An experimental validation test of ecological coexistence theory to forecast extinction under rising temperatures.

J. Christopher D. Terry[1*]

1: Department of Biology, University of Oxford, 11A Mansfield Road, Oxford, OX1 3SZ

christopher.terry@biology.ox.ac.uk 0000-0002-0626-9938

Keywords: coexistence, climate change, Drosophila, Bayesian, ecological forecasting,

Type: Letter

Number of Words in Abstract: 149

Number of Words: **4998**

Number of References: 70

Number of Figures: 4

Number of Tables: 1

Running title: Testing coexistence theory under warming

Data Accessibility: Raw data and code for all analysis steps is available at github.com/jcdterry/MCT_Fly_Val_public and archived on Zenodo doi.org/10.5281/zenodo.10679556

Abstract

Interactions between species pose considerable challenges for forecasting the response of ecological communities to global changes. Coexistence theory could address this challenge by defining the conditions species can or cannot persist alongside competitors. However, although coexistence theory is increasingly deployed for projections, these frameworks have rarely been subjected to critical multigenerational validation tests. Here, using a highly replicated mesocosm experiment, I directly test if the modern coexistence theory approach can predict time-to-extirpation in the face of rising temperatures within the context of competition from a heat-tolerant species. Competition hastened expiration and the modelled point of coexistence breakdown overlapped with mean observations under both steady temperature increases and with additional environmental stochasticity. That said, although the theory identified the interactive effect between the stressors, predictive precision was low even in this simplified system. Nonetheless, these results support the careful use of coexistence modelling for forecasts and understanding drivers of change.

Introduction

Species are already shifting their distributions across the world as environmental conditions rapidly change (Chen *et al.* 2011; Lenoir *et al.* 2020). Forecasting how the identities of species resident in ecosystems will change is both of profound applied importance (Pecl *et al.* 2017) and a key test of our ecological understanding of community assembly. Beyond direct physiological responses, the potentially large influence of interspecific competition in determining whether a species can persist under a different environment has been demonstrated in both mesocosm (Davis *et al.* 1998; Legault *et al.* 2020) and field studies (Alexander *et al.* 2015; Descombes *et al.* 2020; Freeman *et al.* 2022; Suttle *et al.* 2007). Parsing this complexity of multi-species dynamics to make forecasts is a considerable challenge given the range of possible processes, but theoretical ecology has developed streamlined frameworks defining precise conditions for when species are, or are not, expected to coexist (Armitage & Jones 2020; Barabás *et al.* 2018; Chesson 2000; Ellner *et al.* 2019; Ranjan *et al.* 2024).

Modern coexistence theory focuses on a widely applicable currency: invasion growth rate (Chesson 2000; Grainger *et al.* 2019), the per-capita population growth rate from low densities into a stationary community of established competitors. Assuming no positive frequency-dependence ('Allee effects' Courchamp *et al.* 1999), a positive invasion growth rate is taken to imply that a species is able to persist on the basis that it could always recover from low densities. In a two-species system, when both species fulfil the mathematical criteria (Barabás *et al.* 2018), they are claimed to be able to coexist in a locality. This approach is frequently framed as an assessment of whether average fitness differences between pairs of species that would otherwise drive exclusion can be overcome by stabilising niche differences (Chesson 2000).

There have been increasing efforts to deploy the extensive body of theory that has developed around the determinants of coexistence for more applied purposes (Alexander *et al.* 2016; Aoyama *et al.* 2022; Hallett *et al.* 2023) with dozens of applications to empirical systems (reviewed in Buche *et al.* 2022; Terry & Armitage 2024). As part of this, coexistence theory has been proposed as a tool to understand how environmental change will impact communities in the future, largely based on parametrising models from short-term experiments under different environmental conditions (Adler *et al.* 2012; Alexander *et al.* 2015, 2018; Bimler *et al.* 2018; Mordecai 2013; Van Dyke *et al.* 2022; Wainwright *et al.* 2019). Where key model parameters are found to be environmentally dependent, the future environments in which species can (or cannot) coexist can be inferred. An ultimate hope is to be able to make trait or phylogeny-based predictions (Godoy *et al.* 2014; Kraft *et al.* 2015; Van Dyke *et al.* 2022) of niche and fitness differences in different environments which would potentially unlock the capacity to forecast species responses to climatic changes at scale without the need for individual parameterisation.

Since the initial development of 'modern' coexistence theory, it has been criticised from many angles - from the fundamental framing (Simha *et al.* 2022) through to the simplicity of the representation of the dynamics of interspecific interactions (Abrams 2022) and the challenge of accurately estimating the form of competition (Cervantes-Loreto *et al.* 2023). The underlying mathematical assumptions of standard modern coexistence theory are also widely recognised as being some distance from reality: these include the focus on pairwise dynamics (Chesson 2018; Spaak & Schreiber 2023) and the potential misalignment of the spatial scales of observation and coexistence (Clark *et al.* 2019; Hart *et al.* 2017). To calculate long-term growth rates the standard approach implicitly assumes fixed traits with infinite time and space horizons (Schreiber *et al.* 2023), as well as continuous population counts without the demographic stochasticity that discrete individuals at small population sizes bring (Pande *et al.* 2020). However, the extent to which these issues preclude the applicability of the framework to contemporary challenges remains unclear, as although it is applied to empirical systems with increasing frequency, its predictions have rarely been directly tested with multi-generational experiments. As such, it remains an unsettled question the extent to which this framework is adequate (Getz *et al.* 2018) for ecological questions given natural complexities.

To assess the capacity for modern coexistence theory to forecast the time of extirpations (i.e. local extinctions) under changing environments, I conducted a highly replicated *Drosophila* mesocosm experiment where populations were tracked through discrete generations in small mesocosms (Figure 1). The controlled environment isolates several of the potentially troublesome simplifications by limiting the diversity and defining the spatial scale. This directly exposes to testing the assumptions of competition model sufficiency, infinite time and space horizons, as well as the potential impact of positive density dependence and adaptation.

Focal species are drawn from a well-characterised community (Hangartner *et al.* 2015; Jeffs *et al.* 2021; Kellermann *et al.* 2012; Terry *et al.* 2021) from a montane Australian tropical rainforest that shows pronounced species turnover over an elevational gradient. Previous empirical work using this community has demonstrated roles for both temperature and competition in determining species range boundaries (Chen & Lewis 2024; O'Brien *et al.* 2017). I focussed on predicting the survival of populations of *Drosophila pallidifrons*, a highland-distributed species with a comparatively cool thermal optimum, in the context of competition from a lowland-distributed species *Drosophila pandora* with a comparatively warm thermal optimum.

Methods

Species

The two focal species were *Drosophila pallidifrons* (*nasuta* species complex) and *Drosophila pandora* (*ananassae* subgroup, McEvey & Schiffer 2015), both drawn from a well characterised Australian tropical forest community that shows distinct elevational community turnover (Jeffs *et al.*

2021). Laboratory populations had been maintained at Oxford University since 2018, originally drawn from multiple isofemale lines sampled from rainforest sites in northern Queensland, Australia (permit PWS2016-AU-002018, Australian Department of Energy and Environment). These cultures had been maintained for several years at large population sizes in a 25°C controlled temperature room. While some laboratory adaptation over this period cannot not be ruled out, preceding work using these populations had demonstrated that species had maintained distinct thermal physiologies (Chen & Lewis 2024).

Experimental Procedures

Each generation was initiated by transferring the founder flies into a standard 25 mm diameter *Drosophila* vial with 5 ml of cornflour–sugar-yeast–agar *Drosophila* medium. The founders were allowed approximately 48 hours to lay eggs before being removed and frozen for later censusing. After 10 days incubation any emerged flies became the founders of the next generation in that line, giving a total tip-to-tip generation time of 12 days. This length was chosen to match the development time across the temperatures used in the experiment. During censusing, all individuals were identified by species, sexed and counted under a stereo microscope. Flies that had clearly died before freezing were not counted. Each generation 1 (G1) vial was initiated with 3 female and 2 male *D. pallidifrons* counted under light CO_2 anaesthesia that had been raised in 180 ml Drosophila flasks under moderate density at 24°C. Populations were censused each generation and the experiment was run for 10 discrete generations (at which point 98.75% of *D. pallidifrons* populations were extinct). Sixty replicates were run for each of two crossed treatment combinations: monoculture or the intermittent introduction of small numbers of *D. pandora*, (Figure 1a), and steady temperature rise or with additional generational-scale thermal variability (either +1.5°C, -1.5°C or 0, with equal chance, Figure 1B).

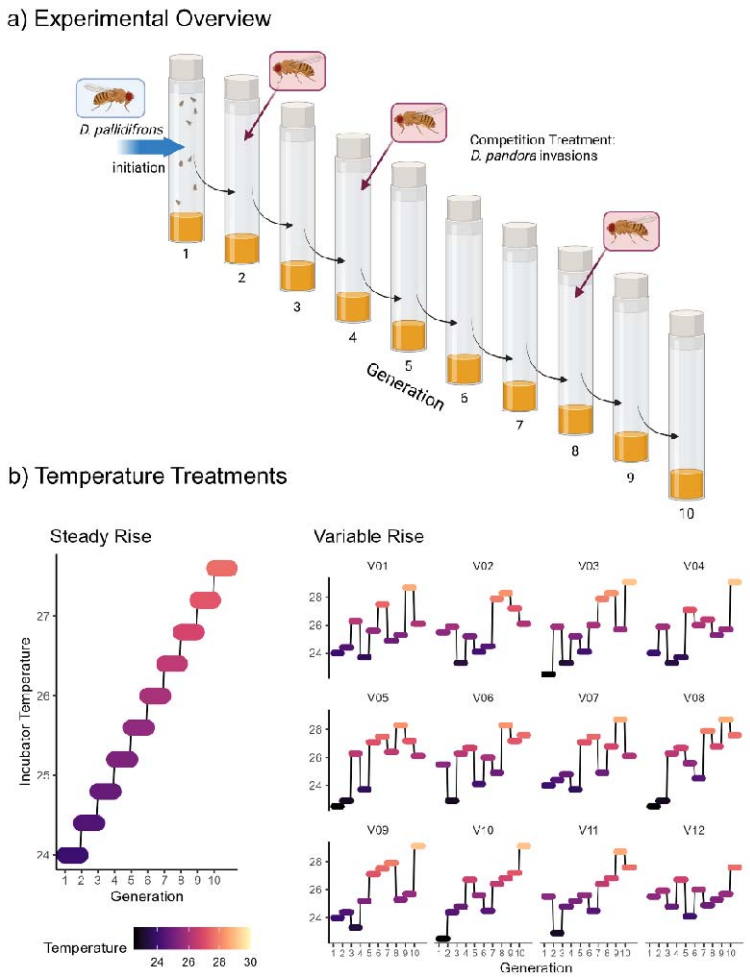

Figure 1. Outline of experiment and time-to-extinction results a) Overall experiment schematic. After initiation of population vials with *D. pallidifrons*, the flies were allowed two days to lay eggs before removal and censusing. After a further 10 days, all emerged flies were transferred to a new vial for two days to start the next discrete generation. In the 'competition' treatment, small numbers of *D. pandora* were added alongside the founders of generation 2, 4 and 8. b) Temperature treatments. In the 'steady rise' scenario, the temperature increases 0.4°C at the start of each generation, while in the variable rise treatment, the temperature imposed on each population follows one of 12 trajectories (V01 – V12) that superimpose variability on top of the steady rise.

In the 'Steady' rise temperature treatment, G1 was founded at a constant temperature of 24°C and the temperature was raised by a 0.4°C step at the start of each generation, implying a 4°C rise across the experiment. This was expected to shift the competitive balance from *D. pallidifrons* to *D. pandora*. For the 'Variable' rise treatment, 12 different variable temperature trajectories (Figure 1b) were generated by randomly assigning a third of the trajectories to each of lowered (-1.5), mean (=) or raised (+1.5) temperature conditions for each generation. Each trajectory had 5 replicates within each treatment. Temperature values were chosen to span the expected shift in coexistence, while also allowing the use of a fixed generation duration across the experiment (since colder temperatures delay development significantly). Experimental populations were maintained in three incubators of two

equivalent models (Sanyo MIR-154 & Sanyo MIR-153) under a 12-12hr light-dark cycle. Temperature and humidity loggers indicated these incubators were able to successfully maintain a controlled environment throughout the experiment.

In the competition biotic treatment, *D. pandora* were intermittently added alongside the founders from the preceding generation. At the start of G2, approximately 15 *D. pandora* were added to each treatment vial. At the start of G4 and G8 4 female and 2 male *D. pandora* were added to all lines that still had a *D. pallidifrons* population. These additions were subtracted from the subsequent census to infer the reproductive rate in the previous generation.

Population Data

In total 37,826 *D. pallidifrons* and 20,786 *D. pandora* flies were counted from 1,045 censuses with non-zero counts (Figure S1). Population dynamics were based just on female counts to better account for the reproductive capacity of small populations and since the sex ratio of both species skewed towards females at higher densities (Figure S2, likely due to slower emergence of males and increased male mortality). Across all temperatures, there were 1,223 *D. pallidifrons* and 592 *D. pandora* observed transitions from non-zero female populations. Because there were so few (6) *D. pallidifrons* data points from temperatures above 28°C, these were excluded to leave a total data set of 1,217 transitions. This allowed the lower bound of T_{max} (see below) to be reduced without introducing numerical errors.

Population Model Fitting

Models representing the generation-generation population dynamics of the two species were built following recently established best practices (Terry & Armitage 2024), including propagation of uncertainty in model parameters (Bowler *et al.* 2022) and consideration of alternative model forms (Cervantes-Loreto *et al.* 2023). A suite of phenomenological models were tested that systematically varied the model form used to represent the fundamental thermal performance curve, the reduction in fecundity caused by competition and the error distribution. These models were fit and compared separately for each species.

All models were fit in a Bayesian framework using the *brms* version 2.19 (Bürkner 2018) R interface to STAN version 2.26.1(Carpenter *et al.* 2017). Priors on the core growth rate function parameters were chosen to be weakly informative and bounded only to prevent numerical errors. Priors on the competition model parameters were based on existing expectations of the strength of competition (Chen & Lewis 2024; Terry *et al.* 2021) and to enforce competition rather than positive interactions. For each model, 4 chains were each run for 4000 total iterations (2000 used as a warm up) on standard settings. Model code in R (*brms*) and STAN is available in the code supplement. Runs were checked for convergence through the Gelman diagnostic $\hat{r}$. Posterior distributions and predictive plots are

available in the code supplement. Model performance was compared using differences in ELPD-LOO (Vehtari *et al.* 2017), which assesses each model on its capacity to make predictions on unseen further data across the whole predictive distribution. Where models could not be clearly distinguished on the basis of predictive performance (with an ELPD difference significantly greater than 4) the simpler model was used.

Each model followed the overall structure:

$$N_{i,g+1} \sim N_{i,g} \times B(T_g) \times C(N_{i,g}, N_{j,g},\ T_g)$$

where N is the number of individual females and T the temperature indexed by the generation g, with i and j indicating focal and competitor species respectively. Based on existing expectations of the shape of thermal performance curves from this *Drosophila* community (Chen & Lewis 2024), four phenomenological asymmetric thermal performance curve models $B(T_t)$ were trialled, selected for having a moderate number of parameters and their distinctiveness (Kontopoulos *et al.* 2023): 'Taylor-Sexton', 'Atkin', 'Simplified β-type' and 'Simplified Brière'. Only the Simplified Brière thermal performance function (Briere *et al.* 1999) was able to consistently converge:

$$B(T_g) = a(T_g - T_{min})\sqrt{T_{max} - T_g}$$

where a, T_{min} and T_{max} are parameters of the thermal performance function defining the low-density growth rate.

Eight phenomenological competition response models (Table S1) were tested for $C(N_{i,t}, N_{j,t},\ T_t)$. These were based on generalisations of the widely used Beverton-Holt model and included temperature-dependent competition coefficients. Linear relationships (i.e. a Lotka-Volterra model) were also tested but converged poorly as they frequently generated negative expectations and were excluded from consideration. For both species, the selected competition kernel was the Beverton-Holt model of density-dependent competition with the competitive pressure exerted by *D. pallidifrons* negatively impacted by temperature:

$$C(N_{PAL,t}, N_{PAN,t},\ T_t) = \frac{1}{1 + (\alpha_{0i,PAL} + \alpha_{Ti,PAL}T_t)N_{PAL,t} + \alpha_{i,PAN}N_{PAN,t}}.$$

where α terms are competition coefficients. The α_{ij} terms represent constant competition coefficients, whereas α_{0ij} is an intercept and α_{Tij} a slope term to linearly link a competition coefficient to temperature. Note that the number of intraspecific competitors of species was taken to be $max(0, N - 1)$, incorporating the assumption that individuals do not compete with themselves. Models including temperature dependent competition consistently performed better than models that did not (Table S3-

4). For *D. pallidifrons*, the highest-ranked model included an additional exponent term but this did not justify the additional model complexity (ΔELPD =6.3, SE = 2.5).

Six error models for count data were tested: Poisson, Negative Binomial, Zero-inflated Negative Binomial, Temperature-dependent-zero-inflated Negative Binomial, Zero-inflated Negative Binomial with competition dependent shape, and Zero-inflated Negative Binomial with temperature dependent shape (Table S2). The differences in error functions had a greater impact on ELPD-LOO than choice of competition model (Table S3-4). Error for *D. pallidifrons* was best represented by a zero-inflated negative binomial model, where the extent of zero inflation was temperature dependent (via a logit link function). For *D. pandora*, error was best represented by a negative binomial function where the shape parameter (ϕ) was competition dependent. See Table S5 and S6 for model parameter values and Figure S4 temperature dependence of emergent parameters.

Predictions of coexistence at different temperatures

The selected population dynamic models were used to forecast the temperature at which *D. pallidifrons* would be extirpated (excluded by *D. pandora*) by three alternative approaches: 1) calculation of mutual invasibility using analytical modern coexistence theory via analysis of niche and fitness differences ('classic' modern coexistence theory), 2) Calculation of long-term growth rates from simulations following the method of Ellner et al. (2019), and 3) Direct replicate simulations of expected extinctions in the specified scenario.

Predictions from analytic coexistence theory

Following the original specification of niche and fitness differences (Chesson 1990; Spaak *et al.* 2023) derived for the Beverton-Holt model in Godoy et al. (2014), I defined the fitness difference between the two species as:

$$\frac{\kappa_i}{\kappa_j} = \frac{\lambda_i}{\lambda_j}\sqrt{\frac{\alpha_{ji}\alpha_{jj}}{\alpha_{ij}\alpha_{ii}}}$$

and the niche overlap as

$$\rho = \sqrt{\frac{\alpha_{ij}\alpha_{ji}}{\alpha_{ii}\alpha_{jj}}}\ .$$

The temperature dependent zero-inflation was incorporated into the overall growth rate term λ by scaling the mean expectation of the number of offspring by the zero inflation: $\lambda(T_t) = B(T_t) \times \left(1 - inv.logit(z_0 + z_T T_t)\right)$. Predicted coexistence at a steady state of each temperature was inferred based on whether the key inequality was fulfilled (Chesson 2000): $\rho < \frac{\kappa_i}{\kappa_j} < \frac{1}{\rho}$. The full posterior parameter estimates of the best performing model was propagated through to give relative support for

each of the four coexistence outcomes (exclusion of either species, coexistence or priority effects) at each temperature.

Prediction of coexistence through calculation of long-term-growth rates in simulations.

Predictions of the mean temperature at which *D. pallidifrons* is excluded under either fixed or stochastic temperature regimes used the flexible simulation-based method of Ellner *et al.* (2019). This makes asymptotic estimates of the long-term invasion (low-density) population growth rate (IGR) by averaging over long synthetic time series of population dynamics. For each mean temperature, a series of 1000 temperatures were generated: either fixed or with fluctuations (+1.5, 0 or -1.5) following the two experimental scenarios.

The population density of the resident *D. pandora* in monoculture was found from simulation using the best-fit population model under both static and stochastic temperatures regimes (discarding the first 50 steps as a burn-in). The population growth rates from rare of *D. pallidifrons* were calculated for each of these sequences of competitors and temperatures, and the long-term average calculated by the geometric mean. Where for some parameter draw combinations the modelled growth rate was either incalculable ($T > T_{max}$) or below 0.1, a surrogate value of 0.1 was used to facilitate the averaging. These steps were repeated across 1000 draws of the posterior distribution of the model fits and at each central temperature in the range 24-27.6°C in steps of 0.2°C.

Subsequently, to partition out how the environmental fluctuations and competition act together to influence the IGR, further simulation scenarios were explored with and without various processes to isolate the individual and interactive effects. The IGR calculated for the full model was partitioned into five components ($r_0 + \Delta_C + \Delta_\sigma + \Delta_* + \Delta_{cov}$) by calculating the differences between models that switched off various processes. To rule out 'storage-effects' emerging from covariance between the environmental and competition (Δ_{cov}), the temperature time series was reshuffled (Model D). To isolate the effect of competition (Δ_C), fluctuations (Δ_σ) and their interaction (Δ_*), the IGR was calculated with neither competition or fluctuations (Model A), and in competition without environmental fluctuations impacting either species (Model B), as a monoculture with fluctuations (Model C). Table S7 lists the constituent partitions of each model. Note that to cap the total number of partitions, Δ_* here incorporates (but is not exactly equivalent to) terms labelled 'relative-non-linearity' in similar analyses (e.g. Letten *et al.* 2018).

Expected time of extinction through direct simulation.

To generate a baseline, non-theory derived, forecast for the extinction of *D. pallidifrons*, the selected population models were used in direct simulations of the population dynamics in all four experimental scenarios (with and without competition and environmental variability). Parameter uncertainty was propagated by running the simulations using 1000 distinct posterior draws from both species'

population dynamics models. To assess the role of how stochasticity is modelled, this test was repeated using a model with the same competition kernel, but fit with a Poisson error term.

Results

Impact of treatments on time to extinction

Treatment combination had a significant overall effect on the generation at which extirpation was observed (Figure 2, ANOVA: = 8.152, p = 0.0003) but only the slower 'steady rise monoculture' treatment was statistically distinct (Tukey HSD test: p< 0.005 for all comparisons with this treatment, p> 0.7 for all others) with a mean expiration 1.73 (SE: 0.358) generations later. When viewed as concurrent stressors, environmental stochasticity and competition had antagonistic interactive effects (Orr *et al.* 2020), in the sense that their combined effect was less than the sum of their individual impacts. The particular trajectory of temperature variability had a significant impact on time-to-extirpation (ANOVA model comparison of Extirpation~Treatment with Extirpation~Treatment +(1|Trajectory): = 6.4029, p = 0.011, n = 120, 12 groups). Trajectories with a lower temperature in the first generation were correlated with earlier expiration, but across the 12 trajectories this was not statistically significant (post-hoc linear regression, p = 0.145, =2.498, adjusted $R^2 = 0.12$).

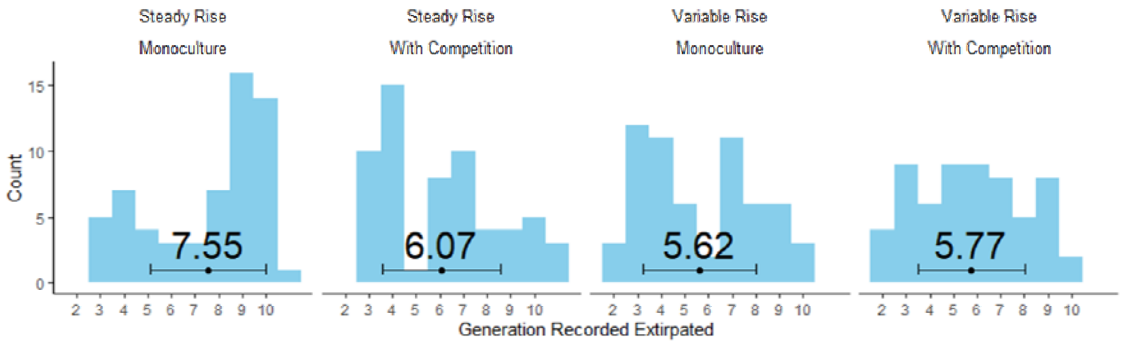

Figure 2 Distribution of observed extirpation points across experimental treatments. Extirpation defined as the first generation where zero *D. pallidifrons* individual were censused. Values shown are means and error bars show +/- 1 SD. n=60 per treatment combination (240 total).

Prediction of time-to-extinction through coexistence theory

With increased temperature, the selected model predicted a shift from *D. pallidifrons* (the resident, highland species) excluding *D. pandora* (the lowland, invader species), through the zone of likely coexistence, to the exclusion of *D. pallidifrons* by *D. pandora* (Figure 3a). Using mean parameter estimates, the threshold for predicted exclusion of *D. pallidifrons* is crossed at approximately 26.7°C, although inspecting the full posterior shows there is support for predicted exclusion somewhat earlier (Figure 3b, red area). Converting these temperatures to generations in the experiment this threshold is

first breached in generation 8, just under two generations (corresponding to a 0.8°C difference) after the mean generation of observed extirpation (6.07).

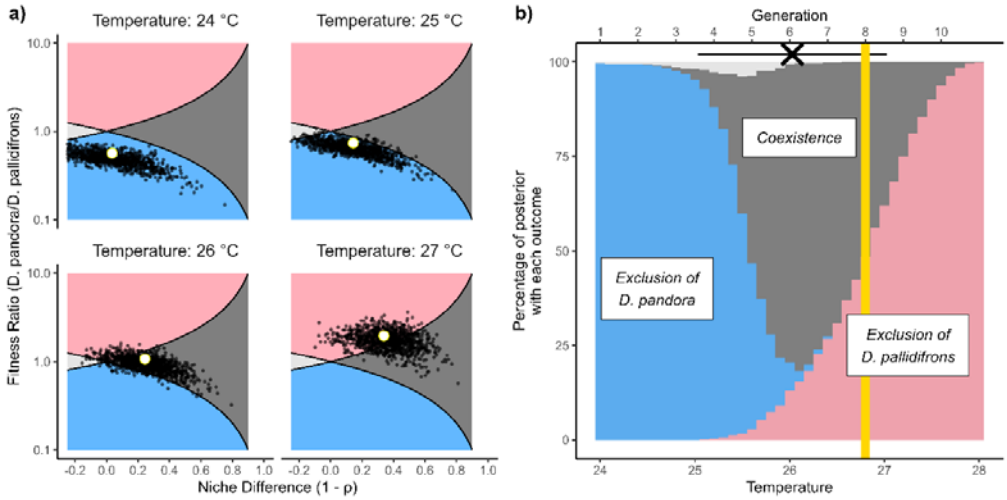

Figure 3: Predictions of coexistence in the steady rise and competition scenario based on analysis of fitted models. a) Full posterior distribution of the niche difference ( ) and fitness differences calculated based on parameter distributions from the selected models. White circles show results based on average parameters. Small black dots show draws from posterior distribution. Shading indicates regions on the plane demarking predicted coexistence outcomes: blue corresponds to predicted exclusion of *D. pandora* at low temperatures, dark grey to the region of predicted coexistence, red to exclusion of *D. pallidifrons* and light grey corresponds to 'priority effects', where each species can exclude the other. b) Allocation of posterior support for each of the four possible predicted coexistence outcomes based on inferred niche and fitness differences. Cross at top of figure shows mean observed extirpation (+/- 1 SD) of *D. pallidifrons* in the competition and steady rise treatment (see Figure 2 for full distribution).

Modelling the system with simpler models that did not well represent the observed variability gave substantially different predictions. Using a Poisson error distribution coupled with the best fit competition model described above the system is predicted to instead move through a region of priority effects to a sharp shift in predicted dominance of *D. pandora* at an earlier temperature (Figure S6). The reuse of the same experiments to both make and test predictions reduces parameter uncertainty but introduces a certain degree of circularity. Refitting the *D. pallidifrons* model excluding the data from the 'steady-rise, competition' treatment increases the uncertainty of parameter estimates but gives comparable predictions (posterior support for exclusion at 27.4°C, central 66% interval: 26.5- 28.3°C, Figure S7).

Predictions based on the simulation-derived invasion growth rate closely aligned with the analytic prediction (Figure 4a, Model B). The predicted exclusion temperature of *D. pallidifrons* under fluctuating temperatures was 26.23°C (Figure 4a Model E, interpolating linearly between trialled temperature steps), slightly lower than without fluctuations. Predicted IGRs under variable environments show a sharp hinge beyond 26.6°C (Figure 4a). This stemmed from 'high-temperature' temperatures breaching with the sharp decline in modelled fecundity at high temperatures that

necessitated the introduction of a lower threshold. However, by this point, the key IGR threshold had already been breached and so the choice of threshold does not impact the key predictions.

Partitioning impact of different processes.

The partitioned additive contribution of both competition and temperature fluctuations had negative impacts on the IGR (Figure 4b). At temperatures closer to the transition point the negative pressure from competition reduced while the detrimental impact of the fluctuation increased. Combined, the effects had less joint negative impact – the interaction term between competition and fluctuations was positive, particularly at higher temperatures. As expected from the model structure, there was no detectable 'storage effect' component (magnitude of median values of $\Delta_{cov} < 0.0125$ for all central temperatures).

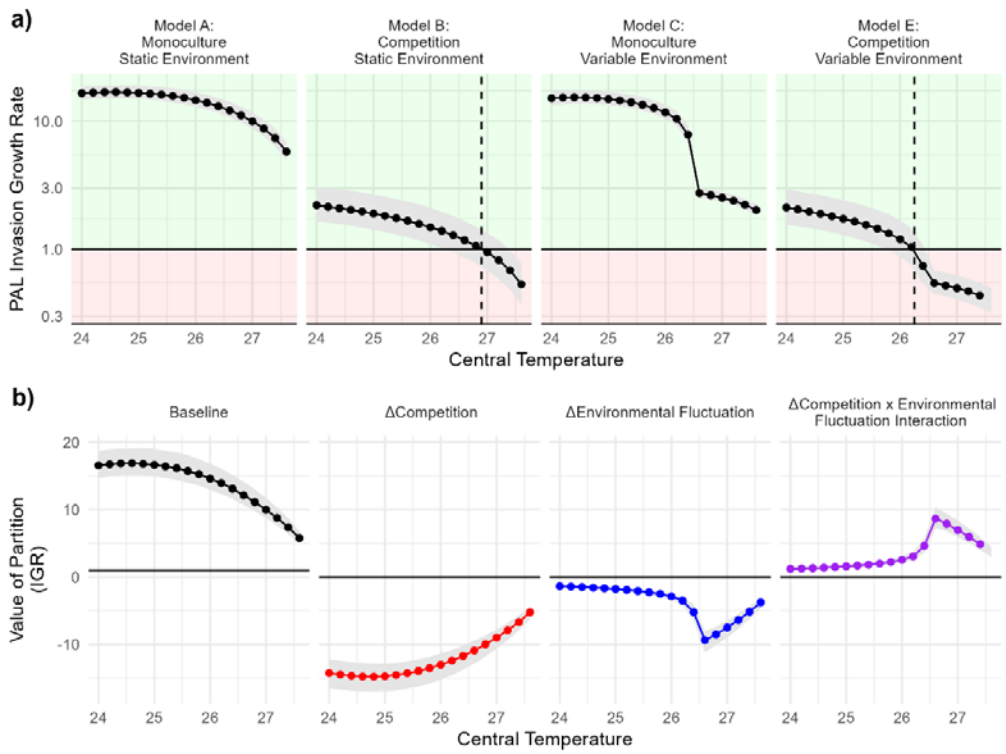

Figure 4. Simulation based decomposition of contributors to the long-term growth rate. Shaded areas show 66% central interval. a) Invasion growth rate (IGR) of *D. pallidifrons* inferred under different simulation scenarios. Results from model D (breaking the covariance between fluctuations in the competitors and the environment) were practically indistinguishable from Model E. b) Value of four key additive partitions from differences in models: $r_0, \Delta_C, \Delta_\sigma$ and Δ_* (Δ_{cov} was neglible).

Predictions of extinction by simulation methods.

The predictions from the direct simulation of the selected models were generally closer to the observed values than the predictions from coexistence theory (Table 1). However, the direct

simulations predict that competition has quite little marginal impact on extinction (hastening extinction by ~0.6 generations on average), contrary to the observed experimental data and the simulations. Extinctions from predictions under the simulation model were heavily influenced by the zero-inflated error distribution – simulations derived from a model using Poisson error predicted essentially no extinctions until thermal maximum was reached in G12.

Scenario		**Mean Generation of Observed Extinction**	**Predictions**		
				Direct Simulation	
Temperature Trajectory	**Competition**		**Coexistence theory**	**(ZINB)**	**(Poisson)**
Steady -rise	Competition	6.07 (3.83-9)	8.04 (6.25-9.375)	6.29 (3-9)	11.90 (12-12)
Variable-rise	Competition	5.77 (3 -8.17)	6.57 (5.39 - 7.12)	5.69 (3-9)	10.18 (19-12)
Steady -rise	Monoculture	7.55 (4 -10)		6.31 (3-9)	11.99 (12-12)
Variable-rise	Monoculture	5.62 (3-8)		5.73 (3-9)	10.38 (9-12)

Table 1. Predicted generations-until extinction under each experimental scenario. Predictions from the coexistence theory models have been converted from °C to generations for comparability with data. All intervals are central 66%. In the steady rise treatment, simulated extinctions very rarely occurred before the thermal limit causing mean observed generation of extinction to fall outside this central interval.

Discussion

In this experimental mesocosm climate-change scenario, coexistence theory was able to make predictions that approximately corresponded with the observations, despite many of the core assumptions not being fulfilled. Although spatial and multi-species effects were excluded by the experimental design, the theory was challenged to deal with non-stationary temporal dynamics, non-finite populations and timeframes, model adequacy as well as potential trait adaptation and positive density dependence. This correspondence appears to also extend to the impact of fluctuating environments, successfully capturing the impact of environmental variability including a strongly antagonistic interaction with competition.

This experiment directly tests coexistence theory both as a predictive tool and as a representative summary of complex ecological dynamics. Nonetheless, assessing the value of the accuracy of the model's predictions is inevitably somewhat subjective. An optimistic perspective of the alignment would highlight that an estimate within two generations is close enough to be useful. Given the discordance between the theoretical assumptions of the modern coexistence theory and the empirical reality of the experiment, it is surprising and heartening that the results are aligned at all. Further, the experiment does not include any possibility for *D. pallidifrons* recolonising after a stochastic loss,

likely speeding the observed extirpation in the mesocosms compared to a more natural situation. A pessimistic view would highlight the substantial temperature difference between observed and predicted extirpations (akin to several decades under most warming scenarios) and the high uncertainty in the prediction, despite levels of replication greatly exceeding the norms of the field. Furthermore, the mean generation of extinction mapped to the temperature with maximum support for coexistence between the species.

Although the full-simulation approach identifies only a small impact of competition on extinction, overall, it is more accurate than the 'theory-based' approach. Given then the ready alternative of directly building a simulation model, what can coexistence theory bring to the table? The analytic approach can identify that the combined effect of the warming on multiple temperature dependencies (Figure S4) to cause niche differences to increase (the rightwards drift in Figure 3a). This maintains coexistence longer than may be otherwise expected from an analysis focussing just on population fitness differences. Furthermore, the partitioning process is able to provide further insight on the various mechanisms, isolating a strong interactive effect between the two 'stressors', that was observed in the experimental data but not predicted in the full simulations. Given the qualitative dependency of the impact of variability on parameters (Terry *et al.* 2022), the successful assessment of the complex impact of variability in this noisy system is notable.

Aligning the scales of experiment and natural scenarios is a particular challenge for extinction dynamics. Experiments with more generations and shallower temperature gradients would allow finer resolution, but will be limited to an even more restricted repertoire of model species. Despite being a controlled mesocosm experiment, individual population trajectories were characterised by a high degree of variation (Figure S1). Counts were complete censuses with populations individually sexed and counted so minimal measurement error can be assumed. Population counts were of the order of dozens (mean of non-zero females counts = 19.9, σ =12.7) giving considerable scope for demographic stochasticity (Anderson *et al.* 1982) stemming from individual level-differences within the populations. Within-treatment 'external' environmental variability was minimised by the experimental design. Nonetheless, there was likely significant emergent environmental heterogeneity when defined as variation correlated amongst individuals in a vial (Shoemaker *et al.* 2020). Individual vials in the same treatment could be highly divergent, with some vials containing apparently stressed communities with late-developing pupae. Capturing this variation is a major challenge for applying coexistence theory. The best-fit models incorporated zero-inflated negative-binomial error distributions with temperature dependent parameters, a considerably more complex expression than is frequently assumed (Terry & Armitage 2024). Modelling the system with approaches that did not have the flexibility to represent this variability gave different predictions. Although it is likely chance, it is noteworthy that when using a Poisson error model, analytic theory gave more accurate and more

precise predictions of the loss of coexistence (Figure S6) despite direct simulations using Poisson being highly inaccurate.

It must nonetheless be acknowledged that the alignments observed in this experiment likely represent a 'best-case' expectation of accuracy under controlled conditions. Beyond the model and parameter uncertainty captured within ecological modelling, natural world forecasts must also grapple with climate change uncertainty (incorporating both human and earth system responses, Beaumont *et al.* 2008). The extent and accuracy of data available in this test is likely to exceed that available for many applied cases. While the forecasts could be considered acceptably accurate on average, the high heterogeneity among replicates suggests a strong fundamental limits of predictability at local scales. This heterogeneity arises even without incorporating dispersal variation (Melbourne & Hastings 2009) or the true field community complexity, including many more competitive rivals and dynamic interactions with natural enemies such as parasitoids (Chen & Lewis 2023). Rigorously applying coexistence theory to more complex communities requires significant extensions to the analysis approach (Chesson 2018; Ranjan *et al.* 2024; Spaak & Schreiber 2023) to tackle alternative invasion pathways, which are likely to place unrealistic demands on model parameterisation.

Strong Allee effects can challenge the calculation and use of an invasion growth rate, and are not represented in the fitted Beverton-Holt model. Quite strong positive density dependence (Allee effects) have been observed in competition response surface previous work with these species, where solitary mated females reduce the rate of eggs they lay (Terry *et al.* 2021). In *D. melanogaster* this social modulation effect has been well studied and is possibly associated with the need to overcome competition with fungi (Bailly *et al.* 2023; Wertheim *et al.* 2002). Nonetheless, in this experiment there were few cases with only one or two females and Allee effects appear to have had little impact on the dynamics.

The modelling framework used here assumes fixed traits, building in neither potential cross-generational acclimatory responses or adaptation. Acclimatation effects across and within generations in *Drosophila* can be considerable (Cavieres *et al.* 2020; Schiffer *et al.* 2013) and rapid evolutionary changes are also possible (e.g. Rudman *et al.* 2022), although the high intrinsic variation and relatively restricted range of temperatures may be expected to lead to comparatively little evolutionary genetic change over these timescales. While several authors have proposed extensions that incorporate evolutionary feedbacks (Pastore *et al.* 2021; Yamamichi *et al.* 2022), the high parametrisation requirements mean it is rarely incorporated into empirical applications. As such, for the purposes of this study 'no-trait changes' is effectively one of the assumptions under investigation. The temperature trajectory in this experiment will lead to a close match between temperatures across transitions, and as such any effects of acclimation are likely to be 'built-in' to the fitted thermal impacts.

In natural systems, the *Drosophila* exist in a highly dynamic metacommunity as individual ephemeral food resources become available and are consumed. Such a spatially distributed system will likely mitigate the substantial failure rate of individual patches. Further, within the mesocosms there was no distinction between the conditions under which the species' performance was assessed and those for the long-term experiment. In field conditions it can be hard to capture environmental impacts on all life-stages appropriately, and indeed each may undergo distinct environmental trajectories (Sinclair *et al.* 2016). Nonetheless, this study offers support that when carefully applied even quite abstract coexistence models can indeed make informative forecasts for contemporary change beyond the precise bounds of their formal applicability.

As communities reorganise in the face of changing environments, trustworthy tools are needed to offer guidance. There has been substantial and ongoing debate about how best to address the gap between the asymptotic theoretical world and the finite and messy reality that we must confront (Pande *et al.* 2020; Schreiber *et al.* 2023). Ultimately, observation and experimentation are essential to offer direction. Since there is not the luxury of sufficient time to conduct equivalent direct validation experiments in species with longer lifecycles, mesocosm experiments with high-turnover model species will have a major role to play. Validation tests such as those presented here can contextualise the expected accuracy from models to develop an expectation of the appropriate level of confidence in, and hence application and deployment of, coexistence theory.

Acknowledgments

JCDT was funded by a Leverhulme Trust Early Career Fellowship (ECF-2022-666). The help and support from the Oxford Biology Fly Lab and Community Ecology research groups is gratefully acknowledged. Highly constructive comments from three anonymous reviewers substantially improved the manuscript.

References

Abrams, P.A. (2022). *Competition Theory in Ecology*. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford, New York.

Adler, P.B., Dalgleish, H.J. & Ellner, S.P. (2012). Forecasting plant community impacts of climate variability and change: when do competitive interactions matter? *Journal of Ecology*, 100, 478–487.

Alexander, J.M., Chalmandrier, L., Lenoir, J., Burgess, T.I., Essl, F., Haider, S., *et al.* (2018). Lags in the response of mountain plant communities to climate change. *Global Change Biology*, 24, 563–579.

Alexander, J.M., Diez, J.M., Hart, S.P. & Levine, J.M. (2016). When Climate Reshuffles Competitors: A Call for Experimental Macroecology. *Trends in Ecology & Evolution*, 31, 831–841.

Alexander, J.M., Diez, J.M. & Levine, J.M. (2015). Novel competitors shape species' responses to climate change. *Nature*, 525, 515–518.

Anderson, R.M., Gordon, D.M., Crawley, M.J. & Hassell, M.P. (1982). Variability in the abundance of animal and plant species. *Nature*, 296, 245–248.

Aoyama, L., Shoemaker, L.G., Gilbert, B., Collinge, S.K., Faist, A.M., Shackelford, N., *et al.* (2022). Application of modern coexistence theory to rare plant restoration provides early indication of restoration trajectories. *Ecological Applications*, 32, e2649.

Armitage, D.W. & Jones, S.E. (2020). Coexistence barriers confine the poleward range of a globally distributed plant. *Ecology Letters*, 23, 1838–1848.

Bailly, T.P.M., Kohlmeier, P., Etienne, R.S., Wertheim, B. & Billeter, J.-C. (2023). Social modulation of oogenesis and egg laying in *Drosophila melanogaster*. *Current Biology*, 33, 2865-2877.e4.

Barabás, G., D'Andrea, R. & Stump, S.M. (2018). Chesson's coexistence theory. *Ecological Monographs*, 88, 277–303.

Beaumont, L.J., Hughes, L. & Pitman, A.J. (2008). Why is the choice of future climate scenarios for species distribution modelling important? *Ecology Letters*, 11, 1135–1146.

Bimler, M.D., Stouffer, D.B., Lai, H.R. & Mayfield, M.M. (2018). Accurate predictions of coexistence in natural systems require the inclusion of facilitative interactions and environmental dependency. *Journal of Ecology*, 106, 1839–1852.

Bowler, C.H., Weiss-Lehman, C., Towers, I.R., Mayfield, M.M. & Shoemaker, L.G. (2022). Accounting for demographic uncertainty increases predictions for species coexistence: A case study with annual plants. *Ecology Letters*, 25, 1618–1628.

Briere, J.-F., Pracros, P., Le Roux, A.-Y. & Pierre, J.-S. (1999). A Novel Rate Model of Temperature-Dependent Development for Arthropods. *Environmental Entomology*, 28, 22–29.

Buche, L., Spaak, J.W., Jarillo, J. & De Laender, F. (2022). Niche differences, not fitness differences, explain predicted coexistence across ecological groups. *Journal of Ecology*, 110, 2785–2796.

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10, 395–411.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., *et al.* (2017). Stan□: A Probabilistic Programming Language. *Journal of Statistical Software*, 76.

Cavieres, G., Rezende, E.L., Clavijo-Baquet, S., Alruiz, J.M., Rivera-Rebella, C., Boher, F., *et al.* (2020). Rapid within- and transgenerational changes in thermal tolerance and fitness in variable thermal landscapes. *Ecology and Evolution*, 10, 8105–8113.

Cervantes-Loreto, A., Pastore, A.I., Brown, C.R.P., Maraffini, M.L., Aldebert, C., Mayfield, M.M., *et al.* (2023). Environmental context, parameter sensitivity, and structural sensitivity impact predictions of annual-plant coexistence. *Ecological Monographs*, e1592.

Chen, I.-C., Hill, J.K., Ohlemüller, R., Roy, D.B. & Thomas, C.D. (2011). Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, 333, 1024–1026.

Chen, J. & Lewis, O.T. (2023). Experimental heatwaves facilitate invasion and alter species interactions and composition in a tropical host-parasitoid community. *Global Change Biology*, 29, 6261–6275.

Chen, J. & Lewis, O.T. (2024). Limits to species distributions on tropical mountains shift from high temperature to competition as elevation increases. *Ecological Monographs*, 94, e1597.

Chesson, P. (1990). MacArthur's consumer-resource model. *Theoretical Population Biology*, 37, 26–38.

Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology, Evolution, and Systematics*, 31, 343–366.

Chesson, P. (2018). Updates on mechanisms of maintenance of species diversity. *Journal of Ecology*, 106, 1773–1794.

Clark, A., Hillebrand, H. & Harpole, W.S. (2019). Scale Both Confounds and Informs Characterization of Species Coexistence in Empirical Systems. *The American Naturalist*, 194, 794–806.

Courchamp, F., Clutton-Brock, T. & Grenfell, B. (1999). Inverse density dependence and the Allee effect. *Trends in Ecology & Evolution*, 14, 405–410.

Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B. & Wood, S. (1998). Making mistakes when predicting shifts in species range in response to global warming. *Nature*, 391, 783–786.

Descombes, P., Pitteloud, C., Glauser, G., Defossez, E., Kergunteuil, A., Allard, P.-M., *et al.* (2020). Novel trophic interactions under climate change promote alpine plant coexistence. *Science*, 370, 1469–1473.

Ellner, S.P., Snyder, R.E., Adler, P.B. & Hooker, G. (2019). An expanded modern coexistence theory for empirical applications. *Ecology Letters*, 22, 3–18.

Freeman, B.G., Strimas-Mackey, M. & Miller, E.T. (2022). Interspecific competition limits bird species' ranges in tropical mountains. *Science*, 377, 416–420.

Getz, W.M., Marshall, C.R., Carlson, C.J., Giuggioli, L., Ryan, S.J., Romañach, S.S., *et al.* (2018). Making ecological models adequate. *Ecology Letters*, 21, 153–166.

Godoy, O., Kraft, N.J.B. & Levine, J.M. (2014). Phylogenetic relatedness and the determinants of competitive outcomes. *Ecology Letters*, 17, 836–844.

Grainger, T.N., Levine, J.M. & Gilbert, B. (2019). The Invasion Criterion: A Common Currency for Ecological Research. *Trends in ecology & evolution*, 34, 925–935.

Hallett, L.M., Aoyama, L., Barabás, G., Gilbert, B., Larios, L., Shackelford, N., *et al.* (2023). Restoration ecology through the lens of coexistence theory. *Trends in Ecology & Evolution*, 38, 1085–1096.

Hangartner, S.B., Hoffmann, A.A., Smith, A. & Griffin, P.C. (2015). A collection of Australian Drosophila datasets on climate adaptation and species distributions. *Scientific Data*, 2, 150067.

Hart, S.P., Usinowicz, J. & Levine, J.M. (2017). The spatial scales of species coexistence. *Nature Ecology & Evolution*, 1, 1066–1073.

Jeffs, C.T., Terry, J.C.D., Higgie, M., Jandová, A., Konvičková, H., Brown, J.J., *et al.* (2021). Molecular analyses reveal consistent food web structure with elevation in rainforest Drosophila – parasitoid communities. *Ecography*, 44, 403–413.

Kellermann, V., Overgaard, J., Hoffmann, A.A., Flojgaard, C., Svenning, J.-C. & Loeschcke, V. (2012). Upper thermal limits of Drosophila are linked to species distributions and strongly constrained phylogenetically. *Proceedings of the National Academy of Sciences*, 109, 16228–16233.

Kontopoulos, D.-G., Sentis, A., Daufresne, M., Dell, A.I. & Pawar, S. (2023). No model to rule them all: a systematic comparison of 83 thermal performance curve models across traits and taxonomic groups.

Kraft, N.J.B., Godoy, O. & Levine, J.M. (2015). Plant functional traits and the multidimensional nature of species coexistence. *Proceedings of the National Academy of Sciences*, 112, 797–802.

Legault, G., Bitters, M.E., Hastings, A. & Melbourne, B.A. (2020). Interspecific competition slows range expansion and shapes range boundaries. *Proceedings of the National Academy of Sciences*, 117, 26854–26860.

Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Murienne, J., *et al.* (2020). Species better track climate warming in the oceans than on land. *Nature Ecology & Evolution*, 4, 1044–1059.

Letten, A.D., Dhami, M.K., Ke, P.-J. & Fukami, T. (2018). Species coexistence through simultaneous fluctuation-dependent mechanisms. *Proceedings of the National Academy of Sciences*, 115, 6745–6750.

McEvey, S.F. & Schiffer, M. (2015). New species in the Drosophila ananassae subgroup from northern Australia, New Guinea and the South Pacific (Diptera: Drosophilidae), with historical overview. *Records of the Australian Museum*, 67, 129–161.

Melbourne, B.A. & Hastings, A. (2009). Highly Variable Spread Rates in Replicated Biological Invasions: Fundamental Limits to Predictability. *Science*, 325, 1536–1539.

Mordecai, E.A. (2013). Despite spillover, a shared pathogen promotes native plant persistence in a cheatgrass-invaded grassland. *Ecology*, 94, 2744–2753.

O'Brien, E.K., Higgie, M., Reynolds, A., Hoffmann, A.A. & Bridle, J.R. (2017). Testing for local adaptation and evolutionary potential along altitudinal gradients in rainforest Drosophila: beyond laboratory estimates. *Global Change Biology*, 23, 1847–1860.

Orr, J.A., Vinebrooke, R.D., Jackson, M.C., Kroeker, K.J., Kordas, R.L., Mantyka-Pringle, C., *et al.* (2020). Towards a unified study of multiple stressors: divisions and common goals across research disciplines. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20200421.

Pande, J., Fung, T., Chisholm, R. & Shnerb, N.M. (2020). Mean growth rate when rare is not a reliable metric for persistence of species. *Ecology Letters*, 23, 282.

Pastore, A.I., Barabás, G., Bimler, M.D., Mayfield, M.M. & Miller, T.E. (2021). The evolution of niche overlap and competitive differences. *Nat Ecol Evol*, 5, 330–337.

Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I.C., *et al.* (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, 355, eaai9214.

Ranjan, R., Koffel, T. & Klausmeier, C.A. (2024). The three-species problem: Incorporating competitive asymmetry and intransitivity in modern coexistence theory. *Ecology Letters*, 27, e14426.

Rudman, S.M., Greenblum, S.I., Rajpurohit, S., Betancourt, N.J., Hanna, J., Tilk, S., *et al.* (2022). Direct observation of adaptive tracking on ecological time scales in Drosophila. *Science*.

Schiffer, M., Hangartner, S. & Hoffmann, A.A. (2013). Assessing the relative importance of environmental effects, carry-over effects and species differences in thermal stress resistance: a comparison of Drosophilids across field and laboratory generations. *Journal of Experimental Biology*, 216, 3790–3798.

Schreiber, S.J., Levine, J.M., Godoy, O., Kraft, N.J.B. & Hart, S.P. (2023). Does deterministic coexistence theory matter in a finite world? *Ecology*, 104, e3838.

Shoemaker, L.G., Sullivan, L.L., Donohue, I., Cabral, J.S., Williams, R.J., Mayfield, M.M., *et al.* (2020). Integrating the underlying structure of stochasticity into community ecology. *Ecology*, 101, e02652.

Simha, A., Pardo-De la Hoz, C.J. & Carley, L.N. (2022). Moving beyond the "Diversity Paradox": The Limitations of Competition-Based Frameworks in Understanding Species Diversity. *The American Naturalist*, 200, 89–100.

Sinclair, B.J., Marshall, K.E., Sewell, M.A., Levesque, D.L., Willett, C.S., Slotsbo, S., *et al.* (2016). Can we predict ectotherm responses to climate change using thermal performance curves and body temperatures? *Ecology Letters*, 19, 1372–1385.

Spaak, J.W., Ke, P.-J., Letten, A.D. & De Laender, F. (2023). Different measures of niche and fitness differences tell different tales. *Oikos*, 2023, e09573.

Spaak, J.W. & Schreiber, S.J. (2023). Building modern coexistence theory from the ground up: The role of community assembly. *Ecology Letters*, 26, 1840–1861.

Suttle, K.B., Thomsen, M.A. & Power, M.E. (2007). Species Interactions Reverse Grassland Responses to Changing Climate. *Science*, 315, 640–642.

Terry, J.C.D. & Armitage, D.W. (2024). Widespread analytical pitfalls in empirical coexistence studies and a checklist for improving their statistical robustness. *Methods in Ecology and Evolution*, 15, 594–611.

Terry, J.C.D., Chen, J. & Lewis, O.T. (2021). Natural enemies have inconsistent impacts on the coexistence of competing species. *Journal of Animal Ecology*, 90, 2277–2288.

Terry, J.C.D., O'Sullivan, J.D. & Rossberg, A.G. (2022). Synthesising the multiple impacts of climatic variability on community responses to climate change. *Ecography*, 2022, e06123.

Van Dyke, M.N., Levine, J.M. & Kraft, N.J.B. (2022). Small rainfall changes drive substantial changes in plant coexistence. *Nature*, 611, 507–511.

Vehtari, A., Gelman, A. & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*, 27, 1413–1432.

Wainwright, C.E., HilleRisLambers, J., Lai, H.R., Loy, X. & Mayfield, M.M. (2019). Distinct responses of niche and fitness differences to water availability underlie variable

coexistence outcomes in semi-arid annual plant communities. *Journal of Ecology*, 107, 293–306.
Wertheim, B., Marchais, J., Vet, L.E.M. & Dicke, M. (2002). Allee effect in larval resource exploitation in Drosophila: an interaction among density of adults, larvae, and micro-organisms. *Ecological Entomology*, 27, 608–617.
Yamamichi, M., Gibbs, T. & Levine, J.M. (2022). Integrating eco-evolutionary dynamics and modern coexistence theory. *Ecology Letters*, 25, 2091–2106.