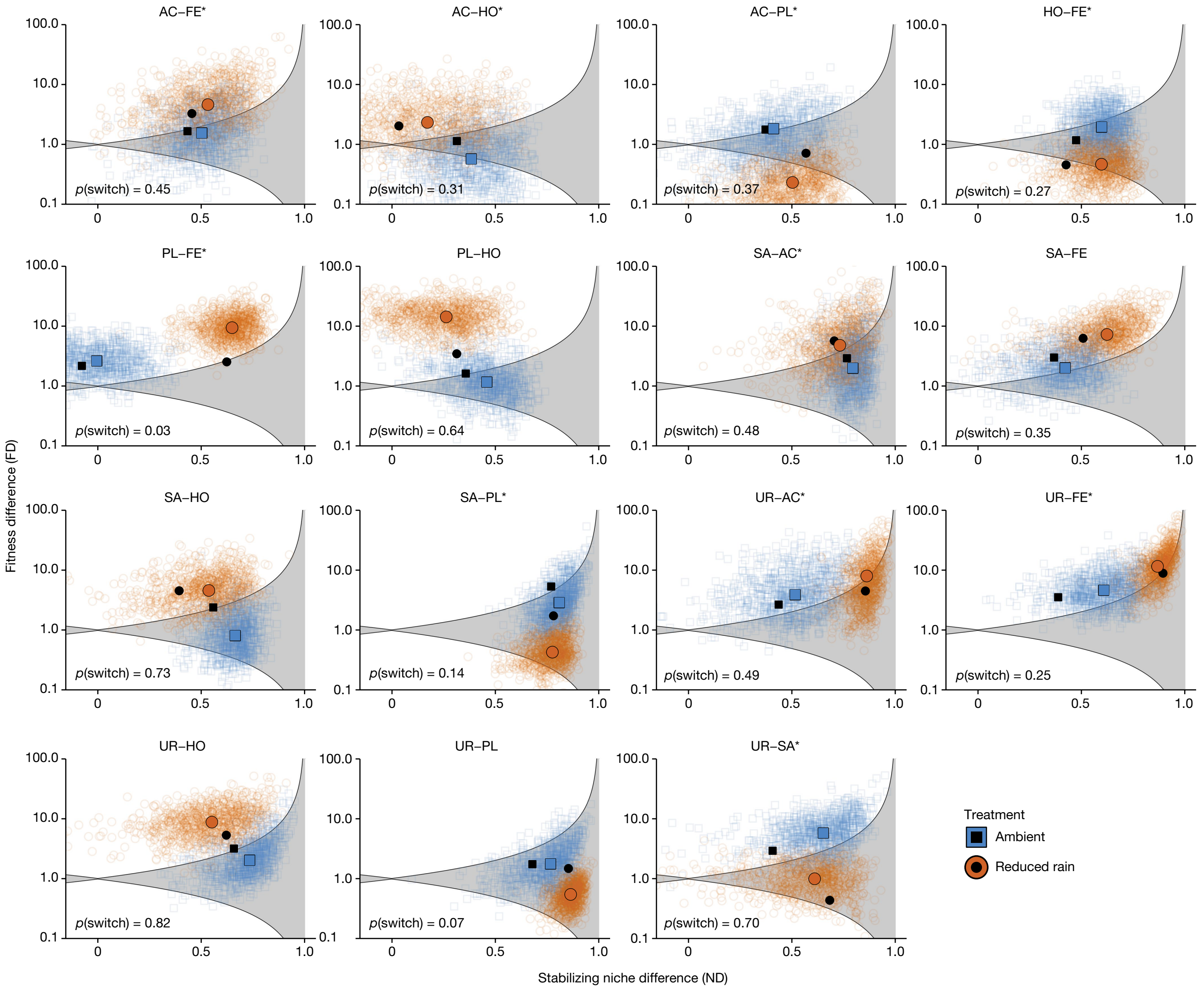


Fig. 1 | Error-aware predictions of coexistence outcomes for 15 species pairs under the best-fit alternative model. Points falling inside of the shaded region are those that satisfy the coexistence inequality. Coloured points represent posterior draws from model 7 ($n = 1000$). Solid coloured shapes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

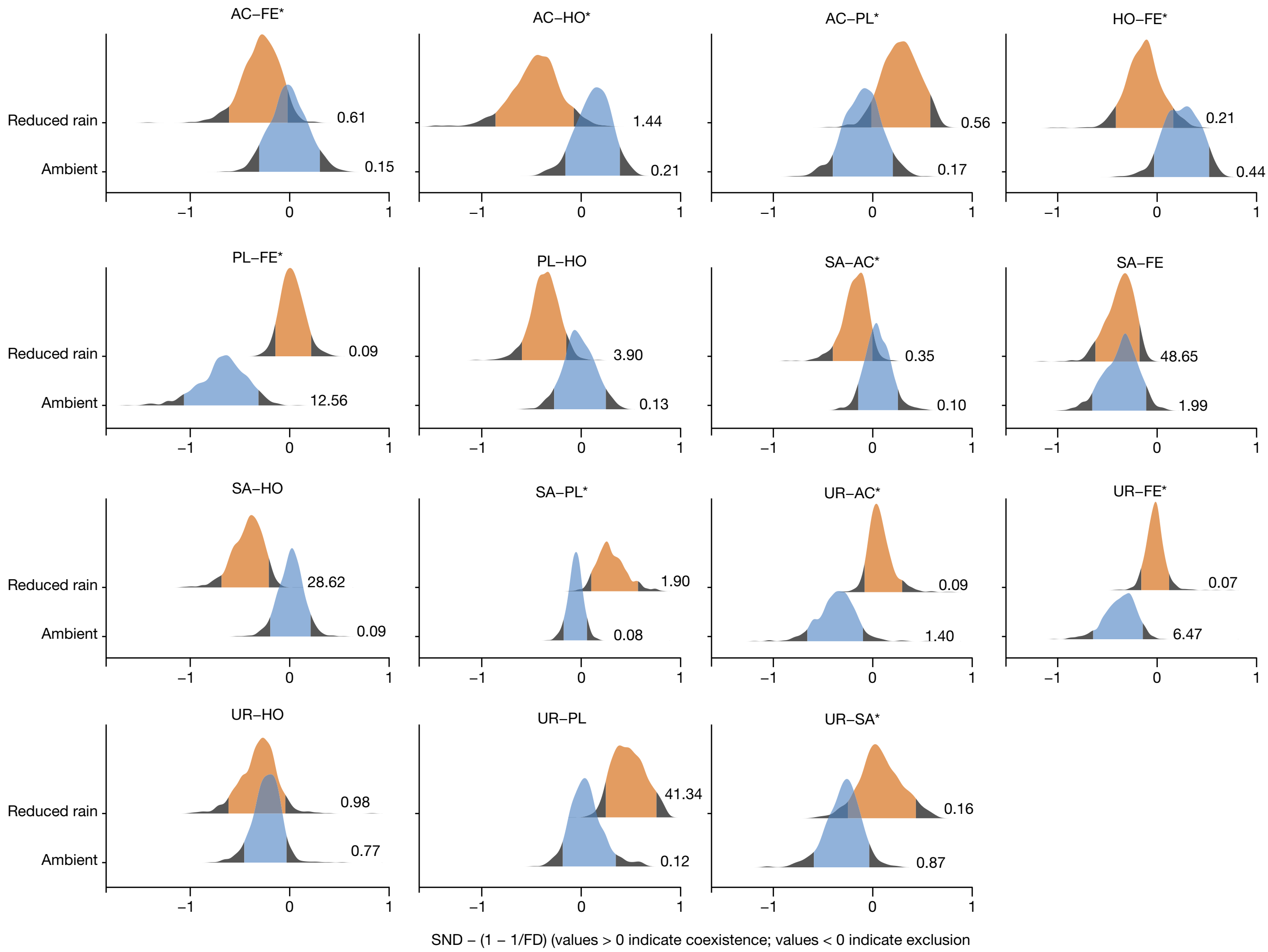


Fig. 2 | Posterior distributions of coexistence predictions for 15 species pairs under the authors' original model. Values greater than zero predict coexistence; values below zero predict exclusion. Grey regions of each distribution fall outside the 89% highest density interval. Bayes factors comparing 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 $f(\mathrm{ND}, \mathrm{FD}) = 0$, where $f(\mathrm{ND}, \mathrm{FD}) = \mathrm{ND} - (1 - \mathrm{FD}^{-1})$ versus the alternative of $f(\mathrm{ND}, \mathrm{FD}) \neq 0$. Support for this null value indicates that a particular scenario could not reliably be classified as coexistence or exclusion, but rather somewhere in-between[10]. 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 factors[11], the alternative hypothesis of $f(\mathrm{ND}, \mathrm{FD}) \neq 0$ 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 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 formulations[15] 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

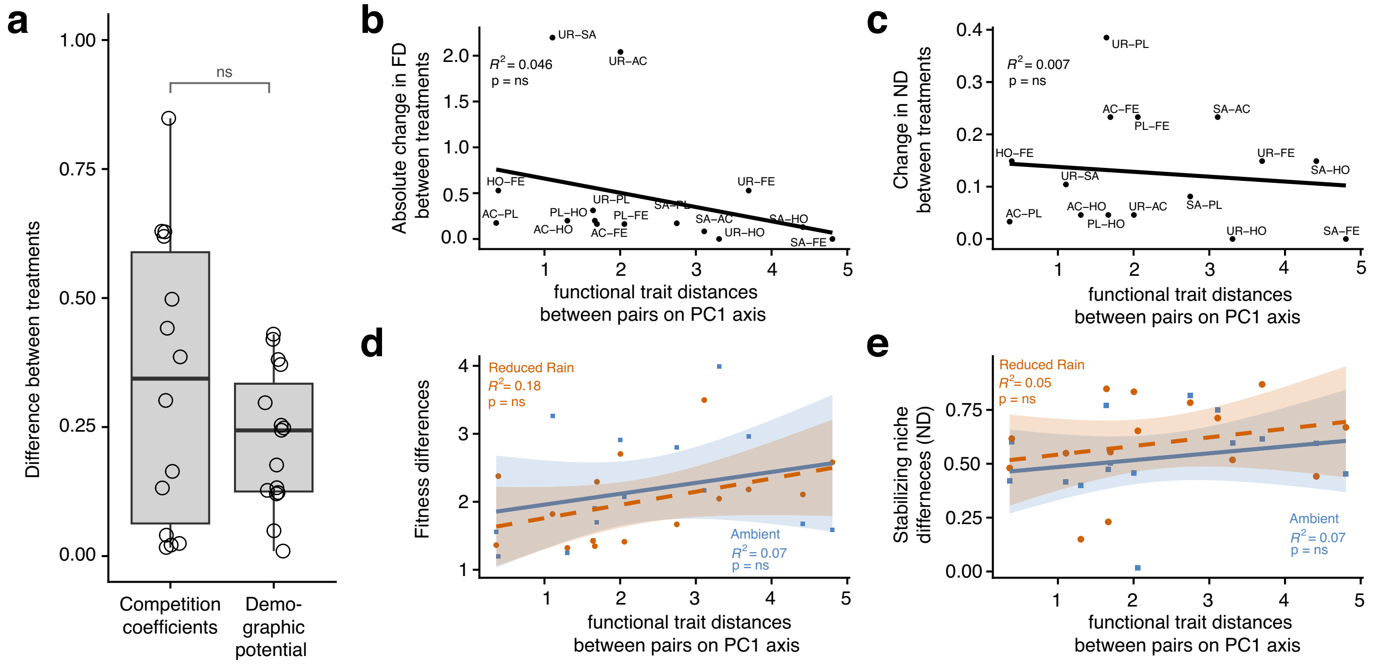
1. Van Dyke, M. N., Levine, J. M. & Kraft, N. J. B. Small rainfall changes drive substantial changes in plant coexistence. *Nature* **611**, 501–511 (2022).
2. Matías, L., Godoy, O., Gómez-Aparicio, L. & Pérez-Ramos, I. M. An experimental extreme drought reduces the likelihood of species to coexist despite 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.



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.

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.

M. N. Van Dyke et al. reply

https://doi.org/10.1038/s41586-024-07777-3

Mary N. Van Dyke[1,2]✉, Jonathan M. Levine[3] & Nathan J. B. Kraft[2]

REPLYING TO: J. C. D. Terry *Nature* https://doi.org/10.1038/s41586-023-06859-y (2024)

REPLYING TO: D. W. Armitage *Nature* https://doi.org/10.1038/s41586-023-06919-3 (2024)

In our paper[1], we fitted a population dynamic model to annual plant demographic data, enabling us to predict the competitive outcome between 15 species pairs and how these change with a rainfall manipulation. Both Terry[2] and Armitage[3] raise several questions about how modern coexistence theory can be best used to make predictions about changes to the outcomes of species interactions. Their concerns focus primarily on model selection and how uncertainty in the estimated parameters in our competition model is used and interpreted. While our analyses[1] closely follow or expand on existing conventions in our field[4–12], the issues Terry[2] and Armitage[3] raise are important to address going forward, especially as statistical methods continue to evolve, and we appreciate the opportunity to contribute to the discussion of these topics.

Terry's[2] first concern relates to model selection and the fact that we fit models that included a low-density fecundity term and interaction coefficients that varied with rainfall treatment. When fitting models to data from different treatments to assess the coexistence implications of treatment, there are two approaches in the literature. The more common approach, and the one that we used, involves fitting models to data in the different treatments and then propagating the error in the fitted parameters through to the competitive outcomes[8,10,12]. An alternative, advocated by Terry[13] is to first test for treatment effects on the fitted parameters and then make projections that include only treatment-specific parameters when a hypothesis test supports doing so. Terry[2] used Akaike information criterion (AIC) estimates to show that, for some of our study species, simpler models in which rainfall affected only a subset or none of the parameters might fit our data better.

While this is interesting context to provide, AIC support for consistent rainfall effects on all interactions affecting a given species in a multispecies population dynamic model does not address our central question of whether the coexistence of species pairs changes with rainfall. This is because coexistence is determined by ratios of several parameters[1] and, therefore, to answer our question, we need to determine whether rainfall differentially affected the parameter ratios between species that determine coexistence, not the individual parameter estimates themselves. For example, it is the fact that intraspecific and interspecific interaction coefficients differentially respond to rainfall that causes the stabilizing niche differences to change. However, none of Terry's[2] three alternative models allow rainfall treatment to differentially affect the competition coefficients experienced by a species, and they therefore cannot accurately predict changes in coexistence outcomes. Treatment does not need to have a significant effect on a species' overall sensitivity to neighbours (Terry's α terms)[2] to alter the ratios of the key interaction coefficients and drive coexistence changes. Rather, only some of the competition coefficients need to be affected by rainfall, and such a response may not generate a lower AIC when comparing models in which all parameters can vary with treatment or none can vary. To most accurately follow Terry's[2] suggestion in the context of our stated goals, one would have to allow all combinations of the seven estimated parameters (six interaction coefficients and one lambda) to vary by treatment or not, resulting in 128 different potential models for each focal species.

However, even if that approach was taken, we disagree with Terry[2] that using a simpler model with common parameter estimates across treatments for some of the species but not others, as his AIC results would suggest, generates a more biologically accurate or defensible prediction. His assessment would argue, for example, that when modelling the interaction between *Hordeum murinum* and *Plantago erecta*, estimated parameters should be rainfall dependent when *P. erecta* is the focal receiving competition from *H. murinum*, but rainfall independent when *H. murinum* is the focal receiving competition from *P. erecta*. This inconsistency is problematic, especially if we make the sensible assumption that the plants are competing for the same water. We argue (as Terry[2] also suggests) that the issue is best addressed by using the more-complete model and propagating uncertainty in the fitted parameters through to the predicted outcomes, as we did in our original analysis with nonparametric bootstrapping[1] (Extended Data Fig. 1). This reflects the fact that some interaction coefficients change with rainfall for some species and not for others, while still providing the information necessary to predict coexistence outcomes and the uncertainty in those predictions.

Armitage[3] raises separate concerns about model selection—what he sees as insufficient justification for the use of the Beverton–Holt (BH) competition model. He asserts that more complex competition models better fit our data, and that these models generate different coexistence outcomes. Armitage[3] argues that "there is no a priori basis to strongly favour the BH model over similar alternatives". In fact, this is not correct. A 2009 study of coexistence in this same annual plant study system[4] assessed all seven of the competition models[14] that Armitage[3] suggests and found that the BH model had the lowest AIC. Moreover, in a later paper in the same system[15], BH parameter fits were used to generate invasion growth rates for nine competitive pairings. The invasion growth rates from this study perfectly lined up with the predicted competitive outcomes, including one case of coexistence, from a purely empirical approach based on a replacement series design[15]. A more recent study assessed several models of annual-plant population dynamics and found that the BH model is the phenomenological model that most consistently aligned with his preferred mechanistic generative model[16]. Thus, the BH model has been supported by past model selection efforts and generated predictions that were empirically validated to the extent possible.

This is not to discount Armitage's[3] result that the Watanabe–Akaike information criteria (WAIC) for his model 7—the exponentiated Beverton–Holt (EBH)—is lower than for the simple BH model on which our results are based. However, WAIC is just one of multiple

[1]Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA, USA. [2]Forestry and Rangeland Stewardship Department, Colorado State University, Fort Collins, CO, USA. [3]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. ✉e-mail: mary.vandyke@colostate.edu

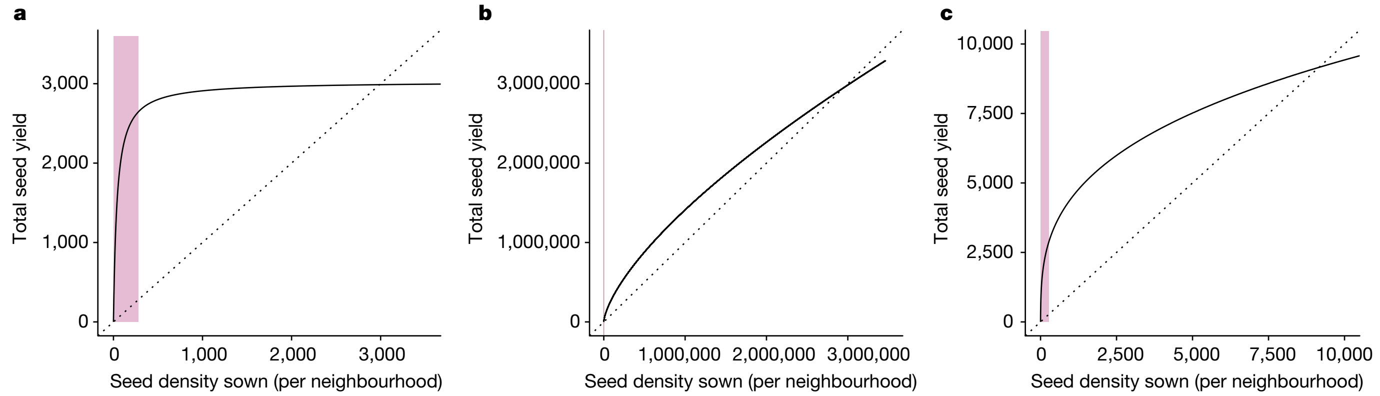


Fig. 1 | Density yield plots for *Uropappus lindleyi* monocultures in the reduced rain treatment, based on three fitted models. **a**, The relationship expected when using the BH model[1]. **b**, The relationship expected when using EBH model (with unconstrained θ). Note the 1,000-fold difference in scales on both of the axes[3]. **c**, The relationship expected when using Armitage's[3] preferred EBH with the constrained $\theta > 0.66$. The shaded pink area indicates the range of neighbourhood densities, which were sown in our experiment (it is almost invisible in **b**). The expected equilibrium density is found where the 1:1 dotted line crosses the yield curve, corresponding to the density at which seed production equals the seeds sown.

considerations that should go into model-selection decisions. For example, candidate models should reproduce observed behaviours of the system of interest and accurately reflect the way that system operates[17–19]. In this context, there are first principles, as well as theoretical and empirical reasons to believe that predictions from the EBH model are less trustworthy than those from the BH. The following arguments focus on the EBH model but are also true for Armitage's[3] other two models that have lower WAIC than the BH (Armitage's models 5 and 6).

The BH model was derived for annual plant populations previously[20], where it was recognized that a population dynamic model could be expressed that reproduces the 'law of constant yield'. The law of constant yield is the canonical observation[21] that, as the density of sown plants increases, so too does the total yield (biomass or seeds), but the yield eventually levels off such that further increases in sowing density generate no increase in biomass or total seed production (Fig. 1a). This 'law' also makes mechanistic sense because the pool of soil resources to make plant material and seeds is finite. While the simple BH model always follows the law of constant yield due to its derivation, the EBH (and Armitage's[3] suggested models 5 and 6) does not (Supplementary Information 1). When the exponential in the denominator (θ) is less than 1, there is no limit to the amount of seeds produced as sowing density increases (Fig. 1b). Critically, if one conducts a classic model comparison with our data and lets the fitting algorithm determine the best-fit parameters, the exponential in the denominator is always less than 1 (Extended Data Table 1), meaning a violation of the law of constant yield (Supplementary Information 1). In what follows, we present the problems arising from following a classic model-selection approach, which leads to the adoption of an unconstrained EBH, and then show that the problems remain even after arbitrarily constraining the EBH as Armitage[3] did.

Violating the law of constant yield by selecting the EBH implies inexhaustible resources, which is problematic from a first-principles standpoint. Moreover, doing so also generates unrealistic coexistence predictions. This is because these predictions, and all modern coexistence theory metrics, are derived from the invasion condition—the state in which one species is invading the other at its single-species carrying capacity. One rarely has enough seed to assess the single-species carrying capacities directly. Instead, one needs to extrapolate beyond the range of sown species densities to equilibrium conditions. While the EBH does fit our data better than the BH within the range of competitor densities sown, because it never saturates, it makes wildly unrealistic predictions when extrapolated to the density at which species reach equilibrium.

The single-species equilibrium for a species can be seen on the density–yield plots of Fig. 1 by identifying the intersection of the yield curve with the 1:1 line (the density at which seed production equals the seed density sown). For *Uropappus*, under reduced rain, our BH model predicts an equilibrium density of 2,988 seeds per competitive neighbourhood (a circle with a 10 cm radius; Table 1). Given that a *Uropappus* seed is approximately a 1-mm-diameter by 10-mm-long cylinder, this corresponds to almost a monolayer of seed if evenly spread, a high but not wholly unrealistic scenario for an equilibrium density. By contrast, the fitted EBH predicts an equilibrium density of approximately 2,949,299 seeds per competitive neighbourhood (Table 1), an increase of three orders of magnitude over the BH. To put this massive equilibrium density in perspective, 2,949,299 *Uropappus* seeds per competitive neighbourhood corresponds to over 1,000 stacked monolayers of *Uropappus* seed, which, if they could be stacked within the area, would correspond to a column of seed over 1 m in height above the surface of the soil. Not only is this wildly unrealistic, so too would be the invasion growth rate of a competitor introduced at low density into this 1 m high pile of seeds. Note also that, although both model forms require an uncomfortable degree of extrapolation from the sown densities to the equilibrium value (compare the purple shaded regions to equilibrium point in Fig. 1), it is egregious with the EBH (Fig. 1b; see Extended Data Fig. 2 for all other species). In summary, while the EBH model does offer marginally better fit than the BH model over the range of the data for some of the species, it is unsuitable for extrapolating to the equilibrium densities required for coexistence analyses.

To avoid the problematic carrying capacities produced by the EBH[3,22], Armitage arbitrarily constrains θ to be greater than 0.66, essentially constraining his EBH to be closer to the BH (where $\theta = 1$). If he did not constrain θ, as would be the natural model-selection approach (and matched his original critique[22]), 5 out of the 12 estimated thetas would be below 0.66 (Extended Data Table 1). Importantly, there are no first-principles reasons to constrain the fit of θ to be greater than 0.66, as that value still allows unbounded seed production with increasing sowing density (Fig. 1c), and still requires particularly egregious extrapolation from the data to get carrying capacities (Fig. 1c). In brief, the problems with the unconstrained EBH remain with Armitage's[3] constrained EBH. In practice, the constraint on θ reflects the investigator's choice to weigh two considerations—low WAIC versus realism. While the unconstrained EBH has the lowest WAIC, the BH follows from first principles, was validated in past studies in the system, and its carrying capacities and invasion growth rates require the least extrapolation. Depending on how the investigator wishes to weigh WAIC versus realism, they can let θ be unconstrained at one extreme, or constrained to

Table 1 | Equilibrium single-species density predictions

Species	Ambient rainfall			Reduced rain		
	BH[a]	EBH[b]	EBH[b] constrained	BH[a]	EBH[b]	EBH[b] constrained
ACWR	8,055	35,414	31,984	4,766	21,531	21,215
FEMI	3,619	9,239	9,197	2,665	17,209	11,835
HOMU	776	1,721	1,657	1,481	19,906	9,913
PLER	1,530	3,450	3,437	645	731	718
SACO	6,722	54,592	37,650	3,962	60,097	22,757
URLI	6,166	1,415,381	13,761	2,988	2,949,299	9,159

Equilibrium single-species density predictions (seeds per competitive neighbourhood, 10-cm-radius circle) for each species under each rainfall treatment using median parameters estimated from the BH model[1], the EBH[3] and the EBH[3] with constrained $\theta > 0.66$. Species' acronyms are the same as in our original paper[1]: ACWR, *Acmispon wrangelianus*; FEMI, *Festuca microstachys*; HOMU, *H. murinum*; PLER, *P. erecta*; SACO, *Salvia columbariae*; URLI, *U. lindleyi*.

[a] BH: equilibrium seeds of species *i* per competitive neighbourhood = $(\lambda_i - 1)/\alpha_{ii}$.

[b] EBH: equilibrium seeds of species *i* per competitive neighbourhood = $(\lambda_i^{1/\theta_i} - 1)/\alpha_{ii}$.

1 at the other. Any choice in between, and especially one made after seeing model predictions, reflects the investigator's arbitrary choice.

Instead of forcing parameters to desired values post hoc when dissatisfied with the predicted dynamics, a better solution is to be consistent for all species and use the model that more accurately reflects the dynamics of the system, as we did in our original paper[1]. In this context and given that an inappropriate choice of model for extrapolation will naturally generate different predictions than a more realistic one, we are unsurprised by the different coexistence outcomes and modern coexistence theory metrics derived from Armitage's[3] EBH model versus our own (the same goes for the trait correlations with the EBH-based metrics[3]).

In addition to both author's model selection concerns, Terry[2] questions our finding that reducing rainfall changed the log-ratio of the competition coefficients to a greater degree than it changed the log-ratio of the demographic potentials. These log-ratios are the two determinants of the invasion growth rate, which determines species coexistence (see Fig. 3 of the original paper[1]). To our knowledge, this is the first use of this type of analysis in the literature, and getting it right is important for advancing studies that decompose treatment effects on species coexistence. Terry[2] correctly shows that, if there is more uncertainty in the estimates of competition coefficients than in the estimates of demographic potentials, this pattern of unequal uncertainty alone could cause the log-ratio of competition coefficients to change more than the log-ratio of the demographic potentials, all else being equal.

Although Terry's[2] concern about significance testing given this phenomenon is important (for example, what is the correct null hypothesis?), the key question is whether we observed rainfall treatment effects over and above the effects of differences in uncertainty. To answer this, Terry[2] assumes that the average difference between the ratios in his simulated data (with no treatment effect) is solely a measure of the effect of differences in uncertainty and can therefore be subtracted from the observed ratios of these parameters to better estimate the true effect of rainfall. However, Terry's[2] 'corrected' *t*-test is inconsistent in how it propagates uncertainty in his simulated data versus the real data (Supplementary Information 2). In fact, even if we accept Terry's[2] subtraction method and use it to correct the mean changes in parameter ratios from the bootstrap samples in a manner that is consistent with his null model approach, the analysis results in a significant *t*-test. The log-ratios of the competition coefficients change more between treatments than the log-ratios of the demographic potentials ($P = 0.054$; Supplementary Information 2). Regardless of this result, we believe that there is a better way to compare across treatment the changes in the ratios of parameters that have different uncertainties, and thank Terry[2] for motivating this follow-up.

A conventional alternative to Terry's[2] subtraction approach involves generating traditional standard effect size (SES) estimates from classic ecological null model approaches. These estimates enable us to test the hypothesis that our observed changes in the ratios of the demographic and competitive parameters with rainfall (Fig. 3 of the original paper[1]) are different from what would be expected assuming no treatment effect (based on Terry's[2] randomizations). When using the mean estimates from the 1,000 bootstraps of our original data, we find that the changes in both the log-ratios of demographic potentials and log-ratios of competition coefficients (Fig. 2a,b) are larger than expected based on the differential uncertainty alone (SES = 2.66, $P < 0.05$, SES = 3.77, $P < 0.05$, respectively). Importantly, the effect of rainfall treatment on the log-ratios of competition coefficients is almost one and a half times its effect on the log-ratios of demographic potentials (3.77 versus 2.66), supporting our original claim. Moreover, we find that the observed mean difference between the treatment-induced change in the log-ratio of competition coefficients and the change in log-ratio of demographic potentials is greater than expected (at $P < 0.05$) based on Terry's[2] simulated data with no treatment effect (Fig. 2c and Supplementary Information 2; Terry obtains $P = 0.095$ for this difference using a slightly different method). Together, these analyses are now more robust and strongly support our original assertion that reduced rainfall had greater effects on the log-ratio of competition coefficients than on the log-ratio of demographic potentials. Still, both changes probably contributed to the changes in coexistence outcomes that we found—a valuable addition to our original analyses.

Finally, both Terry[2] and Armitage[3] assert that we did not use appropriate statistical rigour when drawing conclusions about changes in competitive interactions in our experiment. They argue that using the median value of stabilizing niche and fitness differences from the bootstrapped analyses to determine coexistence somehow misrepresents our experimental results because it does not account for the uncertainty in the coexistence outcomes. We completely agree that presenting uncertainty in these outcomes is important, hence our decision to carry uncertainty through to our calculations of stabilizing niche and fitness differences as presented in Extended Data Fig. 1 of our original paper[1], analogous to Fig. 1 of Armitage's paper[3] and Extended Data Fig. 2 of Terry's paper[2]. Both authors suggest propagating error through the nonparametric bootstrap samples, which is in fact what we did to create the error bars in Extended Data Fig. 1 of our original paper[1] (also shown here in Extended Data Fig. 1). Yet, both responses suggest that pairs for which coexistence changed with treatment in >50% of the bootstrapped outcomes did not change enough to constitute a shift in coexistence outcome. This critique raises the interesting philosophical question for the field to grapple with of how much change in coexistence outcome is substantial (as in the title of our original paper) or how much ecologists find biologically meaningful.

The answer is not clear. Consider the case in which a pair lies exactly on the border of the coexistence region, and therefore has a 50% chance of coexistence with any uncertainty in the underlying parameters. If we chose, for example, a critical threshold of 95% change in coexistence outcome to be significant (or even 51% change), no force in nature can significantly affect the coexistence of the pair, which seems unreasonable. Clearly, we need to establish conventions for this in the field. Returning to our data, we have plotted the probability of coexistence change with treatment for each pair based on our bootstrapped estimates of the competitive outcome (Fig. 3 and Extended Data Fig. 1). Each of the ten pairs for which the median coexistence outcome changed with treatment in our original paper[1] has a minimum probability of competitive outcome change with treatment of 50%. What should be the correct threshold? One option is to use precedent. A recent paper found an approximately 40% change in coexistence outcome, and treated that shift as a substantial effect of the tested factor[11]. If we use 40% change with treatment as a threshold here, then 12 pairs showed meaningful effects of rainfall treatment on coexistence, more than

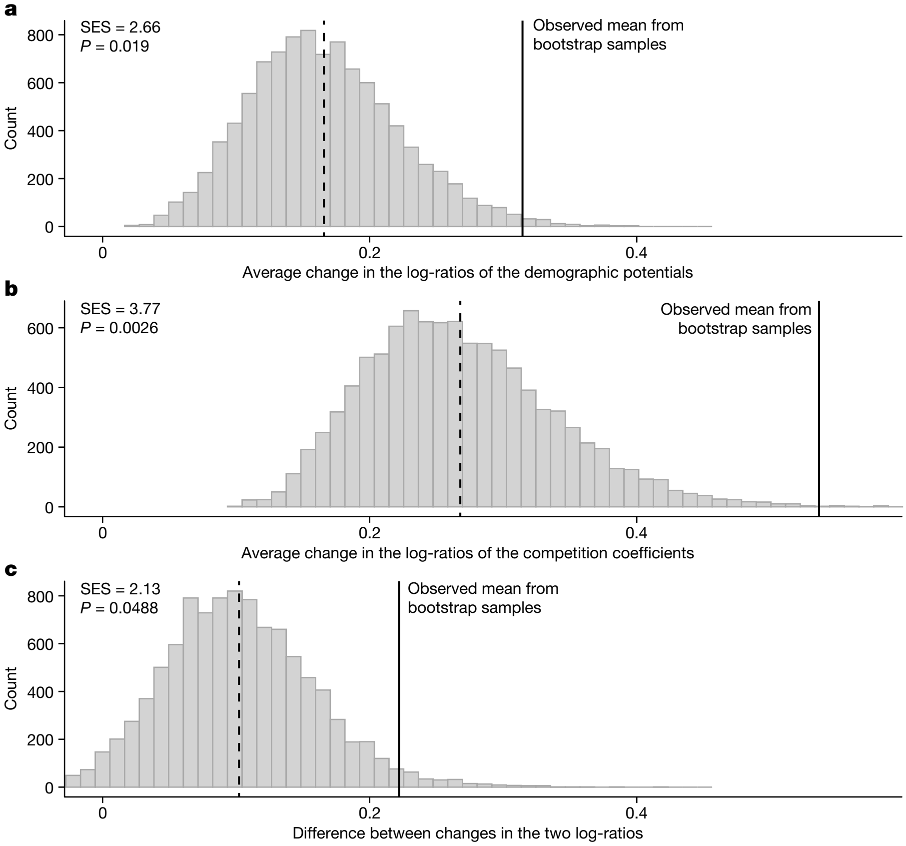


Fig. 2 | The distributions of average changes in the determinants of invasion growth rates from the null model draws compared with the mean from bootstraps of the real data. **a**–**c**, The distributions of the average changes in the two determinants of invasion growth rates for 15 species pairs based on 10,000 draws from Terry's[2] simulated datasets with no true treatment effects (grey) and the average change from 1,000 bootstrap samples of the real data (solid black line). The dashed lines show the means generated from the simulated data. **a**, The distribution of average changes in the log-ratios of the demographic potentials. **b**, The distribution of average changes in the log-ratios of the competition coefficients. **c**, The distribution of the average difference between the two log-ratio changes for each simulated dataset. Assuming that the mean of the changes in the log-ratios generated from the bootstrap samples of our real data represents the observed change with rainfall treatment and the simulated data represent the null distribution, we can calculate SES values and two-tailed *P* values. We find that the observed mean is significantly greater than expected for all three panels.

the 10 pairs suggested by the bootstrapped medians in our original paper[1] (Fig. 3 and Extended Data Fig. 1). Our general point is not that 40% is the correct threshold, but rather that continued discussion of this issue is required for the field to establish conventions. In this case, identifying these probabilities (Fig. 3) and presenting the results with all the bootstrapped outcomes as Terry[2] and Armitage[3] suggest provides additional useful context (Extended Data Fig. 1).

Terry[2] also questions our conclusions by showing that, with his simulated 'no treatment' data, pairs changed coexistence outcomes on average 22.9% of the draws due to uncertainty in parameters and proximity to the coexistence region. However, to put this expected change in context, we can again examine SES values drawn from ecological null model analyses, using Terry's[2] simulated data as the null expectation. We find that our observed average change of 52% (among all 15 pairs) is highly unlikely to be due to chance alone (Extended Data Fig. 3a; SES = 2.8, $P < 0.05$), indicating that the effects of water treatment on coexistence outcomes far outweigh any effect from species pairs' uncertainty in their stabilizing niche and fitness differences.

In conclusion, we welcome the efforts of both Terry[2] and Armitage[3] to make the field application of modern coexistence theory more robust. However, even in light of their concerns, the BH model still appears to be the best choice that we have for prediction in this system, and the statistical models proposed by Terry[2] are not adequate to test how coexistence metrics change with rainfall. As we stated in our original paper[1], moving towards more mechanistic models of competition will bring us more clarity, and our research groups are working towards that end[23,24]. Terry[2] also raised excellent points about how differences in parameter uncertainty may affect our conclusions, and addressing them has provided a better test of how rainfall affected

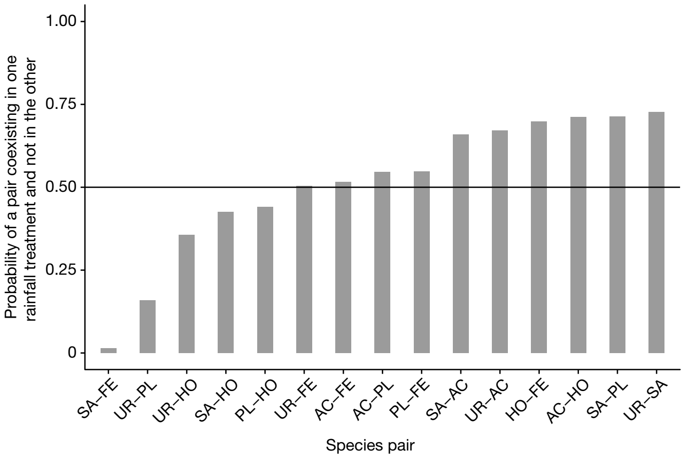


Fig. 3 | The probability of a species pair changing competitive outcomes with rainfall treatment. The probabilities of a changed outcome between treatments are based on outcomes from all bootstrap samples[1]. The solid black line represents the 50% threshold. Species pairs on the *x* axis are indicated by the first two letters of their acronyms as defined in our original paper[1]: AC, ACWR; FE, FEMI; HO, HOMU; PL, PLER; SA, SACO; UR, URLI.

the determinants of invasion growth rates and coexistence. We also agree with both authors that proper error propagation is essential for understanding competitive interactions. Although questions about how to interpret changes in competitive outcomes with uncertainty will need further attention from the field, the results of our original paper remain unchanged. Small changes in rainfall substantially altered coexistence for most species pairs—effects that are mediated by changes in species' interaction strengths. This finding is consequential for understanding how communities and species ranges will be affected by global change.

Reporting summary

Further information on experimental design is available in the Nature Portfolio Reporting Summary linked to this Article.

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/s41586-024-07777-3.

Data availability

Data are all from the original paper[1] and are available at Zenodo (https://doi.org/10.5281/zenodo.7083314)[25].

Code availability

All codes needed for this reply are provided in Supplementary Information 1–3. All code needed for the original paper[1] is available at Zenodo (https://doi.org/10.5281/zenodo.7083314)[25].

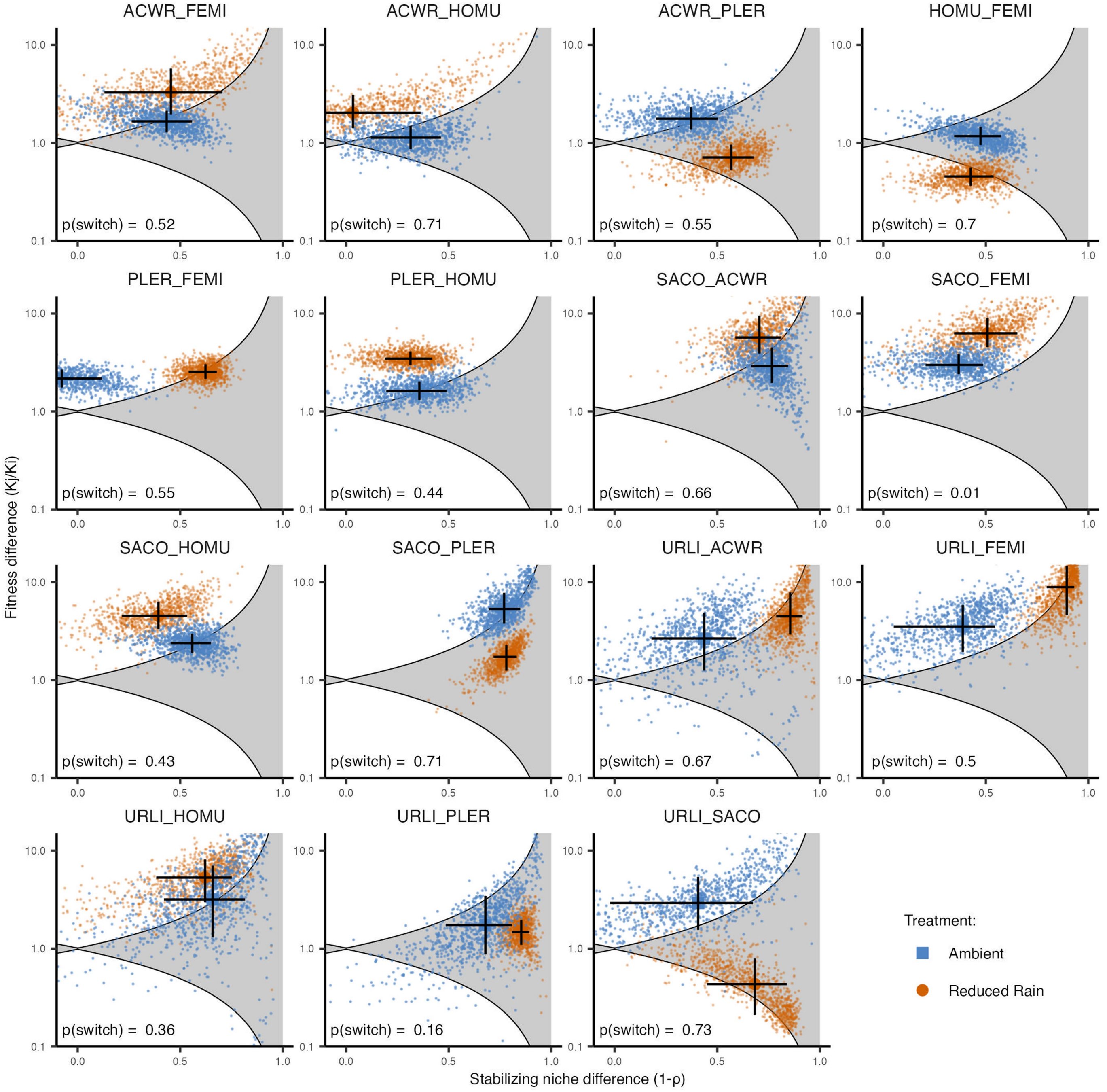
1. Van Dyke, M. N., Levine, J. M. & Kraft, N. J. B. Small rainfall changes drive substantial changes in plant coexistence. *Nature* **611**, 507–511 (2022).
2. Terry, J. C. D. Uncertain competition coefficients undermine inferences about coexistence. *Nature* https://doi.org/10.1038/s41586-023-06859-y (2024).
3. Armitage, D. W. To remain modern, the coexistence program requires modern statistical rigor. *Nature* https://doi.org/10.1038/s41586-023-06919-3 (2024).
4. Levine, J. M. & HilleRisLambers, J. The importance of niches for the maintenance of species diversity. *Nature* **461**, 254–257 (2009).
5. Godoy, O., Kraft, N. J. B. & Levine, J. M. Phylogenetic relatedness and the determinants of competitive outcomes. *Ecol. Lett.* **17**, 836–844 (2014).
6. 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).
7. Germain, R. M., Mayfield, M. M. & Gilbert, B. The 'filtering' metaphor revisited: competition and environment jointly structure invasibility and coexistence. *Biol. Lett.* **14**, 20180460 (2018).
8. Wainwright, C. E., HilleRisLambers, J., Lai, H. R., Loy, X. & Mayfield, M. M. Distinct responses of niche and fitness differences to water availability underlie variable coexistence outcomes in semi-arid annual plant communities. *J. Ecol.* **107**, 293–306 (2019).
9. Kandlikar, G. S., Johnson, C. A., Yan, X., Kraft, N. J. B. & Levine, J. M. Winning and losing with microbes: how microbially mediated fitness differences influence plant diversity. *Ecol. Lett.* **22**, 1178–1191 (2019).
10. Johnson, C. A., Dutt, P. & Levine, J. M. Competition for pollinators destabilizes plant coexistence. *Nature* **607**, 721–725 (2022).
11. Hess, C., Levine, J. M., Turcotte, M. M. & Hart, S. P. Phenotypic plasticity promotes species coexistence. *Nat. Ecol. Evol.* **6**, 1256–1261 (2022).
12. Hart, S. P., Turcotte, M. M. & Levine, J. M. Effects of rapid evolution on species coexistence. *Proc. Natl Acad. Sci. USA* **116**, 2112–2117 (2019).
13. 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).
14. 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).
15. Godoy, O. & Levine, J. M. Phenology effects on invasion success: insights from coupling field experiments to coexistence theory. *Ecology* **95**, 726–736 (2014).
16. Stouffer, D. B. A critical examination of models of annual-plant population dynamics and density-dependent fecundity. *Methods Ecol. Evol.* **13**, 2516–2530 (2022).
17. Klir, G. J. *Architecture of Systems Problem Solving* (Springer, 1985).
18. Zeigler, B. P., Praehofer, H. & Kim, T. *Theory of Modelling and Simulation* (Academic, 2000).
19. Box, G. E. P. & Luceno, A. *Statistical Control: By Monitoring and Feedback Adjustment* (John Wiley & Sons, 1997).
20. Watkinson, A. R. Density-dependence in single-species populations of plants. *J. Theor. Biol.* **83**, 345–357 (1980).
21. Harper, J. L. *Population Biology of Plants* (Academic, London, 1977).
22. Armitage, D. W. To remain modern, the coexistence program requires modern statistical rigor. Preprint at *bioRxiv* https://doi.org/10.1101/2022.12.28.522056 (2022).
23. Levine, J. I., Levine, J. M., Gibbs, T. & Pacala, S. W. Competition for water and species coexistence in phenologically structured annual plant communities. *Ecol. Lett.* **25**, 1110–1125 (2022).
24. Kleinhesselink, A. R., Kraft, N. J. B., Pacala, S. W. & Levine, J. M. Detecting and interpreting higher-order interactions in ecological communities. *Ecol. Lett.* **25**, 1604–1617 (2022).
25. Van Dyke, M. N., Levine, J. M. & Kraft, N. J. B. Data and code for 'Small rainfall changes drive substantial changes in plant coexistence'. *Zenodo* https://doi.org/10.5281/zenodo.7083314 (2022).

Acknowledgements We thank D. W. Armitage and J. C. D. Terry for constructive discussion.

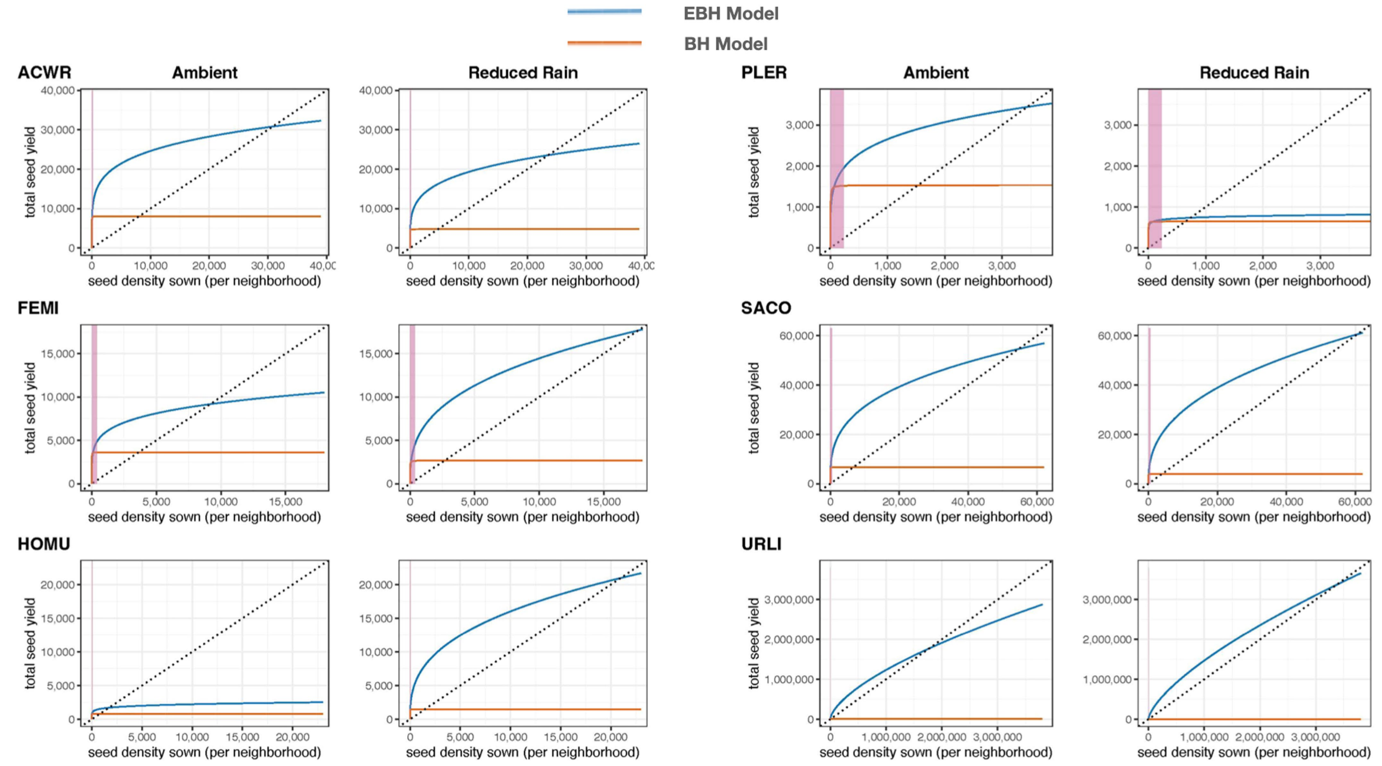
Author contributions M.N.V.D. led the writing with substantial contributions from J.M.L. and N.J.B.K. M.N.V.D. performed the additional analyses and created the figures.

Competing interests The authors declare no competing interests.

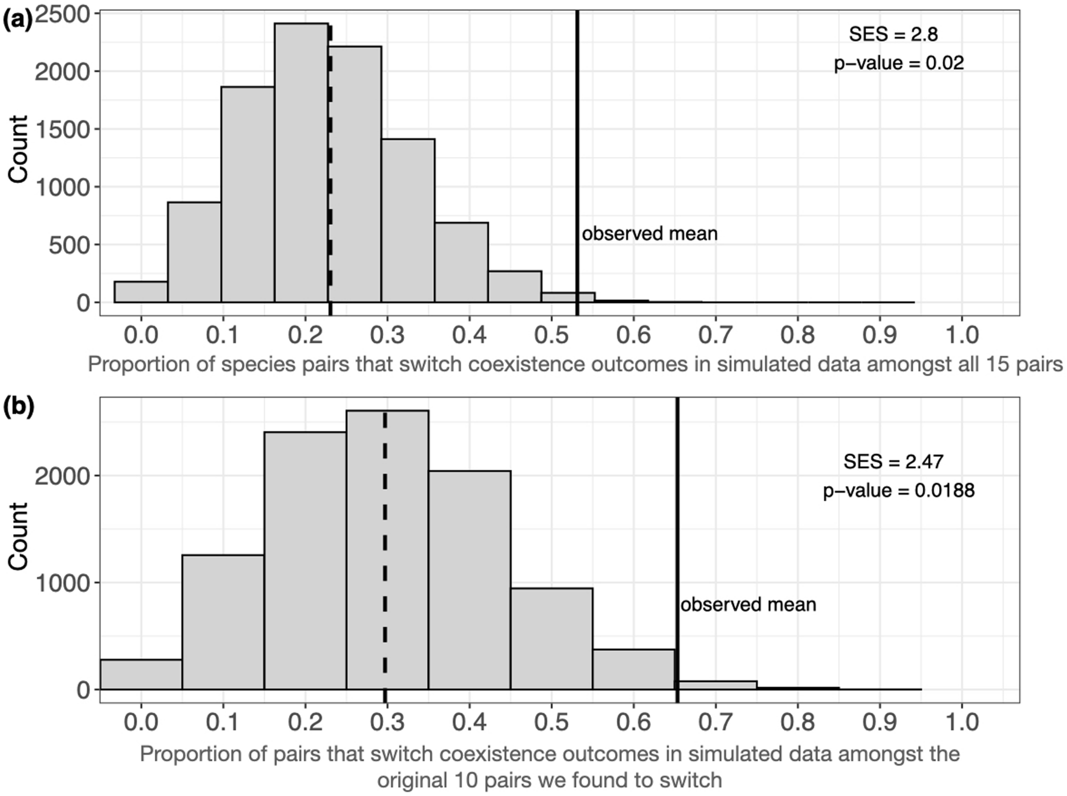
Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07777-3.
Correspondence and requests for materials should be addressed to Mary N. Van Dyke.
Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Predictions of competitive outcomes for the 15 species pairs. Points represent stabilizing niche and fitness differences calculated from non-parametric bootstrap samples. Points lying inside the grey shaded region indicate coexistence, and outside indicate competitive exclusion. Median predictions are where the error bars (+/− 1 SD) cross. p(switch) in each panel reports the probability that pair switched coexistence outcome with treatment.

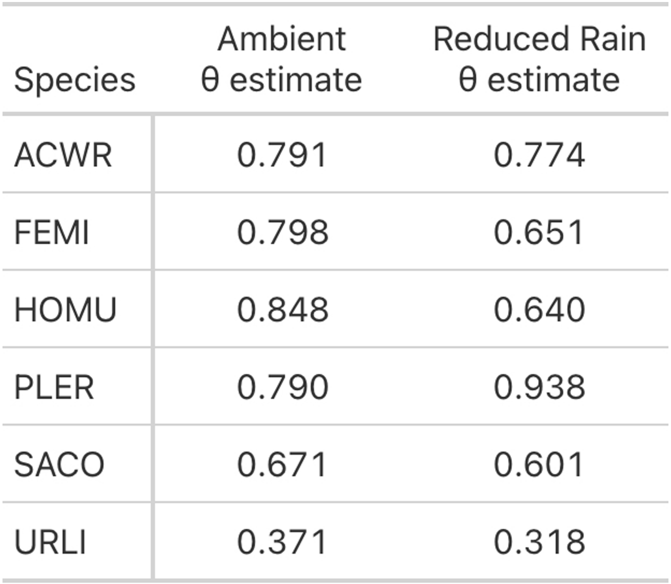


Extended Data Fig. 2 | Density yield plots for each species monocultures under both treatments. The BH model estimates are shown in orange[1] and the EBH model estimates are shown in blue[3]. The shaded pink area indicates the range of neighbourhood densities which were sown in our experiment for each species.



Extended Data Fig. 3 | Distributions of the proportion of pairs that switched coexistence outcomes in the null model draws compared to the mean from the bootstraps of the real data. Panel (a) shows the distribution of the proportion of pairs found to switch coexistence outcomes in Terry's simulated 'no treatment' data and the mean proportion from the bootstrap samples of the real data (solid black line) when we include all 15 pairs, and panel (b) shows the proportion when we only include the ten pairs our original paper found switched coexistence outcomes.

Extended Data Table. 1 | Estimated thetas from the unconstrained EBH model



Species	Ambient θ estimate	Reduced Rain θ estimate
ACWR	0.791	0.774
FEMI	0.798	0.651
HOMU	0.848	0.640
PLER	0.790	0.938
SACO	0.671	0.601
URLI	0.371	0.318

Five of the twelve estimated thetas are below the 0.66 threshold set by Armstrong[3].

nature portfolio

Corresponding author(s): Mary Van Dyke

Last updated by author(s): 2024/04/14

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No new data was used for this analysis. All data from the original paper are available at https://zenodo.org/records/7083314
Data analysis	All code and analysis needed for the original analysis is available at https://zenodo.org/records/7083314. All code needed for the additional analyses in this response are provided in the supplementary materials. All analyses were done in R version 4.2.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data is available at https://zenodo.org/records/7083314.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This paper is a response to a 'matters arising' comment and is not a new study.
Research sample	n/a
Sampling strategy	n/a
Data collection	n/a
Timing and spatial scale	n/a
Data exclusions	n/a
Reproducibility	n/a
Randomization	n/a
Blinding	n/a

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Clinical data
☒	☐	Dual use research of concern
☐	☒	Plants

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	n/a
Novel plant genotypes	N/A
Authentication	N/A