Predicting evolution in experimental range expansions of an aquatic model system

Giacomo Zilio[1*], Sascha Krenek[2], Claire Gougat-Barbera[1], Emanuel A. Fronhofer[1], Oliver Kaltz[1]

Giacomo Zilio[1*], gcm.zilio@gmail.com, ORCID ID: https://orcid.org/0000-0002-4448-3118

Sascha Krenek[2], krenek@bafg.de, ORCID ID https://orcid.org/0000-0002-5062-4140

Claire Gougat-Barbera[1], claire.gougat-barbera@umontpellier.fr, ORCID ID: unavailable

Emanuel A. Fronhofer[1], emanuel.fronhofer@umontpellier.fr, ORCID ID: https://orcid.org/0000-0002-2219-11 784X

Oliver Kaltz[1*], oliver.kaltz@umontpellier.fr, ORCID ID: https://orcid.org/0000-0002-7154-0456

[1] ISEM, University of Montpellier, CNRS, EPHE, IRD, Montpellier, France.

[2] Institute of Hydrobiology, Technische Universität Dresden, Dresden, Germany

[*]Corresponding authors: gcm.zilio@gmail.com, oliver.kaltz@umontpellier.fr

Abstract

Predicting range expansion dynamics is a challenge for both fundamental and applied research in conservation and global change biology. However, if ecological and evolutionary processes occur on the same time scale, predictions are challenging to make. Combining experimental evolution and mathematical modelling, we assessed the predictability of independent realisations of range expansions in a laboratory model system, the freshwater protozoan *Paramecium caudatum*. We followed ecological dynamics and evolutionary change in range core and front populations in the experiment. These settings were recreated in a predictive mathematical model, parametrized with dispersal and growth data of the of the 20 founder strains in the experiment. We find that short-term evolution was driven by selection for increased dispersal at the front and general selection for higher growth rates in all treatments. There was a good quantitative match of predicted and observed trait changes. Phenotypic divergence was mirrored by a complete genotypic divergence, indicating the highly repeatable fixation of strains that also were the most likely winners in our model. Long-term evolution in the experimental range front lines resulted in the emergence of a dispersal syndrome, namely a competition - colonisation trade-off. Altogether, both model and experiment highlight the importance of dispersal evolution as a driver of range expansions. Our study suggests that evolution at range fronts may follow predictable trajectories, at least for simple scenarios, and that predicting these dynamics may be possible from knowledge of few key parameters.

Keywords

Dispersal evolution, reaction-diffusion model, predictability, eco-evolutionary dynamics, invasions

Introduction

Predicting ecological dynamics and species' range shifts has become a major challenge for conservation and management strategies in times of global climate and environmental change (Petchey *et al.*, 2015). Indeed, whether the outcomes of range expansions or biological invasions can be predicted at all remains highly debated in ecology even in simple settings, due to the intrinsic stochasticity of these phenomena (Melbourne & Hastings, 2009; Giometto *et al.*, 2014). Moreover, evolutionary processes occur at the same time scale as ecological dynamics during range expansions (Perkins *et al.*, 2013; Williams *et al.*, 2016), potentially exacerbating the uncertainty of outcomes (Williams *et al.*, 2019).

Theory shows that range expansions can involve the concurrent evolution of dispersal and other traits (Perkins *et al.*, 2013; Kubisch *et al.*, 2014) and lead to the emergence of dispersal syndromes (Clobert *et al.*, 2012; Cote *et al.*, 2017). Individuals with greater dispersal propensity are the first to reach the range front, and they will reproduce with conspecifics that have the same fast spreader characteristics (Thomas *et al.*, 2001; Hughes *et al.*, 2007). Consequently, high dispersal ability and correlated life-history traits evolve in the range front populations due to spatial selection and spatially assortative mating (Phillips *et al.*, 2008; Shine *et al.*, 2011). Since expansion speeds are mainly influenced by dispersal and reproduction (Fisher, 1937; Kolgomorov et al. 1937), the two traits can be rapidly selected and evolve simultaneously. However, if dispersal is costly (Bonte *et al.*, 2012) there may be trade-offs with other traits. Higher reproduction at the range front may come at the expense of lower competitive ability (Burton *et al.*, 2010), recalling the competition-colonisation trade-off in classic species coexistence models (Calcagno *et al.*, 2006).

Fast evolution in range front populations can produce eco-evolutionary feedbacks and thereby speed up the expansion process (Shine *et al.*, 2011; Chuang & Peterson, 2016; Ochocki *et al.*, 2019; Williams *et al.*, 2019; Miller *et al.*, 2020). In the emblematic example of the cane toad (*Rhinella marina*) expansion in Australia, increased dispersal at the range front coincided with evolutionary change in behavioural, morphological and demographic traits, promoting the speed of the toad expansion (Phillips *et al.*, 2006; Perkins *et al.*, 2013; Brown *et al.*, 2014). Growing empirical evidence from other natural populations and biological systems (Simmons & Thomas, 2004; Alford *et al.*, 2009; Leotard *et al.*, 2009; Lombaert *et al.*, 2014) suggest that dispersal evolution at range fronts is a common phenomenon. Recently, experimental evolution and microcosm landscapes have been

used to test fundamental predictions (Friedenberg, 2003) and mimic range expansions in the laboratory. Experiments with ciliates (Fronhofer and Altermatt, 2015), arthropods (Ochocki & Miller, 2017; Szűcs *et al.*, 2017; Weiss-Lehman *et al.*, 2017; Petegem *et al.*, 2018) or plants (Williams *et al.*, 2016) all showed the rapid evolution of dispersal and other dispersal-related traits during the experimental range expansions. However, whether we can accurately predict these eco-evolutionary dynamics from prior information on the genetic or phenotypic characteristics of the expanding populations remains an open question.

Coupling microcosm experiments with mathematical modelling and genetic analyses provides a possible way forward to assess the predictability of range expansions (Nosil *et al.*, 2020). In micro/mesocosm landscapes, we can study the repeatability of range expansions through independent replicates under controlled conditions. Using specifically tailored and parametrised mathematical models, we can then formalise putative processes of range expansion dynamics and confront predicted with observed outcomes. Genetic analysis can further characterise the degree of similarity among experimental replicates and link phenotypic trait change to genetic change.

Here, moving a step forward from previous ecological models (Melbourne & Hastings, 2009; Giometto *et al.*, 2014), we employed such a combined approach to assess the predictability of evolutionary outcomes of range expansions in an aquatic model organism, the freshwater protozoan *Paramecium caudatum*. Following previous studies (Fronhofer & Altermatt (2015) Nørgaard et al. (2021), we used interconnected 2-patch systems to establish a range front treatment, where recurrent episodes of dispersal alternated with intermittent periods of population growth. In the contrasting range core treatment, only the non-dispersing individuals were maintained. We recreated these experimental treatments in a predictive mathematical model, parameterised for dispersal and growth characteristics of the 20 *Paramecium* strains that were used to assemble the founder population in the selection experiment. Based on selection from standing genetic variation in an asexual population, the model predicted the rapid divergence between range core and front populations, mainly driven by positive selection on dispersal at the front. There was a good quantitative match between model predictions and experimental results, and the most likely winner strains identified by the model corresponded to particular genotypes, found to be repeatably fixed in the experimental core and front populations. In the long run (160 dispersal/growth episodes), range core and front populations continued to diverge, resulting in the emergence of a

dispersal syndrome with a competition – colonisation trade-off, hypothesised in many ecological models (Livingston *et al.*, 2012). Our results suggest that, even with evolution occurring over short ecological time scales, range expansions may follow predictable trajectories and predicting these dynamics may be possible from knowledge of a few key parameters.

Material and Methods

Study organism and strains

Paramecium caudatum is a freshwater ciliate with a world-wide distribution, feeding on bacteria and detritus. Asexual reproduction occurs by mitotic division and represents the main mode of population growth. Swimming is accomplished through the coordinated movement of ciliary bands on the cell surface (Wichterman, 1986). Previous work on *P. caudatum* indicated a genetic basis of dispersal propensity, and plastic responses are induced by biotic factors, such as parasitism, chemical predator signals, or population density (Fellous *et al.*, 2012; Fronhofer *et al.*, 2018; Zilio *et al.*, 2021). Here, we used 20 *Paramecium* strains (i.e., clonal mass cultures derived from a single individual) from various geographic origins (Weiler *et al.*, 2020; Zilio *et al.*, 2021) and representing different groups of mitochondrial haplotypes ("COI genotypes" or "genotypes", hereafter; Table S1). All cultures were reared under standard laboratory conditions in lettuce medium with the food bacterium *Serratia marcescens* at 23 °C, allowing up to 3 asexual doublings per day (Nidelet & Kaltz, 2007).

Founder strains measurements

Prior to the start of the long-term experiment, we assayed the 20 founder strains for dispersal and population growth characteristics (Table S1). For the dispersal assay, we placed aliquots of 8 mL of culture (at equilibrium density) in a 2-patch system (see below for additional details) and let the *Paramecium* disperse for 3h. Once connections were blocked, we estimated the number of residents and dispersers by taking 150-600 µL samples from the two tubes and counting the number of individuals under a dissecting microscope. Dispersal was taken as the proportion of dispersers of the total number of individuals in the system. Dispersal rates (Gaussian posteriors) were then estimated using a generalized linear mixed model (binomial error distribution; bobyqa optimizer in function glmer of R package "lme4"; Bates *et al.*, 2015) for each strain with time and observation (to account for overdispersion) as random effects. We tested 4 replicates per strain. For the growth assay, we placed ca. 200 individuals (from cultures at equilibrium density) in 20 mL of fresh medium.

158 Over the course of 6 days, we tracked population density by counting the number of individuals in
159 daily samples of 100-200 µL. We tested 3 replicates per strain. Using a Bayesian approach
160 (Rosenbaum *et al.*, 2019), we estimated the intrinsic population growth rate (r_0) and equilibrium
161 density ($\overline{N}$) for each replicate by fitting a Beverton-Holt population growth model to the time series
162 data. Details of the Bayesian fitting are given in Supplementary Information (Section S1).
163
164 **Selection experiment**
165 The selection experiment comprised a sequence of cycles, where dispersal events alternated with
166 periods of population growth. The founder population was created by mixing the 20 strains at equal
167 proportions in a single mass culture, which was then divided up into 15 replicate selection lines,
168 assigned to the following three treatments. First, in the range front treatment (6 selection lines), we
169 placed the *Paramecium* in one of the two tubes in 2-patch dispersal systems (interconnected 15-mL
170 tubes, Fig. 1A). Connections were opened for 3 h, during which time individuals were allowed to
171 swim to the other tube. We then collected the dispersers and cultured them for 1 week under
172 permissive conditions in 20 mL of fresh medium (in 50-mL plastic tubes), until we initiated a new
173 round of dispersal, again only retaining the dispersers and culturing them for 1 week, and so on.
174 Second, the range core treatment (6 selection lines) followed the same cycles of dispersal and
175 growth, but only the non-dispersing residents were retained after each dispersal episode. Third, in
176 the control treatment (3 selection lines), residents and dispersers were mixed after each dispersal
177 event and then cultured for 1 week, as in the other treatments. In corollary, the range front
178 treatment mimics the advancing cohort of a spatially expanding population, whereas populations
179 from the core treatment remain in place and constantly lose emigrants. The control treatment is
180 similar to the core treatment, except for the loss of emigrants. A total of 161 cycles were
181 accomplished. Prior to each dispersal event, ca. 1800 individuals (median; 25% / 75% quantile range:
182 1400 / 2700) were placed in the dispersal systems. After dispersal, the number of individuals starting
183 the 1-week growth period were matched between treatments. Because dispersal rates were low at
184 the beginning, these starting numbers were initially set to 200 individuals (placed in a total volume
185 of 20 mL of fresh medium). During the following 1-week growth period, stable population sizes were
186 typically reached within 3-4 days, with densities of ca. 240 individuals per ml (median; 25% / 75%
187 quantiles: 180 / 360). After cycle 32, when dispersal had already reached higher levels (see Results),
188 we adjusted the starting numbers to ca. 1500 (median; 25% / 75% quantiles: 1100 / 2000).
189

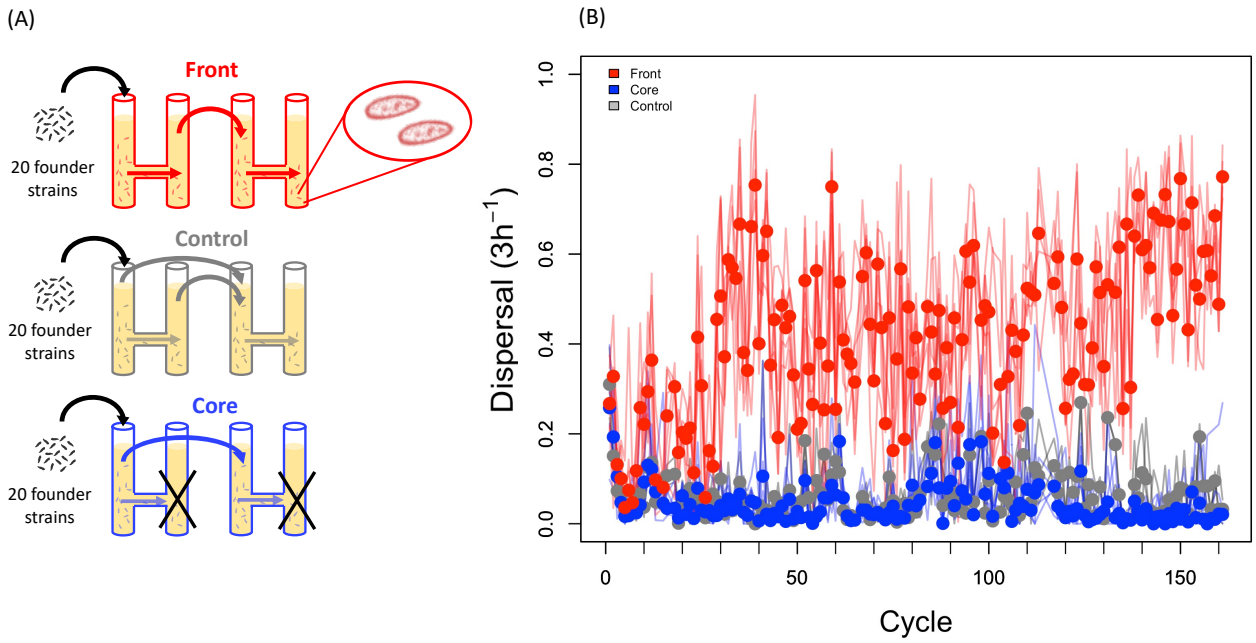


Figure 1 Design of range expansion selection experiment and long-term time series of dispersal in the experimental treatments. (A) Starting from a mix of 20 *Paramecium caudatum* founder strains, experimental population were allowed to disperse in 2-patch dispersal systems. For the range front treatment (red), only the dispersers were maintained and propagated for 1 week, until the next dispersal episode. In the core treatment (blue), only the non-dispersing residents were maintained at each cycle. In the control treatment, both residents and dispersers were maintained. (B) Observed levels of dispersal over the whole duration of the experiment (161 cycles, ca. 3 years). Lines show the trajectories for the individual selection lines (n = 15), the circles indicate the mean dispersal per treatment and cycle.

Data collection

For each selection line, dispersal was measured at each dispersal event and equilibrium densities ($\overline{N}$) taken at the end of the 1-week growth period at each cycle, as described above. Furthermore, growth rate (r_0) was determined in assays conducted at cycle 21 (year 1), 78 (year 2) and 160 (year 3), as described above, with 2-3 replicates per selection line and year. Bayesian model fitting was used to estimate r_0 (Section S1). Measurement of swimming behaviour were also taken in the first two years of the experiment (Section S2).

Genotyping

All founder strains were genotyped for the mitochondrial cytochrome *c* oxidase I (COI) gene, by extracting DNA from 10 cells per strain and applying a PCR amplification protocol and sequence analysis described in Killeen et al. (2017). At cycle 30 in the selection experiment, mixes of 50 cells from each selection line were processed in the same way and analysed for (multiple) COI marker

signals. This method characterises the line for the most frequent COI genotype, and it has a resolution threshold of c. 5%, i.e., it can detect 2-3 cells of a minority genotype in the sample (Killeen *et al.*, 2017). The sequences used are deposited in GenBank under the accession numbers listed in Barth et al. (2006) and Weiler *et al.* (2020).

Range expansion model

Our model is designed to capture the specificities of the selection experiment and the characteristics of the strains in the founder population. Thus, we model the population dynamics of *Paramecium* strains, assuming logistic growth following the Verhulst equation expanded to include both intra- and inter-strain competition:

$$\frac{dN_i}{dt} = (r_{0,i} - \sum_j (\alpha_{ij} N_j))\, N_i$$

where N_i is the population size of strain i, $r_{0,i}$ is its intrinsic rate of increase and α_{ij} as the competition coefficients. The model is parametrized with the posteriors extracted from growth curve fitting (r_0, $\overline{N}$), as described above. We make the simplifying assumption that intra- and interspecific competition is of equal strength. We further assume a quasi-extinction threshold of 0.7 (the bottleneck occurring during dispersal, we have tested the effect of different quasi-extinction thresholds ranging from 0.0001 to 0.9), which implies that strains experience an extinction if they exhibit densities below this value.

We model the community dynamics of the strains for 7 days, followed by a 3h dispersal phase in a 2-patch metapopulation. During this 3h dispersal phase all strains can disperse from their patch of origin to their destination patch according to the dispersal rates estimated from dispersal assay; the model is parametrized with posteriors extracted from the statistical analysis described above. After the dispersal phase we follow the patch of origin (residents in the range core treatment), the destination patch (dispersers in the front treatment) or the combined patches (dispersers and residents mixed in the control treatment). We repeat this procedure for a total of 10 iterations. As in the experiment, we control for densities between rounds of iteration by selecting the equivalent of 10 mL samples.

This approach allows us to predict, based only on measurements of growth parameters and dispersal rates of the founder strains, which strains should predominate in each of the three treatments at the end of the experiment. It is important to keep in mind that the underlying model is deterministic. However, since we parametrize the model with draws from posteriors, our approach takes into account the uncertainty associated with the data and yields a distribution of likely outcomes, given these uncertainties. Note that our model depicts a scenario of selection from standing genetic variation; it does not include mutational change.

Statistical analysis

Statistical analyses were performed in R (ver. 4.1.2; R CoreTeam 2021 and JMP 14 (SAS Institute Inc. 2018). We analysed dispersal (proportion of dispersers), using generalised linear models (GLM) with binomial error distribution. We considered selection treatment (core, front, control), experimental cycle and selection line (nested within treatment) as explanatory variables. We analysed variation in intrinsic population growth rate (r_0) and equilibrium density ($\overline{N}$; averages per selection line and year) using GLMs, with selection treatment, year and selection line as explanatory variables. To illustrate how selection acts on standing genetic variation in our model, we associated the winning probability of each of the 20 founder strains with their respective median values of dispersal, r_0, and $\overline{N}$ from the distributions used by the model. For each treatment, we then performed multiple regressions, with winning probability as response variable and the three traits as explanatory variables. To investigate associations between dispersal, r_0 and $\overline{N}$, we constructed a data matrix based on trait means per year and selection line (3 years x 15 selection lines, n = 45), after centering and scaling trait distributions. One range-front line was lost in year 3, leading to n = 44. We used a Bayesian approach with the "rstan" package version 2.21.2 (Carpenter *et al.*, 2017) to estimate pairwise trait correlations (see section S3). We also performed a principal component analysis (PCA), considering the joint variation in all three traits.

Results

Predicted and observed short-term trait evolution

Over the first 25 cycles of the selection experiment, we observed a strong increase in dispersal in the range front treatment. Dispersal reached 22.3% (± 0.012 SE, averaged over cycles 15-25) at the front, compared to only 4.4% (± 0.004 SE) in the core treatment and 6 % (± 0.012 SE) in the control treatment (effect of selection treatment: $\chi_2^2 = 119.7$; $p < 0.001$). Increased dispersal at the front

established within only a few cycles, and was formally significant for the first time at cycle 8 (cycle-by-cycle analysis: $p < 0.001$). Our parametrized model captured this rapid increase of dispersal in the range front treatment (Fig. 1B), and there was a quantitative match between the distribution of endpoint levels of dispersal in the model and observed values in the experiments (Fig. 2A). The model further predicted general increases in growth rate (r_0) and equilibrium density ($\overline{N}$) in all treatments, from 0.07 in the ancestral mix to 0.08 in core and front end-point populations. This is consistent with results from the growth assay conducted at cycle 21, where estimates of r_0 for the 15 selection lines are well within the central range of predicted values in the model (Fig. 2B). As predicted, selection treatments did not significantly differ in r_0 (treatment: $F_{2,12} = 1.2$; $p = 0.354$). Unlike in the model, range front lines produced nearly 20% higher equilibrium densities than did range core and control lines (treatment: $F_{2,12} = 11.1$; $p = 0.003$).

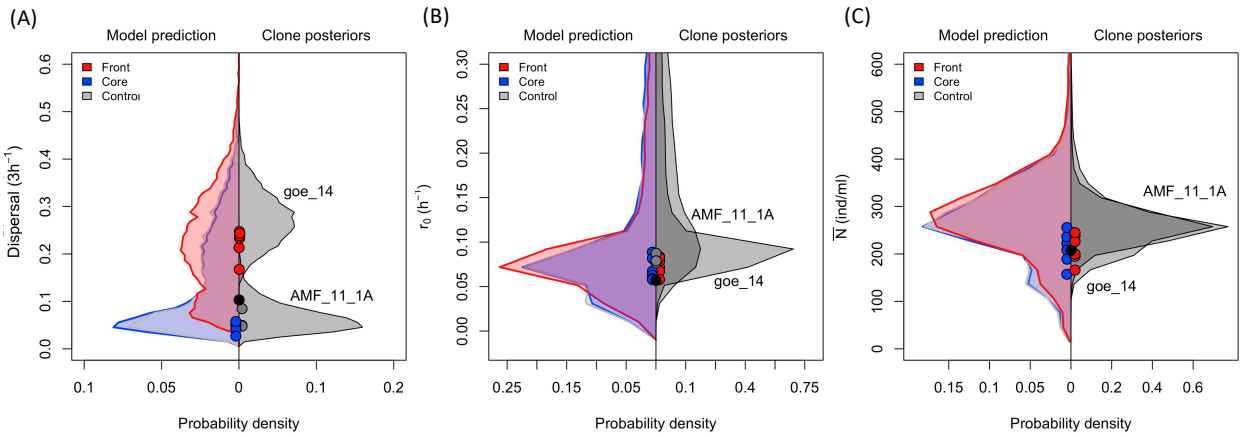


Figure 2 Model endpoint predictions for (A) dispersal, (B) growth rate (r_0), and (C) equilibrium density ($\overline{N}$). In each panel, left: model predictions for the 3 treatments; right: posteriors distributions of the most likely winner strains in the range core (AMF_11_11A) and range front (goe_14) treatment. Circles are the average values measured for each experimental selection lines after 15-25 cycles ("short term"). Different colours represent the different treatments. The black circles represent the ancestral means (founder population).

Predicted and observed short-term changes in strain composition

Our model finds strong variation in the fixation probability among the 20 strains, and different treatments have different most likely winners (range: 0.7-16.8 %; Fig. S7).

For the range front treatment, multiple regression analysis (Table S1) shows that both dispersal and r_0 are positively associated with strain winning probability, and this with equal strength

(standardised beta (β) regression coefficients: +0.55 and +0.64, respectively; Fig. 3). Thus, selection is predicted to favour strains that both disperse more and grow faster. In contrast, in the core and control treatments, strain winning probability is mainly associated with high growth rate ($\beta > +0.96$), accompanied by weak selection against clones with higher dispersal ($\beta \leq -0.27$) or equilibrium density ($\beta \leq -0.33$).

Molecular analysis of the 15 selection lines indicates complete genetic divergence between selection treatments. For all 9 range core and control lines, only the b05 COI genotype was detected. The two strains in the founder population that carry this genotype (Table S1) have very high growth rates and very low dispersal, the trait combination favoured in the model. Indeed, the candidate strain AMF_11_1A has the highest growth rate overall and is the most likely winner in core and control treatments according to our model (Fig. 3). In contrast, all 6 range front selection lines appear to be fixed for the b07 COI genotype. This genotype is shared by 13 founder strains (Table S1), which may thus have gone to fixation in groups or individually. Among these candidate strains is the most likely winner (goe_14) predicted by the model: it has the highest growth rate and the third-highest dispersal, in line with the prediction of the two traits being under joint positive selection in this treatment. As shown in Fig. 2, trait values of the most likely front and core winner strains (goe14 vs AMF_11_1A; strain posterior distributions on the right) show a good match with both the predicted model outcomes (distributions on the left; Fig. 2) and the experimental data.

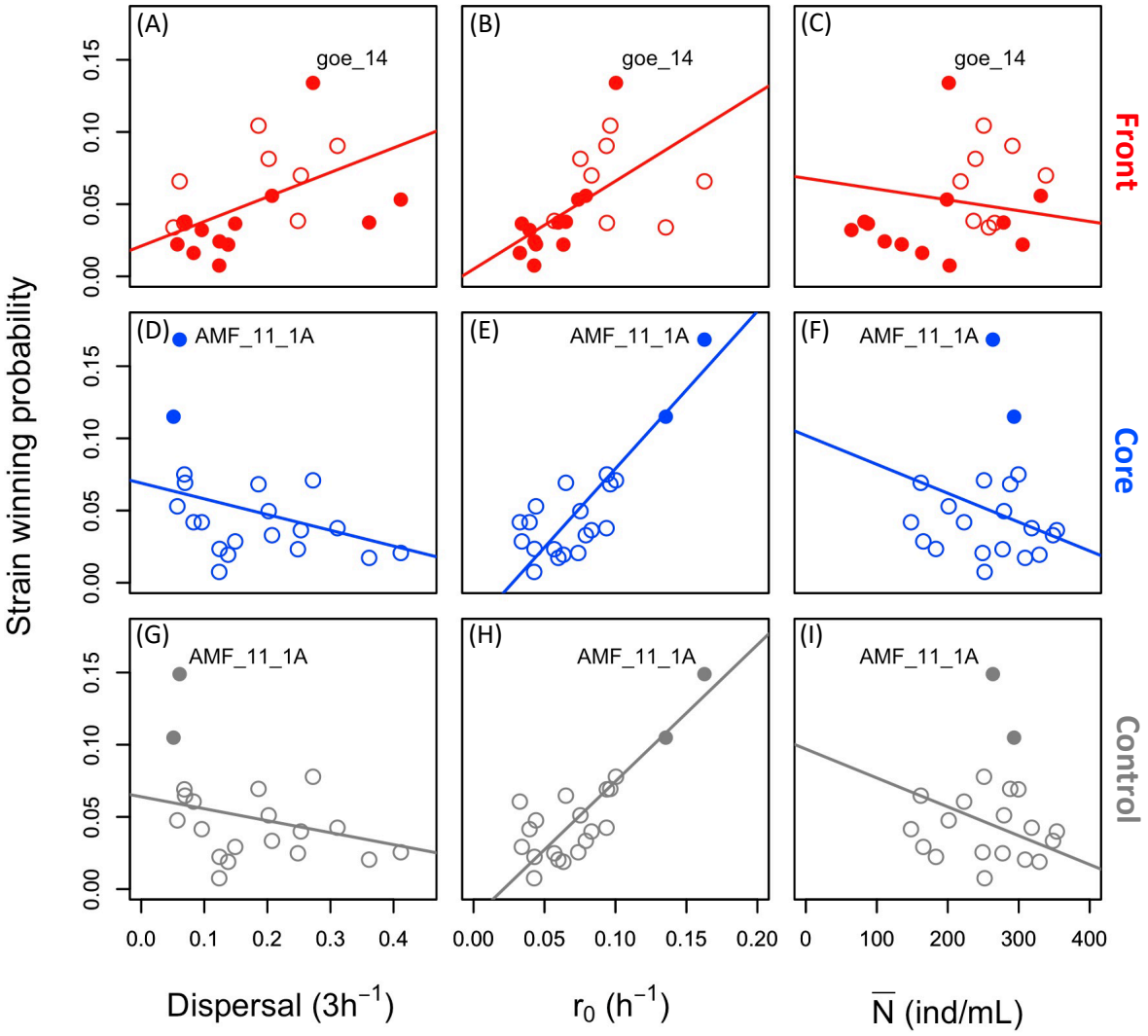


Figure 3 Winning probability (frequency of going to fixation in 10k model runs) of each of 20 strains from the founder population, as a function of its dispersal, growth rate (r_0) and equilibrium density ($\overline{N}$), shown for range front (A-C), range core (D-F) and control (G-I) treatments. Full circles denote the potentially fixed and open circles the eliminated strains, according to genetic analysis (COI genotype). Regression lines obtained from multiple regression models. Different colours represent the different treatments.

Long-term changes

Dispersal

In addition to the short-term evolution (see above), we also observed a long-term increase in dispersal in the range front treatment over the entire time span of the three years of the experiment (cycle x treatment interaction: $\chi_2^2 = 88.8$; $p < 0.001$; Fig. 1). This trend is significant, even when omitting the first 50 cycles ($\chi_2^2 = 51.7$; $p < 0.001$). We found little evidence for a dispersal difference between range core and control lines, neither overall (contrast core vs control: $p > 0.68$) nor when

considering individual cycles (11 cycle-by-cycle contrasts with $0.0078 < p < 0.09$, none significant after correction for multiple testing).

Demographic traits.

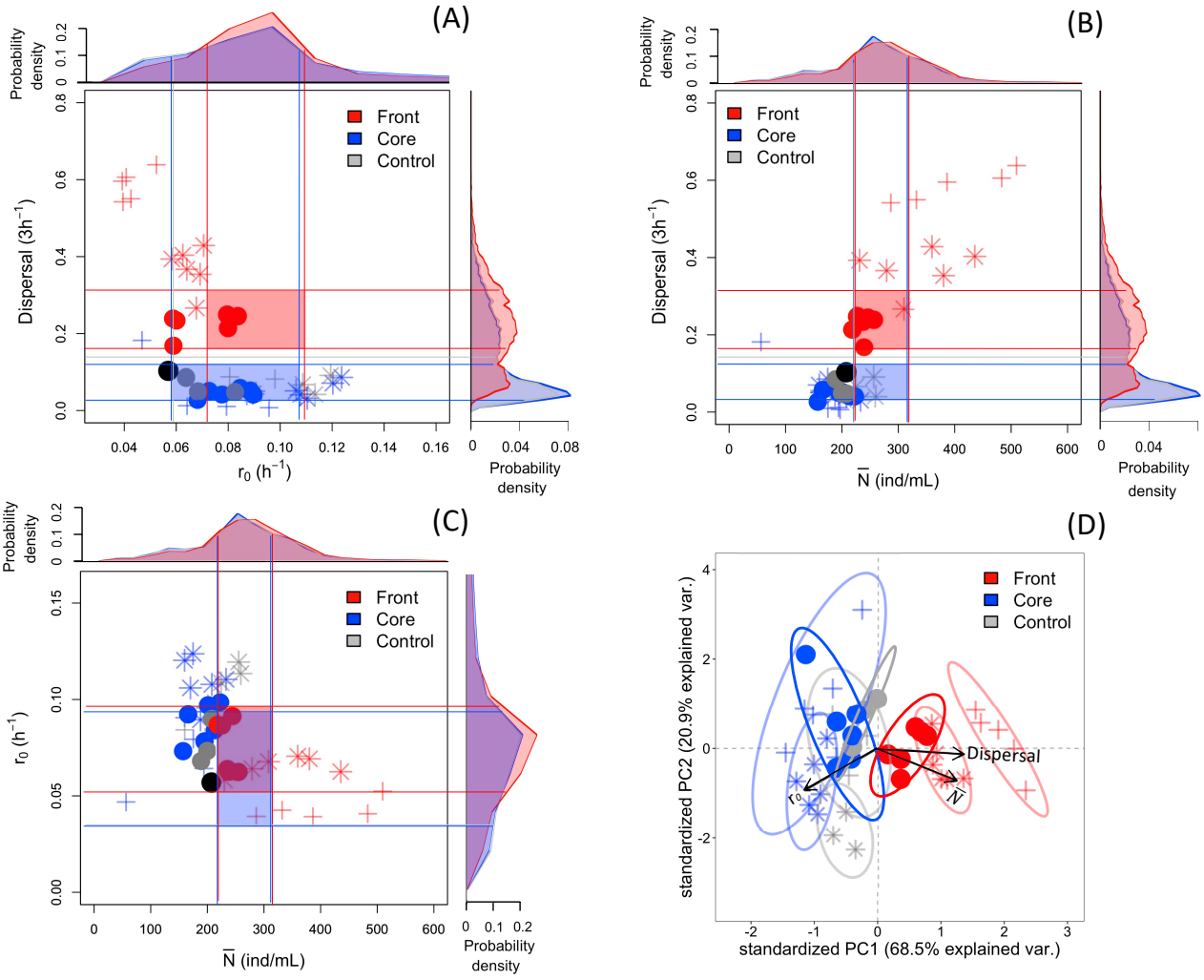
While no significant treatment effects were detected in the first growth assay (cycle 21, see above), range front lines had nearly 2-fold lower values of r_0 than range core lines in assays conducted in year 2 and 3 (year x treatment: $F_4 = 6.66$; $p < 0.001$; Fig. SI 1A). Furthermore, while beginning to grow more slowly, range front lines continued to produce up to 2-fold higher $\overline{N}$ than range core and control lines (treatment: $F_2 = 34.21$; $p < 0.001$; Fig. SI 1B).

Trait associations.

Fig. 4A-C illustrates short- and long-term trends in pairwise trait associations, in relation to the model predictions. For dispersal and r_0 (Fig. 4A), there was no clear relationship between the two traits after short-term selection (year 1). However, in year 2 and 3, observed data points tend to fall outside the main predicted ranges, and a negative relationship between dispersal and r_0 emerged (Fig. 4A). This negative association is highly significant over all lines and years combined ($r = -0.627$, 95% CI [-0.771; -0.434]), but also holds for year 2 and 3 separately (Fig. S3). The positive relationship between dispersal and $\overline{N}$, already observed as a short-term trend, further consolidated in year 2 and 3 (Fig. 4B), again with values mostly falling outside the main predicted short-term ranges. The correlation is significant overall ($r = 0.599$, 95% CI [0.347; 0.725]), as well as for each year separately (Fig. S3). Furthermore, diverging trends in core and front selection lines lead to a negative association between r_0 and $\overline{N}$ (Fig 4C). The negative correlation is of intermediate effect size overall ($r = -0.325$, 95% CI [-0.575; -0.031]), and is significant in all three years separately (Fig. S3). It should be noted that all of these main trends of divergence hold, when we correct for year effects, by expressing front and core line data relative to the control treatment in each year (Fig. S5).

Principal Component Analysis (PCA, Fig. 4D) summarises the patterns of phenotypic divergence. Demography-related traits and dispersal are pulling in approximately equal strength on PC axis 1, but in opposite directions (PC 1 loadings: $r_0 = -0.53$; $\overline{N} = +0.57$; dispersal = +0.62). Thus, range front lines are characterised by a combination of higher equilibrium density and dispersal, but lower intrinsic population growth rate relative to range-core and control lines (MANOVA: $F_{2,37} = 10.85$, $p <$

363 0.001). The separation of clouds indicates the progressive divergence through time, with a
364 maximum in year 3. There is little differentiation between range core and control treatments.



365

366 **Figure 4** Short- and long-term traits associations observed in the experiment, in relation to short-term predictions in the
367 model. (A-C) Bivariate correlations between dispersal, growth rate (r_0) and equilibrium density ($\bar{N}$). Circles are the
368 average values measured for each experimental selection line in year 1 (cycles 15-25; "short term"). Stars refer to year
369 2 (cycles 74-84) and cross symbols to year 3 (cycles 154-161). From the distributions of the model predictions (outer
370 part of graphs), the central range of each trait (50% high density probability interval, HDPI; thin lines) can be defined;
371 the overlap zones of the HDPI (shaded square areas) represent the predicted trait association for each treatment, after
372 short-term evolution. Observations falling outside of the overlap areas indicate deviation from the model, possibly due
373 to de novo evolution (year 2 and 3). The black circles represent the mean ancestral trait association (founder
374 population). (D) Principal Component Analysis (PCA) of all three traits combined according to the first two principal
375 components of the PCA. The arrow length represents the loading value of the trait, while opposite arrow direction
376 indicates opposite trend between traits. Different symbols correspond to the different years (circles, year 1; stars, year
377 2; cross, year 3). The ellipses are the 95 % containment probability region per treatments and year. Different colours
378 represent the different treatments.

Discussion

Predicting range expansions with ecology and evolution occurring on the same timescale is a challenging task. Building on previous ecological range expansion studies (Melbourne & Hastings, 2009; Giometto *et al.*, 2014), we included short-term evolution in a simple model parameterised for our laboratory system and confronted predicted evolutionary outcomes with results from experimental range expansions. Both model and experiment show rapid divergence between range core and front treatments, with selection for higher dispersal at the front. The repeated fixation of particular COI genotypes in the experimental lines corresponded to strains identified as most likely winners in the model. This match between predicted and observed outcomes suggest a certain predictability of range expansions, even when evolutionary change occurs. Over longer time scales, experimental range core and front populations continued to diverge, indicating *de novo* evolution and resulting in the emergence of dispersal syndromes.

Dispersal and growth rate are main targets of selection

In the context of reaction-diffusion models, dispersal (diffusion) and population growth at low densities are the two key traits for understanding and predicting range expansion dynamics (Fisher, 1937; Kolgomorov et al. 1937). Consistent with this view and previous studies (Phillips *et al.*, 2010; Shine *et al.*, 2011), dispersal and population growth were here identified as main targets of selection.

Higher dispersal was immediately selected from standing genetic variation at the range front and weakly selected against in the range core in the model as well as in the experiment, where range front populations showed increased dispersal already after the first few cycles. Such strong and fast selection on dispersal in the vanguard front populations has been found in similar experiments (Fronhofer & Altermatt, 2015; Williams *et al.*, 2016; Ochocki & Miller, 2017; Szűcs *et al.*, 2017; Weiss-Lehman *et al.*, 2017; Petegem *et al.*, 2018), but also in natural populations (Phillips at al. 2006; Perkins et al. 2013). Dispersal evolution might therefore accelerate the speed of range expansion already over very short time scales (Ochoki & Miller, 2017; Miller et al. 2020).

Contrary to more standard views of range expansion with r- and K-selection (Charlesworth, 1971; Roughgarden, 1971; Burton *et al.*, 2010), growth rate was under positive short-term selection in both range core and front treatments. This can be explained by the fact that populations in all

411 treatments experienced regular bottlenecks, thus imposing general selection for increased growth
412 rate, a trait for which there was ample variation among founder strains (Fig. S4 A-C). Importantly,
413 however, our model shows that dispersal and growth rate can be simultaneously selected in the
414 range front treatment (Fig. 3). Whether one or the other trait has more weight depends on the
415 stochasticity introduced by the dispersal bottlenecks, implemented in the model via the quasi-
416 extinction threshold. With small bottlenecks, even weak dispersers make it into the new patch and
417 can subsequently regrow to high density. Indeed, additional model scenarios show that when we
418 decrease the quasi-extinction threshold, selection for growth rate overrides selection for dispersal
419 and the strain with the highest growth rate becomes fixed in all treatments (Section S8). However,
420 the model scenario that fits the observed data indicates a large enough extinction threshold in our
421 experiment, putting equal selective weight on dispersal and growth rate (Fig. 2) and allowing
422 selection to pick the best possible disperser strain that still has a high growth rate.
423
424 **Predictability of outcomes**
425 Genetic analysis indicates that the experimental selection lines became fixed for single COI
426 genotypes. Despite limited resolution (several strains have the same COI), there is a good
427 correspondence with model predictions: Range core and control treatments were fixed for the b05
428 COI genotype, and the two b05 strains in the founder population were the most likely winners in
429 the model, due to their particularly high growth rate (Table S1). In the range front treatment, there
430 is more uncertainty (12 strains carry the b07 genotype fixed in this treatment), but among the
431 possible candidates only the predicted most likely winning strain (goe_14) has both high dispersal
432 and high growth rate (Fig. 3). Additional sequencing would be required to determine whether these
433 selection lines are fixed for the same or different (combinations of) strains.
434
435 Although our model seems to correctly identify the most likely winner strains, it nonetheless
436 predicts the frequent fixation of strains with alternative COI genotypes (Fig. 3). Indeed, according to
437 the model, our exclusive finding of b05 strains in all 9 range core and control lines is highly unlikely
438 ($0.28^9 < 0.0001$). Similarly, even in the front treatment, the expected probability of exclusive fixation
439 of b07 strain(s) is well below 5% ($0.48^6 = 0.01$). In this sense, our experiment was more deterministic
440 than the model. Possibly, when we determined dispersal and growth of the individual strains, a large
441 measurement error was added to biologically relevant variation (Fig. S4). This additional noise then
442 cascades through the model, from the strain posterior distributions (making them wider) to the

443 phenotypic composition of the founder population (making strains more similar) to model outcomes
444 (making them more variable). Alternatively, our model may be missing additional factors, such as
445 direct strain-strain interactions or density-dependent dispersal, which potentially amplify among-
446 strain variation in performance. Regardless, one main conclusion from this model-experiment
447 confrontation is that evolution can be fairly predictable, at least in the short term. As already shown
448 for ecological models (Giometto *et al.*, 2014), realistic predictions can indeed be made about range
449 expansion dynamics, at least in highly controlled laboratory settings. Here we infer trait change from
450 knowledge of standing genetic variation in only a few parameters, suggesting that such models can
451 be readily extended to include evolution.
452
453 **Long term evolution of dispersal syndromes and emergence of trade-offs**
454 Experimental evolution studies show that adaptation to novel conditions may reduce performance
455 in other environments (Kassen, 2014). The emergence of such trade-offs depends on underlying
456 biochemical and life-history constraints (Walsh & Blows, 2009), but also on historical contingency,
457 determining the composition and genetic architecture of the ancestral population, and thus the
458 available trait space for selection to act on.
459
460 In our case, short-term selection from standing genetic variation did not seem to produce clear
461 trade-offs. In the long run, however, range front and core populations continued to diverge in
462 multiple traits (Fig. 4D), and the increase in dispersal in the front treatment was associated with a
463 decrease in growth rate (Fig. 4A). Such coupled responses in dispersal and life-history traits are
464 referred to as dispersal syndrome (Clobert *et al.*, 2012; Cote *et al.*, 2017). Typically, they involve the
465 emergence of a competition-colonisation trade-off, where dispersal evolution coincides with
466 selection for opportunistic growth strategies (r-selection). Theoretical and empirical studies have
467 demonstrated the importance of dispersal syndromes in generating eco-evolutionary feedbacks and
468 accelerating the pace of range expansions and biological invasions (Burton *et al.*, 2010; Perkins *et*
469 *al.*, 2013; Ochocki *et al.*, 2019; Miller *et al.*, 2020).
470
471 Dispersal - growth trade-offs were previously reported for this (Zilio *et al.*, 2020) and another ciliate
472 species (Fronhofer & Altermatt 2015). In these systems, growth rate is a good indicator of
473 competitive ability, and the trade-off with dispersal likely reflects a true life-history constraint,
474 mediated through energy costs of foraging activity (Fronhofer & Altermatt 2015).

The evolved differences between core and front lines are stable, even after switching core and front treatments for multiple cycles (Fig. S6.1). Moreover, mixes of core and front lines readily respond to dispersal selection (Fig. S6.2), making new selection experiments possible, where phenotypic measurements change can easily be combined with the tracking of COI genotype frequencies.

Advantages and limitations of an asexual reproduction scenario

In this study, we consider asexual reproduction in both model and experiment. Hence advantageous allele combinations are not broken apart or reshuffled by sex and recombination (Otto, 2009; Lehtonen *et al.*, 2012), such that strains with favourable trait combinations rapidly increase in frequency in our range and core treatments. Similar results were reported for experimental range expansions of the plant *Arabidopsis thaliana*, where the fastest-dispersing clonal genotype became predominant in multiple replicate selection lines, all starting from the same initial mix of clones (Williams *et al.*, 2016). Thus, asexual reproduction narrows down the variability in the range expansion outcomes and, as we show here, makes predictions possible with relatively simple models.

Clearly, recombination will make predictions more difficult, and replicated range expansion experiments with sexually reproducing organisms already showed higher variability and uncertainty in final outcomes (Ochocki & Miller, 2017; Weiss-Lehman *et al.*, 2017; Petegem *et al.*, 2018). For example, recombination may slow down range expansions in the short term, but speed up longer-term responses by creating novel trait associations not previously available. In our system, sex may have immediate and strong fitness consequences due to the nuclear dimorphism typical of all ciliates. Aside from creating novel genetic variants (in the germline micronucleus), sexual reproduction also involves the recreation of a new somatic macronucleus and thereby the loss of any (somatic) adaptation acquired during asexual life (Verdonck *et al.*, 2021).

Conclusions

Predicting evolution is arduous because of the intrinsic tension between determinism and contingency (Blount *et al.*, 2018), and it demands an adequate theoretical representation of the eco-evolutionary processes in the biological system in question and reliable information on the genetic variation in the relevant traits (Nosil *et al.*, 2020), as we describe in this work. At least in

simple settings as ours, accurate predictions of the evolutionary outcomes of range expansions require surprisingly few parameters, and independent biological realisations can be highly repeatable. Future studies will need to consider, for example, more realistic landscape scenarios and interactions with other species occurring during range expansion. This would imply a more systems-biology approach, with simulations calibrated on the empirical knowledge of the specific ecological scenario and biological players (Papp *et al.*, 2011). More generally, increasing our capacities to make reliable quantitative predictions of invasive eco-evolutionary processes is critical to a variety of issues, from conservation and biocontrol strategies to antibiotic development and disease management.

Acknowledgements

GZ was supported by a grant from the Agence Nationale de Recherche (n° ANR-20-CE02-0023-01) to OK. This is publication ISEM-XXXX-XXX of the Institut des Sciences de l'Evolution. Florent Deshors and Thomas Teissedre assisted with movement and growth assays. Alison Duncan, Flore Zelé and Alexey Potekhin gave helpful comments for the experimental set up and the interpretation of the results.

Author Contributions section

OK conceived the study. OK and CGB performed the experimental work. SK performed the molecular analyses. EAF built the model. GZ, EAF and OK performed the statistical analysis and interpreted the results. GZ, EAF and OK wrote the first draft of the manuscript and all authors commented on the final version.

Conflict of Interest statement

The authors have no conflict of interest to declare.

Data availability statement

The data and model will be made available upon potential acceptance (via Dryad/Figshare repository).

References

Alford, R.A., Brown, G.P., Schwarzkopf, L., Phillips, B.L. & Shine, R. 2009. Comparisons through time and space suggest rapid evolution of dispersal behaviour in an invasive species. *Wildl. Res.* **36**: 23.

Barth, D., Krenek, S., Fokin, S.I. & Berendonk, T.U. 2006. Intraspecific genetic variation in Paramecium revealed by mitochondrial cytochrome C oxidase I sequences. *J Eukaryot Microbiol* **53**: 20–25.

Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**: 1–48.

Blount, Z.D., Lenski, R.E. & Losos, J.B. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**: eaam5979.

Bonte, D., Dyck, H.V., Bullock, J.M., Coulon, A., Delgado, M., Gibbs, M., *et al.* 2012. Costs of dispersal. *Biological Reviews* **87**: 290–312.

Brown, G.P., Phillips, B.L. & Shine, R. 2014. The straight and narrow path: the evolution of straight-line dispersal at a cane toad invasion front. *Proceedings of the Royal Society B: Biological Sciences* **281**: 20141385. Royal Society.

Burton, O.J., Phillips, B.L. & Travis, J.M.J. 2010. Trade-offs and the evolution of life-histories during range expansion: Evolution during range expansion. *Ecology Letters* **13**: 1210–1220.

Calcagno, V., Mouquet, N., Jarne, P. & David, P. 2006. Coexistence in a metacommunity: the competition-colonization trade-off is not dead. *Ecol Lett* **9**: 897–907.

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**: 1–32.

Charlesworth, B. 1971. Selection in Density-Regulated Populations. *Ecology* **52**: 469–474.

Chuang, A. & Peterson, C.R. 2016. Expanding population edges: theories, traits, and trade-offs. *Glob Chang Biol* **22**: 494–512.

564 Clobert, J., Baguette, M., Benton, T.G. & Bullock, J.M. 2012. *Dispersal ecology and evolution*.
565 Oxford University Press.

566 Cote, J., Bestion, E., Jacob, S., Travis, J., Legrand, D. & Baguette, M. 2017. Evolution of dispersal
567 strategies and dispersal syndromes in fragmented landscapes. *ECOGRAPHY* **40**: 56–73.

568 Fellous, S., Duncan, A., Coulon, A. & Kaltz, O. 2012. Quorum Sensing and Density-Dependent
569 Dispersal in an Aquatic Model System. *PLoS ONE* **7**: e48436.

570 Fisher, R.A. 1937. The wave of advance of advantageous genes. *Annals of Eugenics* **7**: 355–369.

571 Friedenberg, N.A. 2003. Experimental evolution of dispersal in spatiotemporally variable
572 microcosms. *Ecology Letters* **6**: 953–959.

573 Fronhofer, E.A. & Altermatt, F. 2015. Eco-evolutionary feedbacks during experimental range
574 expansions. *Nature Communications* **6**.

575 Fronhofer, E.A., Legrand, D., Altermatt, F., Ansart, A., Blanchet, S., Bonte, D., *et al.* 2018. Bottom-
576 up and top-down control of dispersal across major organismal groups. *Nature Ecology &*
577 *Evolution* **2**: 1859–1863. Nature Publishing Group.

578 Giometto, A., Rinaldo, A., Carrara, F. & Altermatt, F. 2014. Emerging predictable features of
579 replicated biological invasion fronts. *PNAS* **111**: 297–301.

580 Hughes, C.L., Dytham, C. & Hill, J.K. 2007. Modelling and analysing evolution of dispersal in
581 populations at expanding range boundaries. *Ecological Entomology* **32**: 437–445.

582 Kassen, R. 2014. *Experimental Evolution and the Nature of Biodiversity*. Macmillan Learning.

583 Killeen, J., Gougat-Barbera, C., Krenek, S. & Kaltz, O. 2017. Evolutionary rescue and local
584 adaptation under different rates of temperature increase: a combined analysis of changes
585 in phenotype expression and genotype frequency in Paramecium microcosms. *Molecular*
586 *Ecology* **26**: 1734–1746.

587 Kolmogorov AN, Petrovskii IG, Piskunov NS (1937). A study of the diffusion equation with increase
588 in the amount of substance, and its application to a biological problem.

589 Kubisch, A., Holt, R.D., Poethke, H.-J. & Fronhofer, E.A. 2014. Where am I and why? Synthesizing
590 range biology and the eco-evolutionary dynamics of dispersal. *Oikos* **123**: 5–22.

591 Lehtonen, J., Jennions, M.D. & Kokko, H. 2012. The many costs of sex. *Trends in Ecology &*
592 *Evolution* **27**: 172–178. Elsevier.

593 Leotard, G., Debout, G., Dalecky, A., Guillot, S., Gaume, L., Mckey, D., *et al.* 2009. Range Expansion
594 Drives Dispersal Evolution In An Equatorial Three-Species Symbiosis. *PLOS ONE* **4**.

595 Livingston, G., Matias, M., Calcagno, V., Barbera, C., Combe, M., Leibold, M.A., *et al.* 2012.
596 Competition–colonization dynamics in experimental bacterial metacommunities. *Nat*
597 *Commun* **3**: 1234.

598 Lombaert, E., Estoup, A., Facon, B., Joubard, B., Grégoire, J.-C., Jannin, A., *et al.* 2014. Rapid
599 increase in dispersal during range expansion in the invasive ladybird *Harmonia axyridis*.
600 *Journal of Evolutionary Biology* **27**: 508–517.

601 Lubina, J.A. & Levin, S.A. 1988. The Spread of a Reinvading Species: Range Expansion in the
602 California Sea Otter. *The American Naturalist* **131**: 526–543. The University of Chicago
603 Press.

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

606 Miller, T.E.X., Angert, A.L., Brown, C.D., Lee-Yaw, J.A., Lewis, M., Lutscher, F., *et al.* 2020. Eco-
607 evolutionary dynamics of range expansion. *Ecology*, doi: 10.1002/ecy.3139.

608 Nidelet, T. & Kaltz, O. 2007. Direct and correlated response to selection in a host-parasite system:
609 testing for the emergence of genotype specificity. *Evolution* **61**: 1803–1811.

610 Nørgaard, L.S., Zilio, G., Saade, C., Gougat-Barbera, C., Hall, M.D., Fronhofer, E.A., *et al.* 2021. An
611 evolutionary trade-off between parasite virulence and dispersal at experimental invasion
612 fronts. *Ecology Letters* ele.13692.

613 Nosil, P., Flaxman, S.M., Feder, J.L. & Gompert, Z. 2020. Increasing our ability to predict
614 contemporary evolution. *Nat Commun* **11**: 5592.

615 Ochocki, B.M. & Miller, T.E.X. 2017. Rapid evolution of dispersal ability makes biological invasions
616 faster and more variable. *Nat Commun* **8**: 1–8.

617 Ochocki, B.M., Saltz, J.B. & Miller, T.E.X. 2019. Demography-Dispersal Trait Correlations Modify the
618 Eco-Evolutionary Dynamics of Range Expansion. *The American Naturalist* **195**: 231–246.
619 The University of Chicago Press.

620 Otto, S.P. 2009. The evolutionary enigma of sex. *Am Nat* **174 Suppl 1**: S1–S14.

621 Papp, B., Notebaart, R.A. & Pál, C. 2011. Systems-biology approaches for predicting genomic
622 evolution. *Nat Rev Genet* **12**: 591–602.

623 Perkins, T.A., Phillips, B.L., Baskett, M.L. & Hastings, A. 2013. Evolution of dispersal and life history
624 interact to drive accelerating spread of an invasive species. *Ecol. Lett.* **16**: 1079–1087.

625 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., *et al.* 2015. The
626 ecological forecast horizon, and examples of its uses and determinants. *Ecol Lett* **18**: 597–
627 611.

628 Petegem, K.V., Moerman, F., Dahirel, M., Fronhofer, E.A., Vandegehuchte, M.L., Leeuwen, T.V., *et*
629 *al.* 2018. Kin competition accelerates experimental range expansion in an arthropod
630 herbivore. *Ecology Letters* **21**: 225–234.

631 Phillips, B.L., Brown, G.P. & Shine, R. 2010. Life-history evolution in range-shifting populations.
632 *Ecology* **91**: 1617–1627.

633 Phillips, B.L., Brown, G.P., Travis, J.M.J. & Shine, R. 2008. Reid's paradox revisited: the evolution of
634 dispersal kernels during range expansion. *Am Nat* **172 Suppl 1**: S34-48.

635 Phillips, B.L., Brown, G.P., Webb, J.K. & Shine, R. 2006. Invasion and the evolution of speed in
636 toads. *Nature* **439**: 803–803.

637 R Core Team. (2021). R: a language and environment for statistical computing (Version 4.1.2).
638 Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-
639 project.org/

640 Rosenbaum, B., Raatz, M., Weithoff, G., Fussmann, G.F. & Gaedke, U. 2019. Estimating Parameters
641 From Multiple Time Series of Population Dynamics Using Bayesian Inference. *Front. Ecol.*
642 *Evol.* **6**.

643 Roughgarden, J. 1971. Density-Dependent Natural Selection. *Ecology* **52**: 453–468.

644 Shine, R., Brown, G.P. & Phillips, B.L. 2011. An evolutionary process that assembles phenotypes
645 through space rather than through time. *Proceedings of the National Academy of Sciences*
646 **108**: 5708–5711.

647 Simmons, A.D. & Thomas, C.D. 2004. Changes in dispersal during species' range expansions. *Am.*
648 *Nat.* **164**: 378–395.

649 Szűcs, M., Vahsen, M.L., Melbourne, B.A., Hoover, C., Weiss-Lehman, C. & Hufbauer, R.A. 2017.
650 Rapid adaptive evolution in novel environments acts as an architect of population range
651 expansion. *PNAS* **114**: 13501–13506.

652 Thomas, C.D., Bodsworth, E.J., Wilson, R.J., Simmons, A.D., Davies, Z.G., Musche, M., *et al.* 2001.
653 Ecological and evolutionary processes at expanding range margins. **411**: 6.

654 Verdonck, R., Legrand, D., Jacob, S. & Philippe, H. 2021. Phenotypic plasticity through disposable
655 genetic adaptation in ciliates. *Trends in Microbiology* S0966842X21001396.

656 Volpert, V. & Petrovskii, S. 2009. Reaction-diffusion waves in biology. *Phys Life Rev* **6**: 267–310.

657 Walsh, B. & Blows, M.W. 2009. Abundant Genetic Variation + Strong Selection = Multivariate
658 Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution,*
659 *and Systematics* **40**: 41–59.

660 Weiler, J., Zilio, G., Zeballos, N., Nørgaard, L., Conce Alberto, W.D., Krenek, S., *et al.* 2020. Among-
661 Strain Variation in Resistance of Paramecium caudatum to the Endonuclear Parasite
662 Holospora undulata: Geographic and Lineage-Specific Patterns. *Front. Microbiol.* **11**:
663 603046.

664 Weiss-Lehman, C., Hufbauer, R.A. & Melbourne, B.A. 2017. Rapid trait evolution drives increased
665 speed and variance in experimental range expansions. *Nat Commun* **8**: 1–7.

Wichterman, R. 1986. *The Biology of Paramecium*. Springer US.

Williams, J.L., Hufbauer, R.A. & Miller, T.E.X. 2019. How Evolution Modifies the Variability of Range Expansion. *Trends in Ecology & Evolution* **34**: 903–913.

Williams, J.L., Kendall, B.E. & Levine, J.M. 2016. Rapid evolution accelerates plant population spread in fragmented experimental landscapes. *Science* **353**: 482–485.

Zilio, G., Nørgaard, L.S., Gougat-Barbera, C., Hall, M.D., Fronhofer, E.A. & Kaltz, O. 2020. Travelling with a parasite: the evolution of resistance and dispersal syndrome during experimental range expansion. *bioRxiv* 2020.01.29.924498.

Zilio, G., Nørgaard, L.S., Petrucci, G., Zeballos, N., Gougat-Barbera, C., Fronhofer, E.A., *et al.* 2021. Parasitism and host dispersal plasticity in an aquatic model system. *J Evol Biol* jeb.13893.

Supplementary Information

S1 Growth characteristics of long-term selection lines: r_0 and $\overline{N}$

In growth assays, intrinsic population growth rate (r_0) was measured after 21 (year 1), 78 (year 2) and 160 (year 3) dispersal / growth cycles. Equilibrium density ($\overline{N}$) was taken for each selection line at the end of the 1-week growth period at each cycle during the long-term experiment. Here we averaged $\overline{N}$ for each line over 9-10 cycles in year 1 (cycle 15-25), year 2 (74-84) and year 3 (154-163).

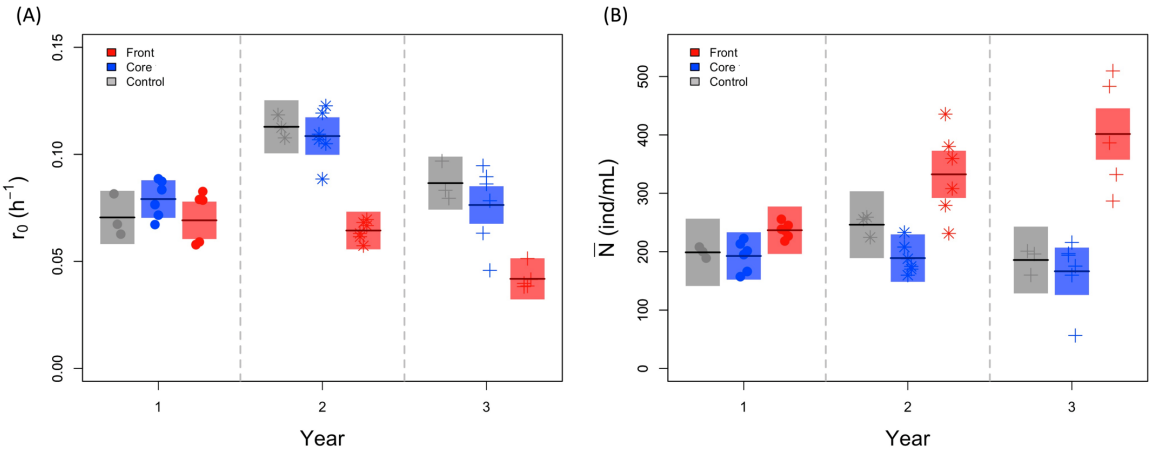


Figure S1 (A) Intrinsic population growth rate (r_0) and (B) mean density at the end of the cycle ($\overline{N}$) from core, front and control treatments (respectively in blue, red and grey). Full symbols represent the mean values for each selection line (n = 15), different symbols (circle, star, cross) refer to the three different years where measurement were taken. Shaded panels show means and 95 % confidence intervals of the model predictions.

Script Beverton-Holt fitting (r_0)

```
################################################################################
rm(list=ls())

# load some packages
library(rstan)
library(deSolve)
library(coda)
library(vioplot)
library(loo)
```

```
721
722  ################################################################################
723  # load data
724  data<- read.table(file="file_name.txt", header=T)
725
726  ################################################################################
727  # declare stan BH model
728  stanmodelcode_BH = '
729    // function that caclulates BH population growth
730  functions{
731    real[] odemodel(real t, real[] N, real[] p, real[] x_r, int[] x_i){
732      // p[1]=r0, p[2]=d, p[3]=K
733      real dNdt[1];
734      dNdt[1] = ((p[1] + p[2])/(1 + ((p[1]/(p[3]*p[2])) * N[1])) - p[2])*N[1];
735      return dNdt;
736    }
737  }
738
739  data{
740    int n;
741    real log_N0;
742    real log_N[n];
743    real t0;
744    real t[n];
745  }
746
747  transformed data {
748    // not used here
749    real x_r[0];
750    int x_i[0];
751  }
752
```

```
753  parameters{
754    real log_r;
755    real log_d;
756    real log_K;
757    real log_N0sim;
758    real<lower=0> sdev;
759  }
760
761
762  transformed parameters{
763    // all of this was in the model section previously
764    // I moved it here to be able to get waic because Nsim needs to be accessible in "generated
765  quantities {}"
766
767    real p[3];
768    real Nsim[n,1]; // simulated values, matrix. dim1 = time, dim2 = dim_ODE = 1
769    real N0sim_dummy[1]; // just a dummy because the ODE solver requires real[] instead of real
770
771    // parameters for integrator
772    p[1] = exp(log_r);
773    p[2] = exp(log_d);
774    p[3] = exp(log_K);
775    N0sim_dummy[1] = exp(log_N0sim);  // see above
776
777    // integrate ODE (maybe try: integrate_ode_bdf()??)
778    Nsim = integrate_ode_rk45(odemodel,N0sim_dummy,t0,t,p,x_r,x_i);
779  }
780
781  model{
782    // priors
783    // note: it can be VERY helpful to estimate parameters on logscale,
784    // especially if they have different orders of magnitude. here it works on regular scale, though.
```

```
  log_r ~ normal(log(0.1), 1.5);
  log_K ~ normal(log(500), 1);
  log_d ~ normal(log(0.1),1.5);
  log_N0sim ~ normal(log(2),1);
  sdev ~ cauchy(0,1);

  // likelihood, normal (maybe lognormal helpful)
  log_N0 ~ normal(log_N0sim,sdev);
  for (i in 1:n){
    log_N[i] ~ normal(log(Nsim[i,1]),sdev);
  }
}

// cacluate log lik to get waic
// from loo R package description
generated quantities {
  real log_lik[n];
  for (nn in 1:n){
    log_lik[nn] = normal_lpdf(log_N[nn] | log(Nsim[nn,1]), sdev);
  }
}
'
################################################################################
# functions
ode.model_BH = function(t,N,p){
  with(as.list(p),{
    dNdt = ((p[1] + p[2])/(1 + ((p[1]/(p[3]*p[2])) * N)) - p[2])*N
    return(list(dNdt))
  })
}

################################################################################
```

```
# stan options
chains = 1
rstan_options(auto_write = TRUE)
options(mc.cores = chains)
iter   =  10000
warmup =  1000
thin   =    1

# compile model
s_model_BH = stan_model(model_code=stanmodelcode_BH)

##############################################################################
 # BH fitting

 # create data object for rstan
 data_BH = list(n = length(act_times)-1,
                log_N0 = log(act_densities[1]),
                log_N  = log(act_densities[2:length(act_times)]),
                t0    = act_times[1],
                t     = act_times[2:length(act_times)])

 # initial values for rK fitting
 init_BH=rep(list(list(log_r=log(0.1),
                       log_K = log(500),
                       log_d = log(0.05),
                       log_N0sim=log(2),
                       sdev=1))
             ,chains)

 # do the rK fit
 fit_BH = sampling(s_model_BH,
                   data=data_BH,
```

```
        iter=iter,
        warmup=warmup,
        thin=thin,
        chains=chains,
        init=init_BH
)

# check model
print(fit_BH)
```

S2 Swimming behaviour

We collected data on swimming behaviour for each selection line in year 1 (cycles 53, 54, 55, 58, 74) and year 2 (cycles 103 and 116), using an automated video analysis pipeline (Fronhofer & Altermatt, 2015; Pennekamp *et al.*, 2015). To this end, 120-µL samples (ca. 20 individuals) from populations at equilibrium were placed on a microscope slide and videos were recorded under a stereomicroscope (Perfex SC38800 camera; settings: frames per second: 15; duration: 5 s; total magnification: 10x). One video per selection line and cycle was recorded, except for cycle 74 (n=4). We analysed the videos using the "bemovi" package (Pennekamp *et al.*, 2015, see script below), which provided estimates of individual *Paramecium* swimming speed and the tortuosity of swimming trajectories, an indicator of changes in the swimming direction (standard deviation of the turning angle distribution). For analysis, averages of swimming speed and tortuosity were calculated for each sample.

Results: *Paramecium* from range front lines had a significantly lower swimming speed (-41%) than those from the range core and control treatments (treatment: $F_{2,12} = 33.5$; $p < 0.001$; Fig. S2). Across selection lines and years combined, swimming speed in the assays was negatively correlated with dispersal rate observed in the selection lines ($r = -0.72$, $n = 30$, $p < 0.0001$), meaning that lines with a higher dispersal generally had a lower swimming speed. Tortuosity of swimming trajectories were negatively correlated with swimming speed (all replicates: $r = -0.29$, $n=135$, $p = 0.0005$), but did not significantly differ among treatments ($p > 0.545$).

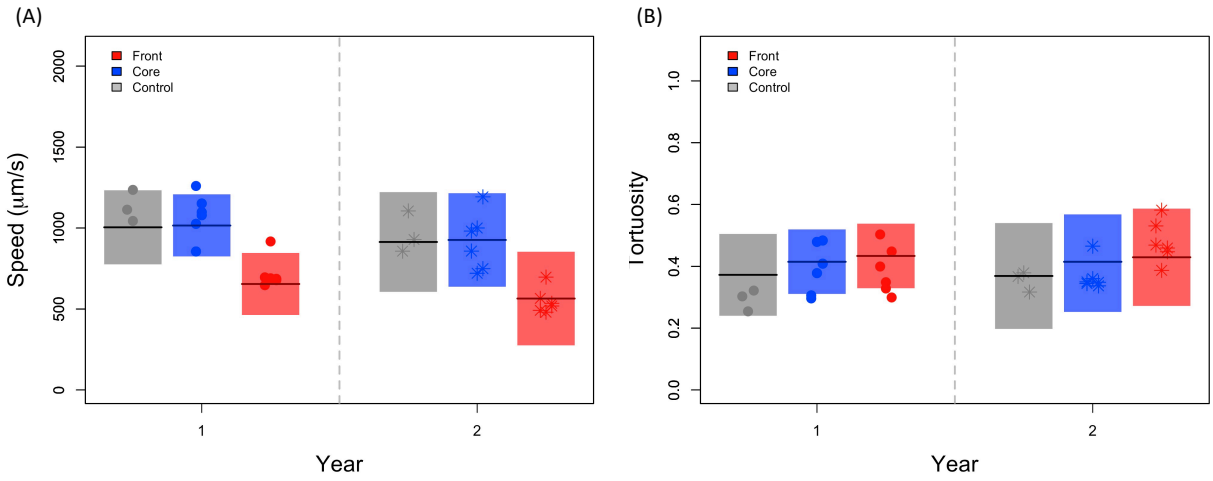


Figure S2 Estimates of the (A) swimming speed and (B) tortuosity of swimming trajectories from core (blue), front (red) and control (grey) treatments. Full symbols represent the mean values for each selection line (*n* = 15), with different symbols (point, star) corresponding to the first two years of the study. No measurements were taken in the year 3. Shaded panels show means and 95 % confidence intervals of the model predictions.

References

Fronhofer, E.A. & Altermatt, F. 2015. Eco-evolutionary feedbacks during experimental range expansions. *Nature Communications* **6**.

Pennekamp, F., Schtickzelle, N. & Petchey, O.L. 2015. BEMOVI, software for extracting behavior and morphology from videos, illustrated with analyses of microbes. Ecol. Evol. 5: 2584–2595.

Script video analysis

```
##########################################################################
# R script for analysing video files with BEMOVI (www.bemovi.info)
rm(list=ls())
# load package
library(devtools)
install_github("efronhofer/bemovi", ref="experimental")
library(bemovi)

##########################################################################
# VIDEO PARAMETERS
```

```
# video frame rate (in frames per second)
fps <- 15
# length of video (in frames)
total_frames <- 150
# magnification
# this parameter sets "measured_volume" and "pixel_to_scale" for Perfex Pro 10 stereomicrocope
with #Perfex SC38800 (IDS UI-3880LE-M-GL) camera and sample height = 0.5mm
# if other devices are used, set the two paramneters manually
# possible values: 0.8, 1, 2, 3
magnification <- 1

# specify video file format (one of "avi","cxd","mov","tiff")
# bemovi only works with avi and cxd. other formats are reformated to avi below
video.format <- "avi"
# setup
difference.lag <- 10
thresholds <- c(90,255) # don't change the second value

#############################################################################
# FILTERING PARAMETERS
# min and max size: area in pixels
particle_min_size <- 5
particle_max_size <- 1000

# number of adjacent frames to be considered for linking particles
trajectory_link_range <- 3
# maximum distance a particle can move between two frames
trajectory_displacement <- 30

# these values are in the units defined by the parameters above: fps (seconds),
#measured_volume (microliters) and pixel_to_scale (micometers)
filter_min_net_disp <- 500
```

```
filter_min_duration <- 1
filter_detection_freq <- 0.1
filter_median_step_length <- 5
#############################################################################
# MORE PARAMETERS (USUALLY NOT CHANGED)

# set paths to ImageJ and particle linker standalone
IJ.path <- "/home/user_name /bin/ImageJ"
to.particlelinker <- "/home/user_name/bin/ParticleLinker"

# directories and file names
to.data <- paste(getwd(),"/",sep="")
video.description.folder <- "0_video_description/"
video.description.file <- "video_description.txt"
raw.video.folder <- "1_raw/"
particle.data.folder <- "2_particle_data/"
trajectory.data.folder <- "3_trajectory_data/"
temp.overlay.folder <- "4a_temp_overlays/"
overlay.folder <- "4_overlays/"
merged.data.folder <- "5_merged_data/"
ijmacs.folder <- "ijmacs/"

# RAM allocation
memory.alloc <- c(60000)

# RAM per particle linker instance (in MB)
memory.alloc.perLinker <- c(3000)

#############################################################################
# VIDEO ANALYSIS

# identify particles
```

```
locate_and_measure_particles(to.data, raw.video.folder, particle.data.folder,
difference.lag, thresholds, min_size = particle_min_size, max_size =
particle_max_size, IJ.path, memory.alloc)

# link the particles
link_particles(to.data, particle.data.folder, trajectory.data.folder, linkrange =
trajectory_link_range, disp = trajectory_displacement, start_vid = 1, memory =
memory.alloc, memory_per_linkerProcess = memory.alloc.perLinker)

# merge info from description file and data
merge_data(to.data, particle.data.folder, trajectory.data.folder,
video.description.folder, video.description.file, merged.data.folder)

# load the merged data
load(paste0(to.data, merged.data.folder, "Master.RData"))

# filter data: minimum net displacement, their duration, the detection
#frequency and the median step length
trajectory.data.filtered <- filter_data(trajectory.data, filter_min_net_disp,
filter_min_duration, filter_detection_freq, filter_median_step_length)

# summarize trajectory data to individual-based data
morph_mvt <- summarize_trajectories(trajectory.data.filtered, calculate.median=F,
write = T, to.data, merged.data.folder)

# get Sample level info
summarize_populations(trajectory.data.filtered, morph_mvt, write=T, to.data,
merged.data.folder, video.description.folder, video.description.file, total_frames)

# create overlays for validation
create_overlays(trajectory.data.filtered, to.data, merged.data.folder,
raw.video.folder, temp.overlay.folder, overlay.folder, 2048, 2048,
```

```
difference.lag, type = "label", predict_spec = F, IJ.path,
contrast.enhancement = 1, memory = memory.alloc)
```

S3 Trait correlations

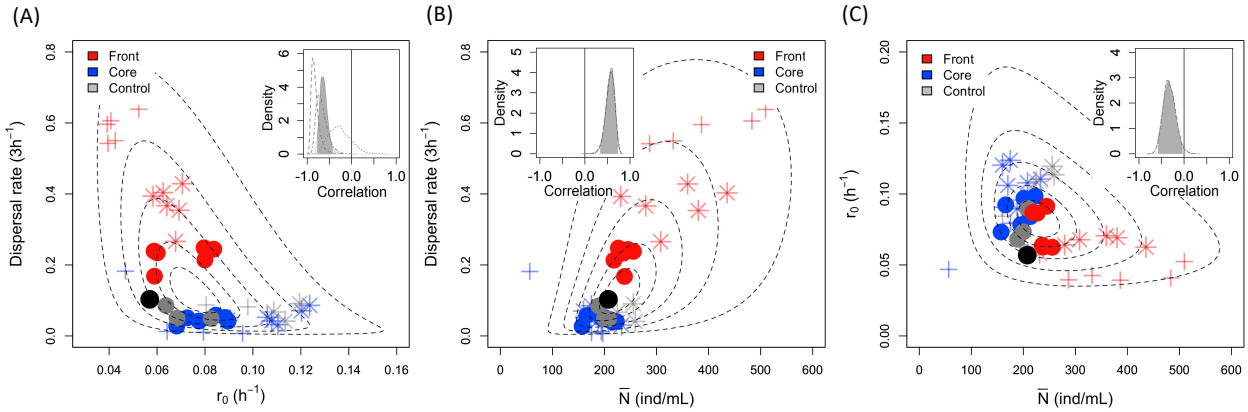


Figure S3 Overall correlation between (A) dispersal - r_0, (B) dispersal - $\overline{N}$ and (C) r_0 - $\overline{N}$ obtained with Bayesian inference. Symbols are the average values for a given selection line and year, with blue, red and grey corresponding to core, front and control treatment, respectively. Different symbols refer to the three different years: circle (year 1), star (year 2), cross (year 3). The black point represents the ancestral values (overall mean) of the founder population. The ellipses are bound to non-linear space and correspond to the 10, 25, 50, 75 an 95 % CI of the correlation of pairs of traits. The shaded areas in the insert panels represent the posterior distribution of the overall correlation coefficients (across selection treatments and years). The dot-dashed lines show the posterior distribution of the year 1 correlation coefficients, dashed lines of the year 2, and dotted lines of the year 3. The black line in the inserts highlights the 0 value, and thus the absence of correlation.

Script correlation analysis

```
##############################################################################
# Stan code
model_simple <- "
// Pearson Correlation
data {
  int<lower=0> n;
  vector[2] x_obs[n];
}
```

```
parameters {
vector[2] mu;
vector<lower=0>[2] lambda;
real<lower=-1,upper=1> r;
}
transformed parameters {
vector<lower=0>[2] sigma;
cov_matrix[2] T;

 // Reparameterization
sigma[1] = sqrt(lambda[1]);
sigma[2] = sqrt(lambda[2]);
 T[1,1] = square(sigma[1]);
T[1,2] = r * sigma[1] * sigma[2];
T[2,1] = r * sigma[1] * sigma[2];
T[2,2] = square(sigma[2]);
}
model {
// Priors
mu ~ normal(0, 10);
lambda ~ normal(0,1);
  // Data
  x_obs ~ multi_normal(mu, T);
}"
# selecting the variable of interest, e.g. trait_1 and trait_2
data <- cbind(dataset$trait_1, dataset$trait_2)
# sampling
samples <- stan(model_code=model_simple,
            data=list(x_obs=data, n = dim(data)[1]),
            init=list(list(r=0, mu=c(0, 0), lambda=c(1, 1))),  # If not specified, gives random inits
            pars= c("r", "mu", "sigma"),
            iter=10000,
```

```
warmup = 2000,
chains=1
)
```

S4 Founder population trait values and correlations

Prior to the start of the long-term experiment, we characterised the 20 founder strains for dispersal and population growth characteristics (r_0, $\overline{N}$), with 3-4 replicates per strain, as described in the main text. Using a Bayesian approach (Rosenbaum *et al.*, 2019), we determined median values and 95% CI for each strain (Table S1, see also section S1). Figure S4 illustrates the (bivariate) trait space occupied by the mix of the strains in the founder population. For example, Fig. S4A shows considerable genotypic variation in both dispersal and r_0. Certain strains have very high r_0 and very low dispersal, and several strains have relatively high dispersal and intermediate levels of r_0. As shown in the main text, these two types of strains are targeted by short-term selection in the range core and front treatments, respectively. There are no strains with very high levels of both dispersal and growth, and such variants also do not seem to evolve in the long term (see Fig. 4), suggesting that this part of the trait is unavailable to the genetic backgrounds used in this experiment.

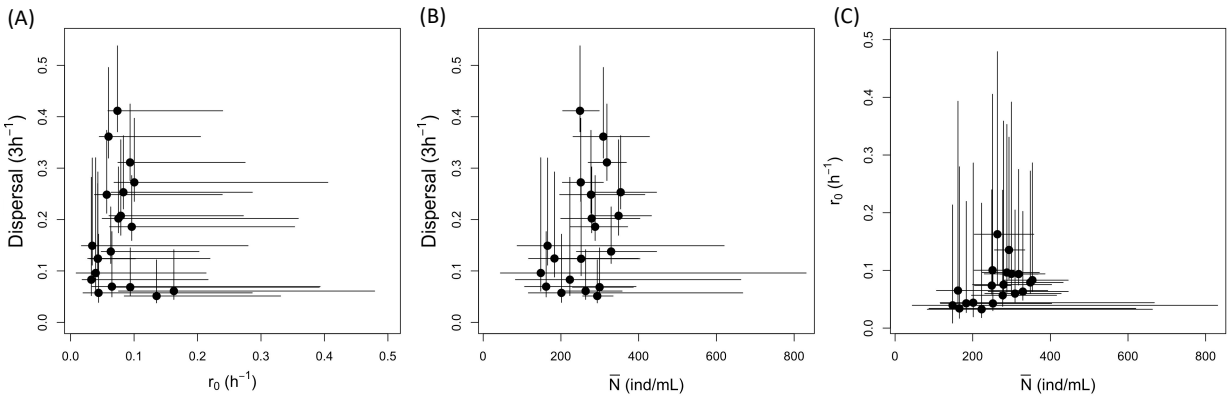


Figure S4 Trait relationships in the base population (mix of the 20 founder strains) for (A) dispersal - r_0, (B) dispersal - $\overline{N}$ and (C) r_0 - $\overline{N}$. Each point represents the median trait values of a strain with the 95% CI (see also Table S1).

References

Rosenbaum, B., Raatz, M., Weithoff, G., Fussmann, G.F. & Gaedke, U. 2019. Estimating Parameters From Multiple Time Series of Population Dynamics Using Bayesian Inference. Front. Ecol. Evol. 6.

S5 Trait relationships relative to the control treatment

In the main text, Figure 4A-C illustrates short- and long-term trends in pairwise trait associations, in relation to the model predictions. The long-term trends are inferred from the comparison of measurements taken at different time points (year 1, 2, 3, see main text), and we can therefore not a priori exclude the possibility that the (evolutionary) change in a trait is confounded with a measurement year effect. Ideally, to avoid this problem, samples would be frozen each year and all samples measured at the same time in a single assay at the end of the long-term experiment. However, freezing of samples was not possible for our lines. Instead, we accounted for potential year effects by expressing the performance of range core and range front lines relative to the control treatment. This was done by subtracting the means of the control lines from the values of individual core or front lines from the same year. Patterns for these standardized trait associations (Fig. S5) are very similar to those shown in Fig. 4A-C. Thus, our main conclusions regarding the divergence of selection lines were unlikely to be affected by measurement year effects.

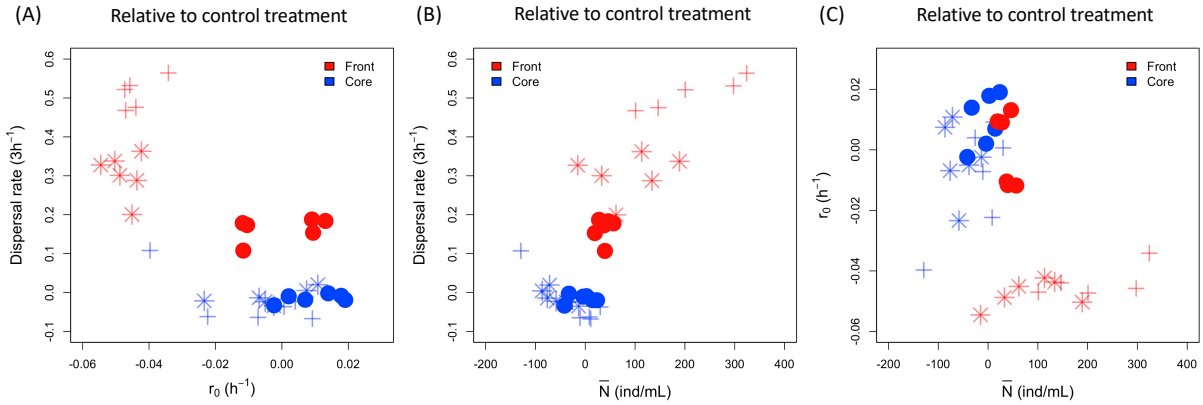


Figure S5 Trait relationships between (A) dispersal - r_0, (B) dispersal - $\overline{N}$ and (C) r_0 - $\overline{N}$, expressed relative to the control treatment (core/front minus control values), for each of three years. Symbols are the average values for each selection line. Negative values correspond to decreased trait values compared to the control treatment of the same year, positive indicated increased values and 0 corresponds to no changes. Different symbols refer to the three different years: circle (year 1), star (year 2), cross (year 3).

S6 Complementary experiments

After the long-term experiment was completed, additional tests were performed with the evolved selection lines. First, we performed a 'treatment-reversal' experiment. To this end, we divided the evolved lines in two new replicates. The first replicate was continued in the original treatment, whereas the second replicate was subjected to the other (opposite) treatment. This experiment was run for 9 cycles and dispersal measured, following the protocols described in the main text.

Results: Range front lines continued to show higher dispersal than range core lines, when switched to 'range core' selection conditions (Fig. S6.1A). Conversely, range core lines continued show very low dispersal, when switched to 'range front' selection conditions (Fig. S6.1B).

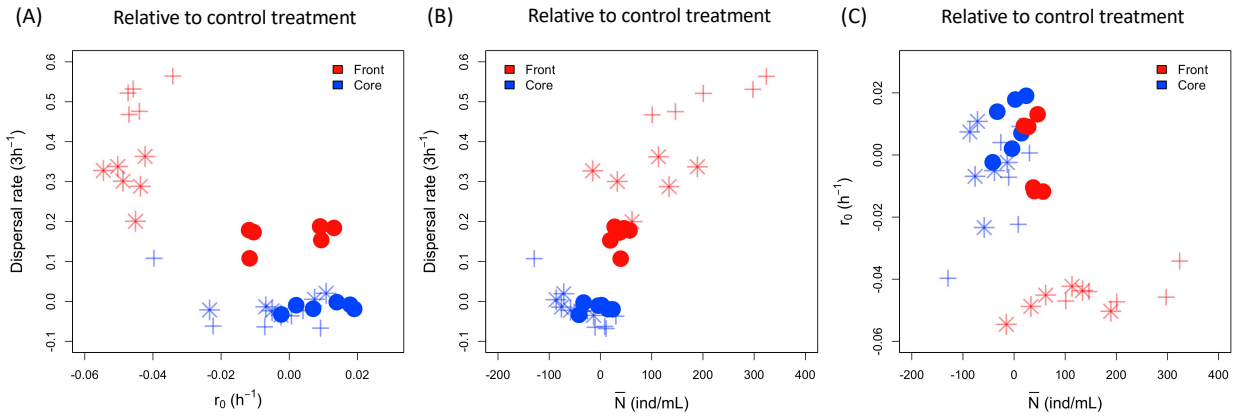


Figure S6.1 Treatment-reversal experiment. (A) Front and core evolved lines exposed for 9 cycles to core treatment. (B) Front and core evolved lines exposed for 9 cycles to front treatment.

Secondly, we wanted to test whether the observed treatment effects in the long-term experiment, as measured in single-line assays, were strong enough to be picked up by selection. To this end, we mixed range core and front lines at different 'initial' proportions, and then exposed these mixes (together with pure 100% core and front controls) to a range core or range front treatment for 3 cycles.

Results: Under range core selection (Fig. S6.1A), we find a decrease in dispersal in the mixes, reaching levels as low as those observed for pure range core lines, whereas pure front lines continue to show high dispersal. Conversely, under range front selection (Fig. S6.1B), we find an increase in dispersal in the mixes, reaching levels comparable to values observed for the pure range front lines. Pure range core lines also show an increase in dispersal, but still disperse less than the mixes or the pure front lines. These results indicate a match between selection history and selection treatment, meaning that front lines have a selective advantage under front selection and core lines under core selection. Indeed, in similar experiments (F. Manzi & O. Kaltz, unpublished), we find that such observed phenotypic changes go hand in hand with the fixation of range core or front lines.

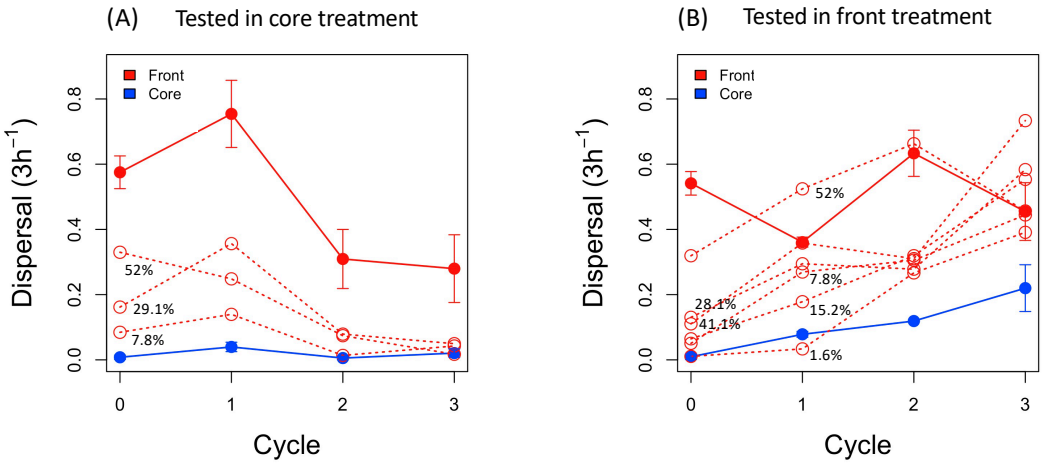


Figure S6.2 Mixed-lines experiment, with (A) range core selection or (B) range front selection treatment. Solid connecting lines are pure (100%) evolved core and front selection lines, respectively. Dashed red lines are mixes of core and front lines, with the initial proportion of front lines ranging from 1.6% to 52%. Each dashed line represents a single experimental 'mixed' replicate, the solid 'pure' lines represent averages (± SE) over 3 experimental replicates.

S7 Strain winning probability

The model predictions show strong variation in the winning probability among the 20 strains, i.e. the probability of strains fixation in the population for each of three treatments at the end of the experiment. Although the model is deterministic, we parametrize the model with draws from posteriors (section S4 above). Thus, the model takes into account the data uncertainty and gives a distribution of likely outcomes. Details of the model are given in the main text.

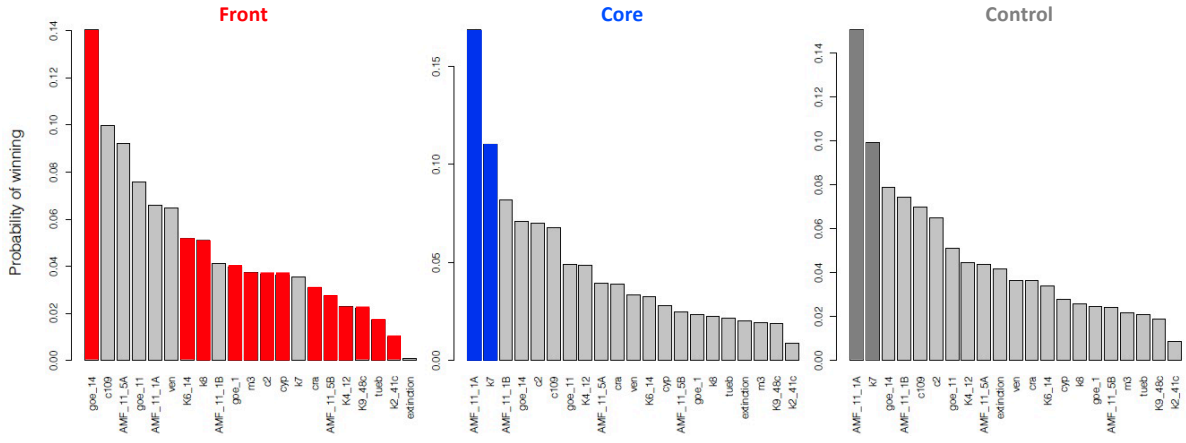


Figure S7 Histograms of strain winning probability for Front, Core and Control treatment with a quasi-extinction threshold of 0.7. Strain winning probability corresponds to the fixation probability among the 20

strains in 10000 model runs. The true potential winner candidates with the right COI genotype are highlighted in red for the Front, blue for the Core and grey for the Control treatment.

S8 Quasi-extinction threshold

The quasi-extinction threshold implies that strains go extinct if they exhibit densities below this value. The model scenario that fits the observed data indicates a large extinction threshold of 0.7, leading to a similar selection on dispersal and growth rate. Under these conditions, selection favours strains with high dispersal but also a relative high growth rate. When running additional model scenarios with decreased quasi-extinction threshold, selection for growth rate overrides selection for dispersal. Despite the bottlenecks occurring during the dispersal phase, strains with low dispersal can still reach the new patch and regrow to high density. Under these alternative conditions few extinctions occur and all strains can reach the new patch, but it is the strain with the highest growth rate (AMF_11_1A) that becomes fixed in all treatments.

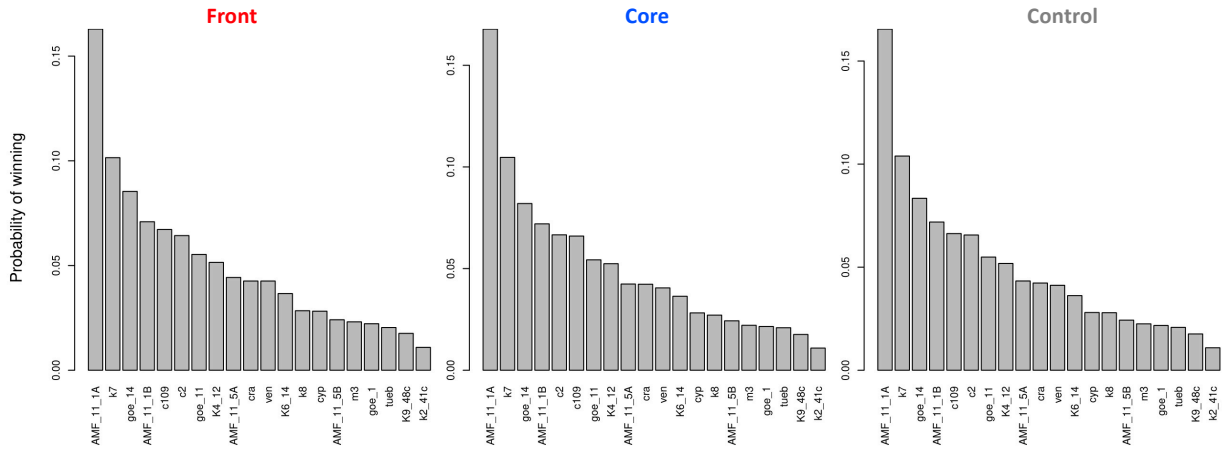


Figure S8.1 Histograms of strain winning probability for Front, Core and Control treatment with a quasi-extinction threshold 0.001.

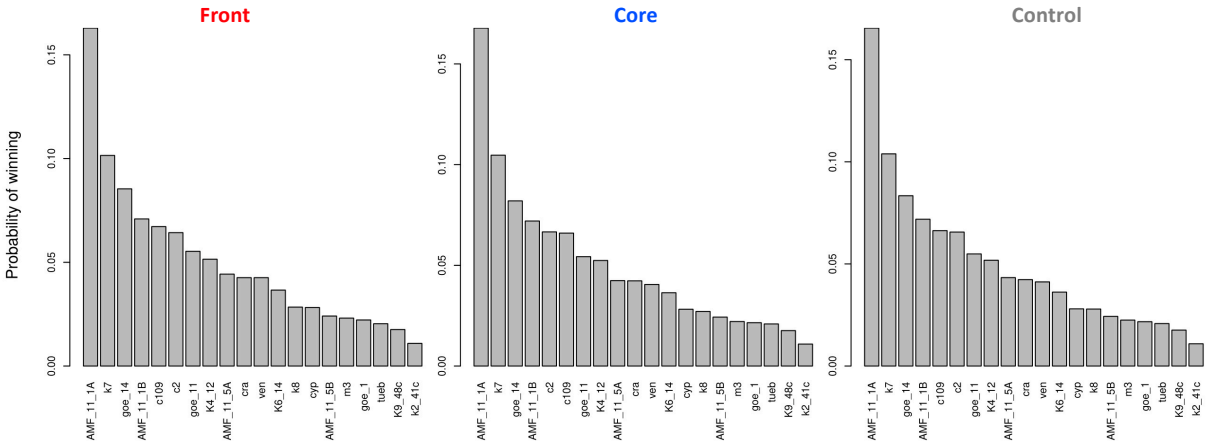


Figure S8.2 Histograms of strain winning probability for Front, Core and Control treatment with a quasi-extinction threshold 0.1.

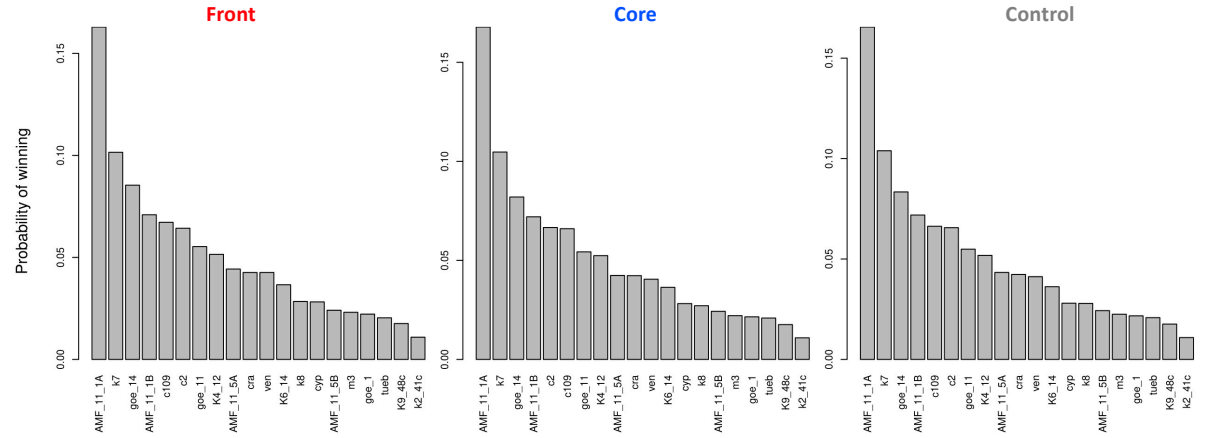


Figure S8.3 Histograms of strain winning probability for Front, Core and Control treatment with a quasi-extinction threshold 0.5.

Table S1 Details and trait values for each of the 20 strains of the founder population.

Strain	Origin (provided by)	In OK lab since	COI genotype	COI genotype detected at cycle 30	Median dispersal	95% CI dispersal	Median r0	95% CI r0	Median $\overline{N}$	95% CI $\overline{N}$
AMF_11_1A	Russia (A. Potekhin)	2014	b05	range core and control	0.060	0.044; 0.141	0.162	0.075; 0.479	263.5	204.0; 357.4
AMF_11_1B	Russia (A. Potekhin)	2014	b04	not detected	0.068	0.052; 0.144	0.093	0.063; 0.391	299.5	229. 7; 386.2
AMF_11_5A	Russia (A. Potekhin)	2014	*	not detected	0.311	0.275; 0.424	0.093	0.074; 0.275	318.3	270.1; 368.4
AMF_11_5B	Russia (A. Potekhin)	2014	b07	range front	0.124	0.088; 0.292	0.042	0.026; 0.219	183.3	116.7; 399.1
c109	Unknown	2014	b01	not detected	0.185	0.159; 0.285	0.096	0.061; 0.353	288.0	223.3; 371.7
c2	Isolated from commercial mix	2004	b07	range front	0.069	0.049; 0.176	0.065	0.033; 0.393	162.0	106.5; 393.3
cra	Kraków, Poland	2006	b07	range front	0.095	0.060; 0.320	0.039	0.008; 0.213	148.2	44.2; 830.9
cyp	Cyprus	2003	b07	range front	0.148	0.111; 0.319	0.034	0.016; 0.279	165.6	87.3; 620.0
goe_1	Stuttgart, Germany	2008	a18**	not detected	0.248	0.212; 0.373	0.056	0.037; 0.239	277.3	196.7; 415.2
goe_11	Stuttgart, Germany	2007	b01	not detected	0.202	0.174; 0.302	0.075	0.049; 0.358	279.3	199.5; 403.2
goe_14	Stuttgart, Germany	2007	b07	range front	0.272	0.235; 0.397	0.100	0.068; 0.405	251.0	203.7; 309.2
k2_41c	Japan, parents KNZ 5 & KNZ 2	2001	b07	range front	0.123	0.090; 0.270	0.042	0.030; 0.102	251.9	181.6; 403.3
K4_12	Japan, parents KNZ 5 & KNZ 2	2001	b07	range front	0.057	0.038; 0.171	0.044	0.019; 0.286	201.3	116.6; 667.8
K6_14	Japan, parents KNZ 5 & KNZ 2	2001	b07	range front	0.207	0.169; 0.355	0.079	0.060; 0.272	348.3	273.9; 432.9
k7	Japan, parents KNZ 5 & KNZ 2	2001	b05	range core and control	0.051	0.037, 0.121	0.135	0.084; 0.330	293.4	256.0; 334. 2
k8	Japan, parents KNZ 5 & KNZ 2	2001	b07	range front	0.411	0.370; 0.537	0.073	0.058; 0.239	249.1	204.4; 298.5
K9_48c	Japan, parents KNZ 5 & KNZ 2	2001	b07	range front	0.137	0.114; 0.224	0.063	0.048; 0.202	329.1	239.9; 446.2
m3	Isolated from commercial mix	2004	b07	range front	0.361	0.319; 0.495	0.059	0.045; 0.204	309.0	231.1; 427.8
tueb	Tübingen, Germany (H-D Görtz)	2001	b07	range front	0.082	0.051; 0.282	0.032	0.018; 0.216	223.0	82.9; 662.9
ven	Venice, Italy	2006	a01**	not detected	0.253	0.220; 0.363	0.083	0.063; 0.286	353.7	276. 8; 446.0

* identified as *Paramecium multinucleatum*

** determined in Killeen, J., Gougat-Barbera, C., Krenek, S. & Kaltz, O. (2017). Evolutionary rescue and local adaptation under different rates of temperature increase: a combined analysis of changes in phenotype expression and genotype frequency in Paramecium microcosms. *Molecular Ecology*, 26, 1734-1746.