Additive effects of environmental and demographic variation shape the repeatability of evolution across replicated experiments

Bisschop, K.; Wortel, M.T.; Kraaijeveld, K.; Marïen, J.; Rathmann, I.; Bonte, D.; ... ; Blankers, T.

Evolution Letters, 2026, **0**(0), qrag017
https://doi.org/10.1093/evlett/qrag017
Letter

Additive effects of environmental and demographic variation shape the repeatability of evolution across replicated experiments

Karen Bisschop [1,2,3,4,5], Meike T. Wortel[2,6], Ken Kraaijeveld[2,7], Janine Mariën[8], Isabel Rathmann[1,6], Dries Bonte[4], Steven Declerck[9,10], Marjon de Vos[5], Rampal S. Etienne[5], Steven Goossens[4], Jan Kammenga [11], Frederik Mortier[4,12], Joost Riksen[11], Maurijn van der Zee[13], Koen Verhoeven [14], Suzanne Wiezer[15], Lars Zandbergen[5], Martijn Egas[1,5], Jacintha Ellers [8], Astrid T. Groot[1], Marcel E. Visser[9], Thomas Blankers[1,2]

[1]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands
[2]Origins Center, The Netherlands
[3]Laboratory of Aquatic Biology, KU Leuven Kulak, Kortrijk, Belgium
[4]Terrestrial Ecology Unit, Ghent University, Ghent, Belgium
[5]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands
[6]Molecular Biology and Microbial Food Safety, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands
[7]Department of Bioinformatics, University of Applied Sciences Leiden, Leiden, The Netherlands
[8]Amsterdam Institute for Life and Environment, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[9]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands
[10]Laboratory of Aquatic Ecology, Evolution and Conservation, Department of Biology, KU Leuven, Leuven, Belgium
[11]Laboratory of Nematology, Wageningen University & Research, Wageningen, The Netherlands
[12]Bioinformatics and Evolutionary Genomics, VIB-UGent, Zwijnaarde, Belgium
[13]Institute of Biology, Leiden University, Leiden, The Netherlands
[14]Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands
[15]Department of Aquatic Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands
Corresponding author: Etienne Sabbelaan 53, 8500 Kortrijk, Belgium. Email: kbisschop.evo@gmail.com

Abstract

The repeatability of evolution is fundamentally important for understanding the origin and diversification of life as well as for developing evolutionary forecasting tools. Repeatability is limited by stochasticity, here defined as changes that are independent of genotypic fitness effects. Over short timescales, the two main sources of stochasticity of evolutionary change are environmental stochasticity and demographic (life-history) stochasticity. Quantifying the effect of these two sources of stochasticity and how they interact in driving fitness outcomes is crucially important for predicting contemporary evolutionary responses. To gain insights in the effects of stochasticity, five institutes replicated an evolutionary experiment exposing *Caenorhabditis elegans* to novel rearing conditions. Replication across the institutes led to variation in selective environments, including through divergent microbiomes among institutes. Replication within institutes was done across demographic treatments that influence the potential for population-size dependent fluctuations in allele frequencies (drift) and genetic hitchhiking (draft). We found high among-institute variation in fitness outcomes, which was partially explained by variation in microbiota. Whereas lab-specific effects explained most of the variance in mean fitness, the repeatability of fitness outcomes depended more on demographic heterogeneity. Specifically, population bottlenecks resulted in high among-replicate variation in fitness. When combined, environmental and demographic stochasticity additively reduced repeatability, underlining their additive importance in developing evolutionary forecasting tools. These results further highlight the importance of statistically integrating heterogeneity in experimental evolution to identify factors constraining outcome repeatability and study replicability.

Keywords caenorhabditis elegans, natural selection, microbiome, evolutionary predictability, replicability

Received: 18 November 2025. **Revised:** 18 February 2026. **Accepted:** 30 March 2026

Introduction

The predictability of evolution has attracted major interest in evolutionary biology (Lässig et al., 2017; Nosil et al., 2021; Wortel et al., 2023). Understanding how predictable evolution is, and the factors shaping predictability, clarifies the origin and present-day patterns of biodiversity, such as whether convergent phenotypes (i.e., similar traits in divergent lineages) arise from similar selective pressures or from constraints and chance (Losos, 2011; Mahler et al., 2017). Identifying mechanisms influencing evolutionary predictability can improve evolutionary forecasting (Wortel et al., 2023). However, accurately predicting evolutionary trajectories and outcomes, even short-term evolutionary responses such as forecasting trait change or fitness outcomes in response to rapid changes, remains challenging.

Two major obstacles are accurately estimating selection strength and accounting for historical contingencies. Selection strength depends on resource and trait distributions, genetic architectures, and species interactions (Nosil et al., 2020; Wortel et al., 2023), requiring detailed ecological and genomic data. Historical contingencies, such as past population structure, shape genomic backgrounds of the targets of selection (Blount et al., 2018; Travisano et al., 1995). Both are influenced by stochasticity, i.e., changes in genomes, individuals, and environments independent of adaptation (Lenormand et al., 2009). Stochasticity inherently limits evolutionary predictability, for example by altering genetic variants and their fitness. Evolutionary genetic changes are often highly sensitive to stochasticity, resulting in low repeatability of genetic trajectories (Lässig et al., 2017). Evolutionary forecasting of fitness should be more accurate, but their sensitivity to stochasticity is unclear. Here, we quantify how stochasticity affects end-point repeatability of fitness ("fitness response"). Although we recognise that repeatability is not same as predictability, quantifying the effect of stochasticity on the repeatability of fitness evolution reveals inherent limits to predictive models and potential solutions.

Three forms of stochasticity shape evolution: mutational, environmental, and demographic (life history) stochasticity (Lenormand et al., 2009). For short-term change based on standing genetic variation, mutational stochasticity is negligible. Environmental stochasticity results from spatio-temporal changes in interactions with the abiotic and biotic environment and results in variation in the selective regime. This drives long-term unpredictability in wild populations (Grant & Grant, 2002; Nosil et al., 2024). In the lab, random fluctuations in selection through time have been found to impact microbial evolution by diversifying the direction and magnitude of the selection response (Cooper & Lenski, 2010; Karve et al., 2015; Lenski, 2023).

Demographic stochasticity affects variation available to selection and the genetic background of selected variants and results from individual-level stochastic events. These events, such as survival and reproduction, are described by a probability that is scaled by population size (drift). Similarly, recombination and the segregation of genotypes are described by probabilities that depend on linkage and genetic hitchhiking (draft) (Lenormand et al., 2009). Evolutionary experiments have shown that bottlenecks or other factors that influence the population size can limit repeatability because drift increasingly drives evolution when effective population size is low (Bisschop et al., 2022; Lachapelle et al., 2015; Rozen et al., 2008; Weber, 1990; Wein & Dagan, 2019; Windels et al., 2021) leading to divergent evolutionary trajectories over time (Lenski & Travisano, 1994). However, as long as population sizes are not excessively small (effective population size, Ne $>$ 10) and selection is not excessively weak (selection coefficient, s $>$ 0.01), variation in selection will be more important than drift for trait repeatability and predictability (Gompert et al., 2022; Ohta, 1972). In contrast, draft is important at normal/high population sizes, in particular in the face of low recombination and sign epistasis (Barton, 1995; Gillespie, 2001; Weinreich et al., 2005). Given this multitude of effects of demographic stochasticity, variation in founder populations and genomic linkage strongly influence repeatability by shaping the response to selection (Blount et al., 2018; Otte et al., 2021; Schlötterer, 2023; Teotónio et al., 2009).

Natural populations will commonly differ in both historical contingencies (through demographic stochasticity) and in selective regimes (through environmental stochasticity). The interaction of these sources of stochasticity is particularly important. Selection targets specific regions of the genome, while demographic effects are thought to affect genomes more homogeneously. However, the extent to which genomic regions are sensitive to demographic effects, such as drift, varies with the amount of genetic variation in those regions (Charlesworth & Jensen, 2021). Thus, when different selective regimes target different regions in genomes with different histories that are also differently sensitive to demographic effects, interactions between demography and environment may have unpredictable consequences (Blount et al., 2018; Otte et al., 2021; Schlötterer, 2023; Teotónio et al., 2009). Quantifying their relative roles and examining how demographic and environmental stochasticity interact will guide data collection and analysis to improve evolutionary predictions.

Rationale and hypotheses

In this study, we address how environmental and demographic stochasticity shape the repeatability of fitness outcomes during adaptation from standing genetic variation. (Lind, 2019) Using a genetically diverse *Caenorhabditis elegans* population (Noble et al., 2017; Teotonio et al., 2012) exposed to a novel environment (16 °C and *Priestia megaterium*), we conducted a replicated evolutionary experiment across five institutes following standardised protocols. Experimental evolution provides a potent tool to determine causes of repeated evolution and conditions under which repeatability fails (Lind, 2019). Minor differences in handling and lab-specific microbiomes introduced undirected environmental heterogeneity (Figure 1). This heterogeneity allows us to test the prediction that repeatability is lower between populations in dissimilar environments versus between populations in similar environments. Secondary treatments—population bottlenecks, reproductive mode, population age synchronisation, and reduced gene flow—introduce demographic heterogeneity and, consequentially, the potential for drift and draft. This heterogeneity allows us to test the prediction that repeatability is lower between populations with different demographic histories versus between populations with the same demographic history (for detailed expectations for the effect on fitness change and repeatability per treatment condition, see the Methods). After 15 weeks (∼20 generations), we quantified evolutionary change by measuring population growth rate in the novel environment, and charac-

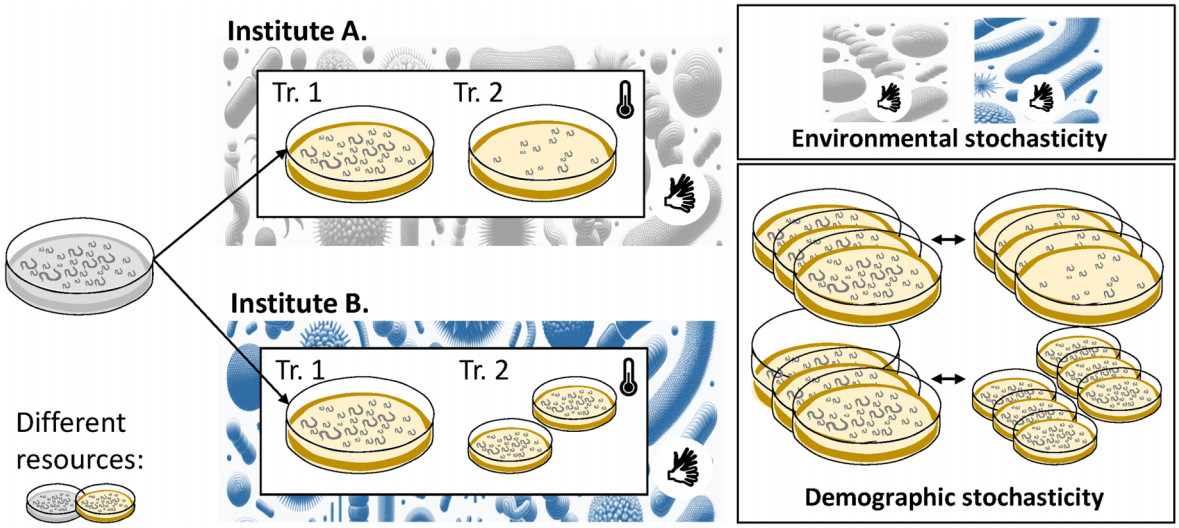

Figure 1 Environmental and demographic stochasticity. All institutes (institute A and B) performed an identical evolutionary experiment where *C. elegans* and microbial communities evolved on a novel resource, the primary treatment (Tr. 1). The controlled conditions during the experiments (temperature and other climate attributes, as well as food source) were kept similar (indicated with the thermometer symbol), but the uncontrolled conditions were different (e.g., institute microbiota in different colours, and institute handling routines as symbolised with the hand icons), resulting in environmental heterogeneity across institutes. Furthermore, all institutes except one conducted a second treatment (Tr. 2) which differed between institutes and resulted in demographic heterogeneity across treatments.

terised microbial communities to assess environmental variation among institutes. Adaptive shifts in *C. elegans* often occur within only a few to tens of generations under strong selection on standing genetic variation (Teotónio et al., 2017), especially when exposed to sudden environmental shifts (Guzella et al., 2018). We thus expect 15 weeks to be sufficient to detect fitness changes even though larger differences typically emerge over longer evolutionary timeframes (Teotónio et al., 2017). We hypothesised that repeatability would be highest within institutes and treatments, and would decline when comparing across treatments, across institutes, or across both. Furthermore, we expected institute-specific microbiota to influence variation in population growth rates.

Repeatability and replicability of experimental evolution

Beyond evolutionary *repeatability*, this study addresses *replicability* and external validity in the context of environmental and demographic stochasticity in experimental evolution. We define *repeatability* as the degree to which independent evolutionary processes lead to similar genetic, phenotypic, or fitness outcomes when subjected to similar selective environments. *Replicability* is defined as the degree to which independent researchers, using the same methods but different datasets, arrive at the same conclusions (National Academies of Sciences (2019)).

Limited replicability of experimental studies warrants the systematic introduction of heterogeneity in genetic and environmental factors (Desjardins et al., 2021; Milcu et al., 2018; Voelkl & Würbel, 2021; Voelkl et al., 2020) to develop more robust theories in ecology and evolution that do justice to the complexity of biological systems. In evolutionary experiments historical contingency can cause divergent outcomes (Blount et al., 2018). Including a priori conceived (Desjardins et al., 2021) or a posteriori characterised heterogeneity (Chesler et al., 2002) across institutes thus help uncover how reliable evolutionary inferences in one institute translate to a broader context (i.e., external validity) (Desjardins et al., 2021).

Methods

Study species and resource

We used the bacterivorous nematode *Caenorhabditis elegans* (Maupas, 1900), a model for experimental evolution due to its small size (adult females: ∼1 mm), high fecundity (∼300 eggs/week), short generation time (min. 50 h), ease of laboratory maintenance and ability to undergo cryopreservation (Gray & Cutter, 2014). The experiment started from a dioecious, genetically diverse population, D00 (Noble et al., 2017; Teotonio et al., 2012), maintained by obligate outcrossing at a 1:1 sex ratio.

The D00 population, shared by the Teotónio laboratory (IBENS-Paris), was expanded on *Escherichia coli* OP50 at 20 °C at the University of Amsterdam (UvA) and distributed as frozen aliquots (hereafter referred to as "ancestral" population) to five institutes. Populations evolved on a novel food source, *Priestia megaterium* (DSM 509). Both bacterial strains were distributed from the same starting population and maintained at 4 °C. Nematodes were cultured on NGM plates seeded with bacterial lawns. Detailed protocols for rearing, media preparation, transfers, and freezing follow standards (Stiernagle, 2006) and are provided in Supplementary Appendix I.

Primary treatment

Shared axis of selection

Populations of obligate outcrossing *C. elegans* evolved for 15 weeks on *P. megaterium* at 16 °C (Figure S1). Cultures were maintained on ± 12 mL NGM agar plates (94 × 16 mm sterile Petri dishes with vents, Greiner Bio-One), prepared centrally using a Petri dish

filling machine (Mediajet and Mediaclave by INTEGRA Biosciences) at the Netherlands Institute of Ecology [NIOO], and distributed to the five institutes (NIOO, University of Groningen [UG], Ghent University [UGent], Vrije Universiteit Amsterdam [VU], and Wageningen University [WUR]). 16S rRNA gene data revealed low-level contamination by *Stenotrophomonas sp.* across plates, adding a uniform selection pressure. Shared axes of selection included the resource change (*P. megaterium*, harder to consume and growth-delaying (Shtonda & Avery, 2006)) and reduced temperature (16 °C instead of 20 °C) to achieve ∼1 generation/week. These factors, plus the contamination, jointly define the shared axis of selection.

At experiment start (Figure S1), each institute thawed ancestral aliquots (10 per institute; WUR: 6; VU: single aliquots per replicate) and expanded populations on *E. coli*. Nematodes were pooled, density estimated, and evolution initiated on *P. megaterium* plates with equal starting numbers (500 nematodes at UGent and UG; 1000 elsewhere). Each replicate (five per institute, except WUR: three) was maintained on three plates to preserve genetic variation, support larger population sizes, and buffer against plate failure. Transfers involved pooling plates, estimating density, and moving 500 nematodes; excess individuals were cryopreserved for later assays.

We checked for unintended founder effects prior to the experiment by measuring genetic differentiation between ancestral populations and week 0 populations to the extent that these samples were available. DNA was extracted from several hundred individuals per population (DNeasy B&T kit, Qiagen) after triple M9 washes and sequenced on Illumina Hi-Seq (150 bp paired-end) by IGAtech (Udine, Italy). Reads were quality checked (fastQC 0.11.9), trimmed (trimmomatic version 0.39), mapped to the WBCel235 reference genome (Consortium, 1998; Davis et al., 2022) using the Burrows-Wheeler Aligner (Li & Durbin, 2009). We then selected only those reads that were properly paired after alignment against the *C. elegans* reference genome. SNPs were called using bcftools (version 1.17) (Danecek et al., 2021). Variants were filtered out if DP was below five or above three times the average coverage; if quality score was below 20; if AD was below two. Allelic frequencies were only considered if larger than 1% and lower than 99%. Indels were excluded in this analysis. We retain 437,071 variant positions. We then examined genetic similarity across the three available ancestral population by scoring the frequency at which variants were unique to a population, or shared between two or all three. We compared that to the same frequencies obtained for the five available week 0 populations from VU for the backbone and both bottleneck treatments, subsampling all combinations of three out of five replicates for comparability with the ancestors. In addition, we examined genetic variation between ancestors and week 0 samples in genetic Principal Component space under the rational that populations with similar genetic backgrounds cluster in multivariate genetic space. We performed PCA using Scikit-learn in python (Pedregosa et al., 2011) and visually examined clustering.

Environmental stochasticity—variation in handling routines and microbiome profiles

In addition to the shared axis of selection, there was environmental heterogeneity across institutes. Institutes differed in handling routines (e.g., plate storage, bacterial growth time, washing protocols). Full details are in Supplementary Appendix I. We also profiled institute-specific microbiota associated with *C. elegans* by sequencing samples from primary and secondary treatments (see below) at week 0/1 and 15. Due to sample limitations, some early timepoints were missing (see Figure S2 for available samples). From the sequence alignment files described above, host reads were removed using samtools (version 1.9). Non-host reads were classified at genus level using the k-mer based kraken2 software (Wood et al., 2019) and obtained species abundances with Bracken (Lu et al., 2017). The database (PlusPf database, latest version 9/4/2024) included bacterial, viral, archaea, protozoa, and fungi references. Genera $< 1\%$ abundance were excluded (31 retained). Abundances were Hellinger-transformed and analysed via PCA and redundancy analysis (vegan package 2.6–10 in R) (Oksanen et al., 2025) to assess variation across institutes and treatments. The loadings in the PCA were generated with the scores function with scaling two.

Secondary treatments

Secondary treatments targeted demographic stochasticity through four variables: population bottlenecks, mixing within-population replicates, reproductive mode, and initial bleaching (Figure S1). Each parameter was varied at a single institute, with other conditions matching primary treatment.

At VU, bottlenecks were introduced before the first week of experimental evolution. Therefore, only the initial population size in week zero differed from the primary treatment. Three treatment levels were investigated: a strong bottleneck (five female founders), a moderate bottleneck (50 female founders), or no bottleneck (500 founders). A single bottleneck reduces the effective population size by $\frac{1}{2N_e}$, with N_e here being the effective size of the bottleneck (Falconer & Mackay, 1996). The expectation was that smaller effective population sizes would thus increase the potential for drift and founder effects. This could lead to a decrease in the adaptive potential (although this depends on several additional factors (e.g., Frankham et al., 1999)), and will limit the repeatability of fitness change (Wahl et al., 2002). Since week 0 samples were available for the bottlenecked populations, we also included these samples in the genetic similarity and founder effect analysis described in the "Shared axis of selection" section. The rationale was that if a bottleneck introduces a founder effect, samples would share less variants with the ancestor compared to non-bottlenecked week 0 samples and would spread out in multivariate genetic PCA space due to divergent stochastic sampling of alleles. This was unfortunately only possible for the bottleneck treatment due to limited availability of week 0 samples.

At UGent, isolation was tested by maintaining five Petri dishes separately instead of mixing three dishes per replicate as in the default protocol (Figure S1). Isolation may reduce overall levels of genetic variation if populations within a single Petri dish do not capture the full extent of genetic variation in the ancestor. This could then limit adaptive potential like a bottleneck. However, isolation also prevents maladaptive recombination of locally adaptive allele combinations (local within a single Petri dish), which can strengthen local adaptation within a Petri dish (Bisschop et al., 2019; Lenormand, 2002). Repeatability will likely be lower in the isolation treatment because more there is a higher potential for population specific, idiosyncratic changes.

At NIOO, reproduction modes were contrasted: obligate dioecious D00, obligately self-fertilising (monoecious populations

or M00) and wildtype populations (androdioecious or A00; Theologidis et al., 2014). In the wildtype populations, sexual reproduction occurred through self-fertilisation and through the production of males that cross with hermaphrodites. All populations share ancestry through A6140 and A00, into which loss-of-female-spermatogenesis [fog-2[q71)] and male-killing [xol-1(mt3055)] alleles were introgressed to create the D00 and M00 populations, respectively (Teotonio et al., 2012; Theologidis et al., 2014). Because these workflows entail distinct introgression paths and potential bottlenecks, we do not assume identical baseline allele frequencies or diversity across reproductive modes. We therefore interpret "reproductive mode" as mating system together with its inherited baseline genomic context. We expect M00 to exhibit a higher growth rate than D00, due to doubled reproductive capacity. The reduced genetic diversity is expected to limit long-term adaptation. D00 should maintain highest diversity and adaptive potential. A00's performance depends on male frequency dynamics: plasticity could confer benefits of both systems, but slow response may impose costs (e.g., reduced mating opportunities or inefficient resource allocation).

At WUR, bleaching was applied before evolution, which is a common procedure in *C. elegans* rearing and kills all organisms (including contaminant bacteria) except the unhatched embryos inside the gravid females. Consequently, it is used to sterilise, and age synchronise populations. It is expected that this treatment mostly affects the demography in the population (as all individuals have the same age at the start of the experiment). Synchronising the population at the embryonic stage may favour alleles beneficial early in development while reducing founder effects, as reproduction is shared across the entire cohort. This could improve adaptive potential by maintaining broader genetic representation. An overview of the treatments, treatment levels and sample sizes can be found in Table S1.

Population growth rate assessments

Ancestral populations were cryopreserved after expansion in the lab, but before exposure to the novel conditions; starting populations after one week on *P. megaterium* at 16 °C; and final populations after 15 weeks (~20 generations). Growth rate assessments were performed at a single location (VU Amsterdam) to prevent assessment bias. Cryopreserved samples were thawed and reared on *E. coli* at 20 °C for one week to create a common garden, remove (grand)parental effects, and produce sufficient nematodes. This assessment thus compares heritable, (epi)genetic effects. Growth was measured for three technical replicates, each starting with 500 nematodes, per replicate unless the expansion step did not yield enough nematodes. For some replicates, the assessment was repeated on different days (resulting in more technical replicates: 3 x the number of measuring days). Population size after seven days was estimated by washing nematodes from plates in a microcentrifuge tube using 2 mL M9 and counting three 5 μL droplets. Growth rate was calculated as $r = \frac{ln(N_t/N_0)}{t}$, with t = 7 days and $N_0 = 500$. Population growth rate is driven both by developmental rate and fertility and increasing growth rates over time are thus interpreted as adaptation to the novel conditions. Counts at earlier timepoints (4, 5, 6 days) validated calculations (Supplementary Appendix I). Manual counts correlated strongly with BioSorter flow cytometer data (Supplementary Appendix I). Some samples with visual contamination (UGent) or poor post-freezing survival (five out of 138 sample measurements) were excluded.

Statistical analysis

Population growth rates—primary treatment

Analysis was performed in R v4.4.2 (R Core Team, 2025). We evaluated whether daily population growth rate differed between week 1 and week 15 across institutes using GLMMs (glmmTMB v1.1.10, Brooks et al., 2017) with Gaussian distribution (Shapiro-Wilk normality test: $W = 0.9928$, P-value $= 0.4329$). The fixed effects were week (categorical variable: week 1 or week 15), institute (categorical variable: NIOO, UG, UGent, VU, and WUR), and their interaction; the random variables were the assessment date and replicate ID (to link the measurements on the same replicate across weeks). Residual diagnostics (DHARMa, Hartig, 2024) showed no violations of model assumptions. Wald χ^2 tests assessed fixed effects; pairwise comparisons used emmeans v1.10.6 (Lenth, 2024) with Tukey's method. Selection response was calculated as the difference in population growth rate between weeks.

Population growth rates—secondary treatment

Early time points (week 0/week 1) were unavailable for some secondary treatments due to unsuccessful freezing and sample loss; therefore, we restricted analyses for secondary treatments to the final time point (week 15). We tested variation in daily population growth rate within institutes using GLMMs (glmmTMB v1.1.10, Brooks et al., 2017). Gaussian distribution was applied for UGent (Shapiro-Wilk normality test: $W = 0.96662$, P-value $= 0.1681$) and WUR (Shapiro-Wilk normality test: $W = 0.95634$, P-value $= 0.6291$), whereas VU and NIOO data were orderNorm transformed (bestNormalize v1.9.1, Peterson & Cavanaugh, 2020). Fixed effects were the secondary treatments; random effects were assessment date and replicate ID. Residual diagnostics (DHARMa, Hartig, 2024) indicated minor issues: in one model, some quantile regressions failed, and in another, the Levene test indicated significant heterogeneity of variance. These deviations were not expected to substantially affect the result interpretation. Pairwise comparisons used emmeans v1.10.6 (Lenth, 2024) with Tukey's method.

The effect of microbiome on population growth rate

To assess whether variation in microbiota is the result of institute-specific accumulation of environmental microbes, we tested if the microbiota diversity increased from week 1 to week 15. Analyses were restricted to primary treatment data. We calculated Shannon diversity (vegan v2.6–10, Oksanen et al., 2025) and applied Kruskal-Wallis rank sum tests per institute with week as fixed effect. We examined whether microbiome composition explained population growth rate using linear models (lme4 v1.1–35.5, Bates et al., 2015). The daily growth rate (averaged over the repeated measurements for each replicate ID) was response variable; the fixed effects were institute, week, and the first three principal components of microbiome variation. We refined models via likelihood ratio-tests. We also assessed the change in the population growth rate versus microbiome composition changes using random forest analysis (ntree = 100000; randomForest v4.7–1.2, Liaw & Wiener, 2002).

Effect of environmental versus demographic stochasticity on repeatability

We investigated the effect of environmental (among-institute variability) and demographic heterogeneity (within-institute, between-treatment variability) on the (endpoint) repeatability of adaptive evolution (population growth rates). We partitioned variance in week 15 growth rates using nested linear models (institute: treatment: replicate: measurement) in MCMCglmm (Hadfield, 2010). Both primary and secondary treatments were included. Greater variance at the institute level, compared to the treatment or replicate term, indicates environmental stochasticity as the main constraint. We iteratively added nested variables to the model, starting with institute, and then calculated for each model the increase in variance explained relative to a null model with no explanatory variables.

We then assessed whether repeatability depended more on environmental stochasticity versus demographic stochasticity. We used linear mixed effect models in MCMCglmm to compare among-replicate variance in population growth rates attributable to institutes with among-replicate variance attributable to treatments. Repeatability was considered to be higher when among-replicate variance in growth rates was lower. If environmental stochasticity is the main factor constraining repeatability, we expect that there is more among-replicate variation partitioned across institutes than across the treatments. We first fitted a global model with treatment as fixed effects and the institute, and technical and biological (evolutionary) replicates as random effects using uninformative priors (degree of belief $\nu = 0.002$). Posterior estimates for the variance among measurements and replicates were then used to define informative priors with prior variance equal to the posterior mode and degree of belief equal to the posterior mode divided by the 95% Highest Posterior Density interval (smaller spread relative to the mean will increase degree of belief). The informative priors were used here to incorporate information from the full data set to help model convergence when analysing small subsets of the data. We fitted the following models: (1) primary treatment only (institute random effect), (2) both primary and secondary (institute and treatment fixed effect), (3) institute-specific subsets (treatment fixed effect). Posterior distributions were estimated from 1000 MCMC samples (1000000 iterations, 100000 burn-in, 1000 thinning interval) and modes with 95% HPD intervals were reported (Brown & Forsythe, 1974).

Results

Adaptive evolution in the primary treatment

Week 0 replicates from UvA and VU exhibited high levels of variant sharing (94% of variants shared among all replicates across institutes) (Figure S3A) and clustered closely with ancestral samples in PCA space (Figure S4). We thus assume that there are no unintended founder effects in the primary treatment by distributing the populations across institutes. We investigated whether the final (week 15) populations of *C. elegans* increased in population growth rate compared to the starting populations (week 1) at the five research institutes (Figures 2A and 2B). Overall, variance in population growth rate was explained by institute ($\chi^2 = 127.2$ and P value < 0.0001) and the interaction between institute and week ($\chi^2 = 32.0$ and P value < 0.0001). Two institutes evolved higher population growth rates (UG: z ratio $= 3.6$ and P value $= 0.0003$; VU: z ratio $= 2.4$ and P value $= 0.0186$) and one institute lower growth rates (UGent: z ratio $= -3.6$ and P value $= 0.0003$) across all replicates from week 1 to week 15 (Figure 2; Table S2).

Adaptive evolution in the secondary treatment

Population growth rates varied among institutes (Figures 2A, B) and among treatments (Figures 2C-F). The final growth rate was higher in the monoecious populations, in the isolated, and in the bleached treatments compared to dioecious (t ratio $= -2.8$; P value $= 0.0218$), mixed (t ratio $= -8.4$; P value < 0.0001), and not-bleached (t ratio $= -4.9$; P value $= 0.0004$) populations in the primary treatment (Figures 2C-F, Table S4). Neither moderate (t ratio $= -4.9$; P value $= 0.5298$) nor strong population bottlenecks (t ratio $= -0.642$; P value $= 0.7973$) affected growth rates, despite replicates having distinct genetic composition relative to each other and relative to ancestral and primary treatment samples (Figures S3B, C; Figure S4).

Environmental stochasticity—institute-specific microbiomes

By replicating the experiment across different institutes, we introduced undirected environmental heterogeneity that was independent of the *C. elegans* genotype. At least two sources of variation contributed to this heterogeneity. The first were small, undirected fluctuations in the protocols used in transfer and treatment of the nematodes (Supplementary Appendix I). The second were undirected differences in the microbial communities that became associated with the nematodes over time.

To assess the microbial community composition and diversity, we leveraged the metagenome sequence data obtained for 93 ancestral, before, and after evolution lines. Prior to sequencing, repeated washing of the sampled nematodes removed most environmental bacteria and fungi. Microbial sequences obtained here are thus assumed to be from bacteria and fungi intimately associated with the nematodes. Between 1% and 91% of trimmed reads were classified as bacterial or fungal genomes (Supplementary Appendix II). Microbiomes identified in sequence pools through metagenomic profiling showed an increase in alpha diversity, meaning higher species richness, over time for three institutes: UGent, UG, and VU (Figure S5A, Table S3) and were divergent among institutes (Figures S2 and S5B).

Marginal effects of microbiome variation on population growth rates

We tested whether the random variation in environment affected population growth rate by modelling the daily growth rate as a response of institute, treatment, and microbiome composition (using its principal components). We identified week, institute, and microbiome PC3 (but not microbiome PC1 and PC2) as predictors for growth rate variation in the set of samples subject to

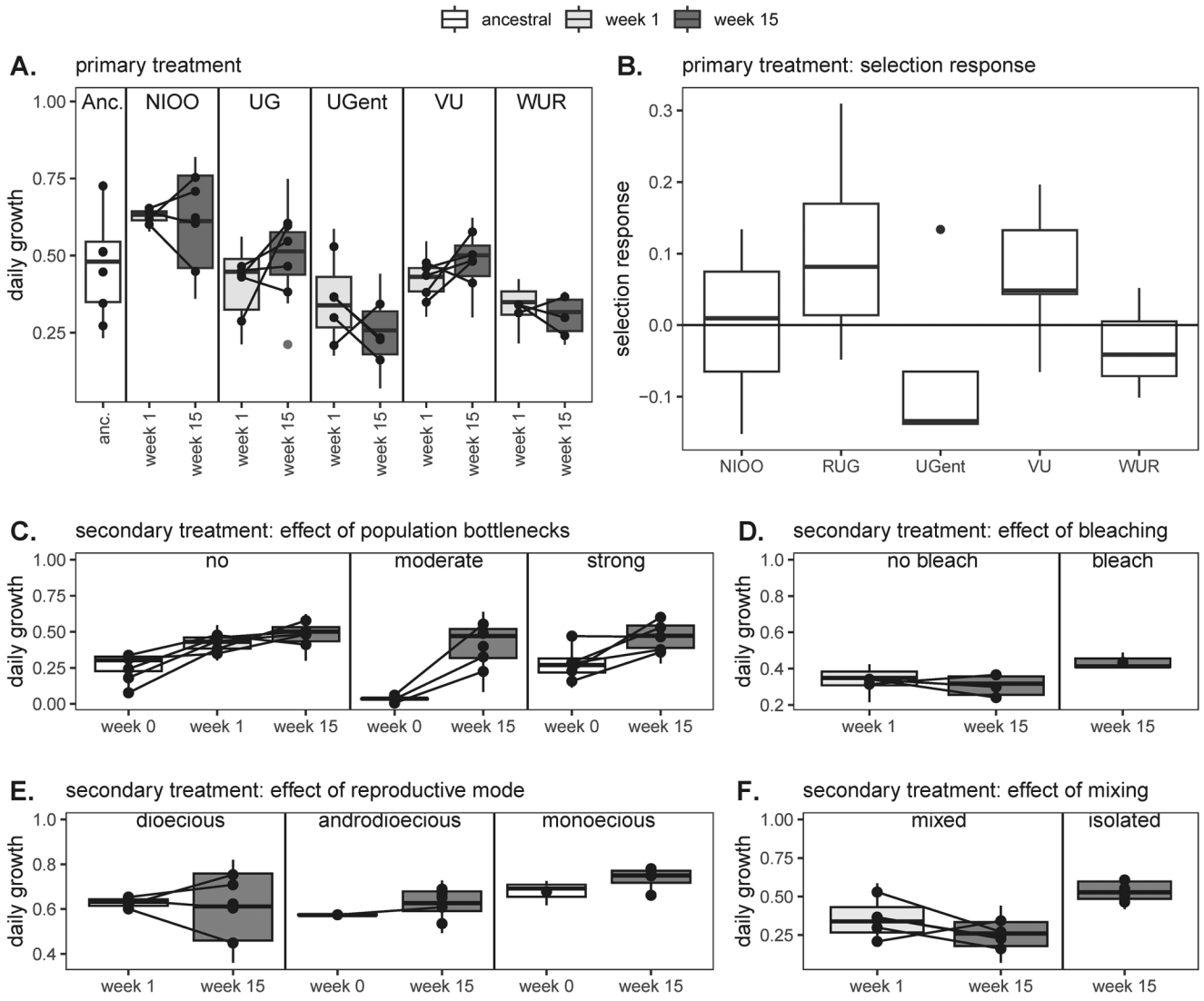

Figure 2 Growth rate evolution. (A) Daily growth is calculated as $r = \frac{ln(N_t/N_0)}{t}$ with t = 7 and $N_0 = 500$. Solid points are per-biological replicate average growth rates (Anc.: n = 5, NIOO: n = 5, but data for one replicate is missing in week 1; UG n = 5, UGent n = 5; VU n = 5; WUR n = 3) across the technical replicates (ranging from 3 to 9 per biological replicate). Lines connect week 1 and week 15 averages for the same replicate. Boxes show inter-quantile range, median; whiskers are 1.5 x interquartile range of the spread in daily growth across the technical replicates (the grey solid point is an outlier from the boxplot). (B) Selection response per replicate line is calculated as $r_{week15} - r_{week1}$. Box-and-whiskers are as in A). (C–F) The y-axis shows the daily growth rates, and the colours indicate the time point during experimental evolution (white: week 0, light grey: week 1, and dark grey: week 15). Solid points are per-biological replicate average growth rates and lines connect biological replicates across time points when multiple time points are available. The effect of population bottlenecks with the different treatment shown in different panels: no bottleneck, moderate bottleneck (50 individuals at the start), or strong bottleneck (5 individuals at the start) is shown in C. The effect of bleaching with the left panel without initial bleaching step and the right panel with initial bleaching step is shown in D. The effect of reproductive mode with the three secondary treatments shown as different panels: dioecious, androdioecious, and monoecious is shown in E. The effect of mixing with the mixing of three petri dishes during each refreshment step or the isolation of the petri dishes is shown in E. Early time points (week 0/week 1) were unavailable for some secondary treatments; therefore, we restricted analyses for secondary treatments to the final time point (week 15).

the primary treatment only (ANOVA: $F_1 = 5.35$; P value = 0.026; when adding PC3 to linear model, ΔAIC = 3.92; Table S5). The bacterial genera mainly driving PC3 are *Serratia*, *Pseudomonas*, and *Brucella* and a total of 8.72% of the variance in microbiome composition across samples is explained by PC3. Using the same data, we tested whether microbiome variation explained variation in population growth rate *change* (from week 1 to week 15) using a random forest regression. Presence/absence of specific taxa (relative abundance > 1% in at least one sample) predicted growth rate change variation in addition to institute specific effects (Figure S6). In the set of samples that also included all secondary treatments, including microbiome PC axes improved model fit and when included in the model, PC axes significantly explained population growth rate variation (PC1: $F_1 = 5.20$; P value = 0.026; PC3: $F_1 = 5.56$; P value = 0.021; Table S6).

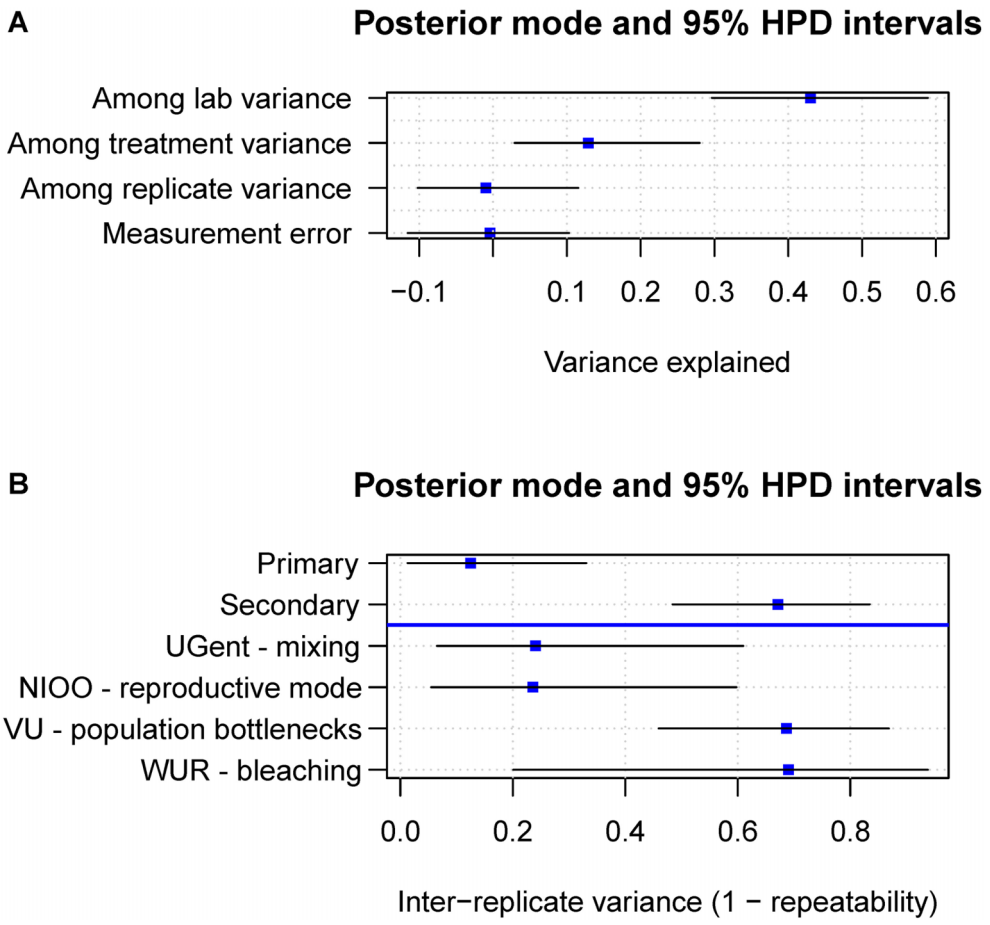

Figure 3 (A) Population growth rate variance partitioning. Population growth rate variance explained by among-institute variation (environmental stochasticity), among-treatment variation (demographic stochasticity), among-replicate, and among-measurement variation. The rectangles indicate the posterior mode, and lines delineate the 95% HPD interval of the posterior distribution for variance components in a MCMC generalised linear model. Some variance estimates are slightly negative, which is mathematically implausible but can occur due to the MCMC model's partitioning approach and prior specification. This suggests that the estimated contribution to trait differences is negligible. (B) Population growth rate repeatability. Repeatability (among-replicate variance relative to total variance) of week 15 population growth rate across institutes and across treatments within institutes (UG is not shown as no secondary treatments were performed in this institute). The rectangles indicate the posterior mode, and lines delineate the 95% HPD interval of the posterior distribution for variance components in a MCMC generalised linear model.

Environmental and demographic stochasticity additively and similarly influence repeatability

We investigated the effect of environmental stochasticity (inter-institute variability) versus demographic stochasticity (intra-institute, between-treatment variability) on the (endpoint) repeatability of population growth rates. For this we fitted generalised linear mixed effect models in an MCMC framework using the R-package MCMCglmm (Hadfield, 2010). We first sequentially added terms (institute, treatment, replicate ID, measurement) to a model explaining daily population growth variation. Only treatment was fitted as a fixed effect because that was a statistically non-random variable. Most of the population growth rate variation (∼50%) was explained by differences among the institutes, followed by differences among treatments (∼10%; Figure 3A). We then used these models to estimate the relative among-replicate variance (which is 1–Repeatability) across institutes within the primary treatment only, across institutes within all treatments, and across treatments within each institute. Repeatability was constrained more by demographic variation than by environmental variation. Among-replicate variance was higher, with non-overlapping 95% HPD intervals, across treatments within VU and WUR versus across institutes (Figure 3B). Among-replicate variance across treatments within UGent and NIOO was also higher than across institutes, but in this case 95% HPD intervals overlapped (Figure 3B). Environmental and demographic heterogeneity appeared to mostly additively affect repeatability: inter-replicate variance across secondary treatments and institutes (the line marked secondary in Figure 3B with posterior mode = 0.68) was similar to the inter-replicate variance across institutes within the primary treatment (posterior mode = 0.10) plus the average of the inter-replicate variances across treatments in each of the institutes (posterior mode = 0.53).

Discussion

Parallel evolution in nature shows that environmental heterogeneity and demographic history limit evolutionary repeatability (Grant & Grant, 2002; Losos, 2011; Mahler et al., 2017; Nosil et al., 2024). Without experiments quantifying these effects, it remains unclear how repeatable evolution is across ecological contexts (Wortel et al., 2023) and how predictions can be optimised by better estimates of selection and population genetic structure (Gompert et al., 2022; Nosil et al., 2020, 2021). Most natural populations will differ in selective regimes as well as demographic history across space and time and these sources of stochasticity interact to determine repeatability (Charlesworth & Jensen, 2021; Simões et al., 2008). Therefore, to overcome the major obstacles to (more) accurate predictions, i.e., estimating selection and accounting for historical contingencies, both should be varied experimentally. Our multi-institute experiment introduced environmental and demography heterogeneity to measure random limits on repeatability and assess replicability of evolutionary experiments across heterogeneous contexts.

Only two of five institutes showed consistent increases in population growth rates on *P. megaterium* at 16 °C relative to their ancestor (Figure 2). This novel food source was chosen because it is more difficult for nematodes to consume and slows their growth (Shtonda & Avery, 2006). Most variance was explained by institute differences (Figure 3A), likely due to divergent selective regimes and microbiota (Figure S5), which had weak but detectable effects on population growth rate evolution (Figure S6, Table S5,6). Within institutes, population genetic and demographic properties were varied. The treatments likely inflated life-history variation relative to the primary treatment alone. This variation in reproductive modes, bottlenecks, demographic distribution, and the random mixing of local genotypes influences several aspects of evolutionary change, including the probability that a given genotype transfers to the next generation (i.e., drift), the rate of inbreeding (i.e., local drift), and the strength of linkage between selected and neutral loci (i.e., draft) (Charlesworth et al., 1993). Together, environmental and demographic treatments introduced heterogeneity that we quantified as among-replicate variation in our mixed model framework. We conclude that both environmental and demographic heterogeneity affect repeatability and that they do so additively. This heterogeneity acts on both sides of the predictability problem: (1) altering the strength and direction of selection (Nosil et al., 2020; Wortel et al., 2023), and (2) imposing distinct historical contingencies through institute-specific demographic and genetic trajectories (Blount et al., 2018; Travisano et al., 1995). Demographic stochasticity appears to have stronger effects on repeatability than environmental stochasticity, depending on treatment specifics. We suggest this is because demographic stochasticity directly modulates realised selection strength through variation in reproductive modes, bottlenecks, and sampling, while simultaneously shaping historical contingencies through founder effects. Consequently, forecasting evolution requires not only estimating variation in selection across environments, such as that arising from ecological differences or microbiota, but also accounting for life history, which encode historical contingencies through demographic histories and linkage patterns. Moreover, stochasticity acting across environments and populations inevitably introduces noise, leading to substantial individual variation and may limit explanatory power, even when predicting redundant traits such as fitness over brief time scales.

Variation in fitness means among institutes and treatments

The temporal change in growth rates across institutes—increasing in two institutes, two staying roughly the same, and one decreasing—reflects multiple possible processes. Identifying divergent microbiota among institutes confirmed one aspect of the expected heterogeneity in selection environments. Because the central experimental paradigm involved adaptation to a novel bacterial food source, we expected the microbial background environment to be relevant for the adaptive process. Even though our analyses revealed some statistically significant effects from microbiome diversity on fitness and fitness change through time, we conclude that these effects are either weak or institute specific. An alternative, non-mutually exclusive explanation for the among-institute variation is that adaptation to the experimental conditions in one institute may translate poorly to fitness measured in a different institute. Growth rates for the evolved populations from all institutes were measured at VU. This institute was one of the sites where population growth rates increased over time. Conducting all measurements at a single location ensured that fitness means and variances were independent of the experimenter and equipment. However, it is not unusual that replicate populations with little fitness variation under the conditions in which they evolved exhibit higher fitness variation in other environments (Blount et al., 2018). For example, local adaptation and environmental trade-offs may lower the growth rate measured in populations at the VU that have evolved elsewhere (Hereford, 2009; Kawecki & Ebert, 2004).

The variation in growth rates across treatments depended on multiple factors. As expected, final growth rates were higher in monoecious populations than in dioecious populations, most likely driven by a greater proportion of reproductive individuals; although, contributions from differences in genomic backgrounds across the A00, M00, and D00 ancestors cannot be excluded. M00 showed a higher growth rates, while A00 and D00 were similar. Because M00 differs strongly in selfing rate, affecting recombination, LD, and genetic variation, whereas the three lines have similarly divergent genomic backgrounds, the results suggest that mating system and the number of reproducing individuals matter here more for fitness than genomic background alone. Growth rates were also higher in the isolation and bleach treatments compared to the mixed and non-bleached populations. Although isolation reduces genetic variation, isolating rather than mixing the three Petri dishes within each biological replicate may be beneficial by reducing outbreeding depression (Dolgin et al., 2007) and lowering migration load (Bisschop et al., 2019; Lenormand, 2002). By synchronising developmental stages and minimising founder effects, the bleach treatment may have enhanced the expression and selection of alleles beneficial early in life, thereby maintaining adaptive potential. Contrary to our expectation that smaller effective population sizes would increase drift and reduce adaptive potential, we found no effect of the bottlenecks on mean final growth rates (see also Bisschop et al., 2022). Interestingly, despite the absence of fitness effects, the bottleneck leaves a clear genetic signature, especially the strong bottleneck. Replicates that underwent

a strong bottleneck were genetically divergent before exposure to the new environment (week 0), as evident in the broad distribution of these samples in genetic PCA space (Figure S3B, C).

Overall, we found institute differences to explain most of the variation in population final growth rates, which is likely due to a combination of the divergent background microbiota and lab-specific adaptive processes. Treatments explained less variation in mean growth rates, after accounting for institute differences. Based on these results, we conclude that treatments likely mostly affected population genetic structure with mixed effects on fitness, while institute differences likely affected selection coefficients with significant effects on fitness.

Environmental and demographic heterogeneity additively impact fitness repeatability

The D00 ancestor of our experimental populations harboured high levels of genetic variation, showed a rapid decay of linkage disequilibrium (LD) over short genetic distances, and exhibited a high degree of sign epistasis, all as a consequence of the multi-parental hybridisation (Noble et al., 2017; Teotonio et al., 2012; Theologidis et al., 2014). Selection likely acts on many different loci that may be expected to be involved in feeding on specific microbial prey. In a context of high genetic redundancy and extensive epistasis, the extent to which populations may "find" the same fitness-improving solutions within the limited time of approximately 20 generations likely depends on several factors. It is shaped not only by selection but especially by the genetic background and genetic structure within which selection operates (Weinreich et al., 2005, 2013). In line with this expectation, we show an important role for demographic stochasticity. This finding highlights that predictability of adaptive evolution across different loci is inherently limited by the contingencies associated with founder populations, echoing previous results from studies in microbes, flies, and nematodes (Blount et al., 2018; Schlötterer, 2023). A forthcoming analysis of the genetic selection response can reveal whether founder effects, variation in LD, and epistasis across the genome in response to the demographic treatments may indeed explain the observed effects on repeatability of fitness outcomes. The current results on the repeatability of fitness outcomes imply that both sources of stochasticity must be considered when predicting short-term adaptive responses to environmental change. This stochasticity likely introduces substantial noise, reducing the accuracy of such predictions.

Replicability and heterogeneity in experimental evolution

An important advantage of our approach is that we do not limit our analysis of repeatability to a single institute or a universally shared demographic and genetic background. Most experimental evolution studies are conducted in a single institute, using standardised media and equipment. These studies typically apply either constant selection coefficients (i.e., constant selection) or constant selection differentials, where the coefficient increases proportionally with the selection response (Van den Bergh et al., 2018). Although this ensures precise estimates of repeatability, it does not reflect the broader biological reality of evolution occurring across different ecological and demographic contexts. Whenever spatiotemporal heterogeneity is introduced, evolutionary outcomes diverge among replicates that differ in environment more than among replicates that experience similar conditions. This is evident from our results and supported by findings from both highly controlled and more "natural" experiments. For example, seemingly minor differences in media, glassware, and atmospheric conditions can substantially influence the outcomes of microbial evolutionary experiments (Desjardins et al., 2021; Lenski, 2023; Worthan et al., 2023). Only 3 out of 80 quantitative trait loci for biomass traits in rapeseed (*Brassica napus L.*) were stable across two experimental locations (Haelterman et al., 2024), and just 53 out of 152 behavioural responses in mice were reproducible across five institutes (Jaljuli et al., 2023). Moreover, the population of origin (genetic background) can itself be a major driving force, as shown both in this study and in previous experimental evolution work using diverse founders (Barghi et al., 2019; Otte et al., 2021; Seabra et al., 2018). Our study illustrates how environmental and demographic heterogeneity can be incorporated to increase the external validity, which makes experimental results more generalisable beyond a single laboratory setting.

In our data, we assessed among-replicate variance across similar conditions (within-institute, within-treatment), across intermediate dissimilar conditions (replicates across institutes *or* treatments), and across most dissimilar conditions (replicates across institute *and* treatments). We find that repeatability decreases accordingly and is not uniformly dependent on whether replicates vary by treatment or by institute. Statistically integrating heterogeneity in experimental evolution through the mixed model approach helped identify factors and quantify their influence on evolutionary repeatability and experimental replicability. We thus were able to compare different sources of heterogeneity quantitively to consider how they impact the uncertainty associated with extrapolating findings to broader contexts. Testing repeatability across heterogeneous contexts aids in understanding (the limits to) evolutionary predictability. Replicating evolutionary experiments across locations and experimenters, and including heterogeneous contexts, adds broader biological context and probes the external validity of replays of the tape of life.

Data and code availability

Data and R code used for analysis and visualisation are publicly available in Zenodo [https://doi.org/10.5281/zenodo.17497646] and all genetic data have been deposited in the NCBI Sequence Read Archive under BioProject accession [PRJNA1252273].

Supplementary material

Supplementary material is available online at *Evolution Letters*.

Author contributions

Conceptualization: K.B., M.T.W., K.K., M.E., J.E., A.T.G., M.E.V., T.B.. Data curation: K.B., T.B.. Formal analysis: K.B., I.R., T.B.. Funding

acquisition: K.B., I.R., M.E., J.E., A.T.G., M.E.V.. Investigation: K.B., M.T.W., K.K., J.M., M.E., J.E., A.T.G., M.E.V., T.B.. Methodology: K.B., M.T.W., K.K., J.M., S.G., J.R., L.Z.. Project administration: M.E., J.E., A.T.G., M.E.V.. Validation: K.B., T.B.. Visualization: K.B., I.R., T.B.. Writing—original draft: K.B., I.R., T.B.. Writing—review & editing: all authors.

Funding

This work was funded by The Dutch Research Council National Science Agenda (NWA-ORC 400.17.606/4175), a Flemish Research Foundation fellowship awarded to K.B. (FWO 12T5622N), I.R. was supported by NWA.1418.22.013, and L.E.Z. was supported by the Adaptive Life program of the Faculty of Science and Engineering, within GELIFES.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We would like to thank the Teotónio lab (IBENS-Paris) for providing the nematodes, and J. Teapal and the Utrecht University Large Particle Flow Cytometry Facility (UU-LPC) for their help with the BioSorter. We also acknowledge The Predicting Evolution consortium (M. Bosse, T. de Jager, L. de Jeu, A. de Visser, M. Groenen, A. Heerdink, P. Hogeweg, M. Maan, I. R. Pen, I. Smallegange, S. van der Steen, S. van Doorn, B. Voetdijk, and B. Wertheim), and the Origins Center for helpful discussions.

References

Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., Dolezal, M., Taus, T., Kofler, R., & Schlötterer , C. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biology*, *17*, e3000128. https://doi.org/10.1371/journal.pbio.3000128

Barton, N. H. (1995). Linkage and the limits to natural selection. *Genetics*, *140*, 821–841. https://doi.org/10.1093/genetics/140.2.821

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bisschop, K., Blankers, T., Mariën, J., Wortel, MT., Egas, M., Groot, A. T., Visser„ M. E., & Ellers„ J. (2022). Population bottleneck has only marginal effect on fitness evolution and its repeatability in dioecious *Caenorhabditis elegans*. *Evolution*, *76*, 1896–1904. https://doi.org/10.1111/evo.14556

Bisschop, K., Mortier, F., Etienne, R. S., & Bonte, D. (2019). Transient local adaptation and source–sink dynamics in experimental populations experiencing spatially heterogeneous environments. *Proceedings of the Royal Society B: Biological Sciences*, *286*, 20190738. https://doi.org/10.1098/rspb.2019.0738

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: replaying life's tape. *Science*, *362*, eaam5979. https://doi.org/10.1126/science.aam5979

Brooks, M. E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M., & Bolker, B.M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*, 378–400. https://doi.org/10.32614/RJ-2017-066

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*, 364–367. https://doi.org/10.1080/01621459.1974.10482955

Charlesworth, B., & Jensen, J. D. (2021). Effects of Selection at Linked Sites on Patterns of Genetic Variability. *Annual Review of Ecology, Evolution, and Systematics*, *52*, 177–197. https://doi.org/10.1146/annurev-ecolsys-010621-044528

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*, 1289–1303. https://doi.org/10.1093/genetics/134.4.1289

Chesler, E. J., Wilson, S. G., Lariviere, W. R., Rodriguez-Zas, S. L., & Mogil, J. S. (2002). Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neuroscience & Biobehavioral Reviews*, *26*, 907–923. https://doi.org/10.1016/S0149-7634(02)00103-3

Consortium, C., & elegans, S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science*, *282*, 2012–2018.

Cooper, T. F., & Lenski, R. E. (2010). Experimental evolution with *E. coli* in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evolutionary Biology*, *10*, 11 . https://doi.org/10.1186/1471-2148-10-11

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R.M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*, giab008. https://doi.org/10.1093/gigascience/giab008

Davis, P., Zarowiecki, M., Arnaboldi, V., Becerra, A., Cain, S., Chan, J., Chen, W.J., Cho, J., da Veiga Beltrame, E., Diamantakis, S., Grigoriadis, D., Grove, C.A., Harris, T.W., Kishore, R., Le, T., Lee, R.Y.N., Luypaert, M., Müller, H.M., Nakamura, C.,...Sternberg, P.W. (2022). WormBase in 2022-data, processes, and tools for analyzing *Caenorhabditis elegans*. *Genetics*, *220*, iyac003. https://doi.org/10.1093/genetics/iyac003

Desjardins, E., Kurtz, J., Kranke, N., Lindeza, A., & Richter, S. H. (2021). Beyond standardization: improving external validity and reproducibility in experimental evolution. *Bioscience*, *71*, 543–552. https://doi.org/10.1093/biosci/biab008

Dolgin, E. S., Charlesworth, B., Baird, S. E., & Cutter, A. D. (2007). Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. *Evolution*, *61*, 1339–1352. https://doi.org/10.1111/j.1558-5646.2007.00118.x

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. 4th edn. Longman.

Frankham, R., Lees, K., Montgomery, ME., England, PR., Lowe, EH., & Briscoe, D. A. (1999). Do population size bottlenecks reduce evolutionary potential? *Animal Conservation*, *2*, 255–260. https://doi.org/10.1111/j.1469-1795.1999.tb00071.x

Gillespie, J. H. (2001). Is the population size of a species relevant to its evolution? *Evolution; International Journal of Organic Evolution*, *55*, 2161–2169. https://doi.org/10.1111/j.0014-3820.2001.tb00732.x

Gompert, Z., Flaxman, S. M., Feder, J. L., Chevin, L. M., & Nosil, P. (2022). Laplace's demon in biology: models of evolutionary prediction. *Evolution*, *76*, 2794–2810.

Grant, P. R., & Grant, B. R. (2002). Unpredictable evolution in a 30-year study of Darwin's finches. *Science*, *296*, 707–711. https://doi.org/10.1126/science.1070315

Gray, J. C., & Cutter, A. D. (2014). Mainstreaming *Caenorhabditis elegans* in experimental evolution. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20133055. https://doi.org/10.1098/rspb.2013.3055

Guzella, T.S., Dey, S., Chelo, I. M., Pino-Querido, A., Pereira, VF., Proulx, S. R., & Teotónio, H. (2018). Slower environmental change hinders adaptation from standing genetic variation. *PLoS Genetics*, *14*, e1007731. https://doi.org/10.1371/journal.pgen.1007731

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, *33*, 1–22. https://doi.org/10.18637/jss.v033.i02

Haelterman, L., Louvieaux, J., Chiodi, C., Bouchet, A. S., Kupcsik, L., Stahl, A., Rousseau-Gueutin, M., Snowden, R., Laperche, A., Nesi, N., & Hermans, C. (2024). Genetic control of root morphology in response to nitrogen across rapeseed diversity. *Physiologia Plantarum*, *176*, e14315. https://doi.org/10.1111/ppl.14315

Hartig, F. (2024). DHARMa: residual diagnostics for hierarchical (multi-level /mixed) regression models. *R package version 0.4.7*.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, *173*, 579–588. https://doi.org/10.1086/597611

Jaljuli, I., Kafkafi, N., Giladi, E., Golani, I., Gozes, I., Chesler, E. J., Bogue, M.A., & Benjamini, Y., 2023). A multi-lab experimental assessment reveals that replicability can be improved by using empirical estimates of genotype-by-lab interaction. *PLoS Biology*, *21*, e3002082. https://doi.org/10.1371/journal.pbio.3002082

Karve, S. M., Daniel, S., Chavhan, Y. D., Anand, A., Kharola, S. S., & Dey, S. (2015). *Escherichia coli* populations in unpredictably fluctuating environments evolve to face novel stresses through enhanced efflux activity. *Journal of Evolutionary Biology*, *28*, 1131–1143. https://doi.org/10.1111/jeb.12640

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Lachapelle, J., Reid, J., & Colegrave, N. (2015). Repeatability of adaptation in experimental populations of different sizes. *Proceedings of the Royal Society B: Biological Sciences*, *282*, 20143033. https://doi.org/10.1098/rspb.2014.3033

Lässig, M., Mustonen, V., & Walczak, A. M. (2017). Predicting evolution. *Nature Ecology & Evolution*, *1*, 77. https://doi.org/10.1038/s41559-017-0077

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*, 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Lenormand, T., Roze, D., & Rousset, F. (2009). Stochasticity in evolution. *Trends in Ecology & Evolution*, *24*, 157–165. https://doi.org/10.1016/j.tree.2008.09.014

Lenski, R. E. (2023). Revisiting the design of the long-term evolution experiment with *Escherichia coli*. *Journal of Molecular Evolution*, *91*, 241–253. https://doi.org/10.1007/s00239-023-10095-3

Lenski, R. E., & Travisano, M. (1994). Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences*, *91*, 6808–6814. https://doi.org/10.1073/pnas.91.15.6808

Lenth, R. (2024). emmeans: estimated marginal means, aka least-squares means. *R package version 1.10.7*, [WWW Document]. URL https://cran.r-project.org/package=emmeans [accessed on 2024].

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*, (3), 18–22.

Lind, P. A. (2019). Repeatability and predictability in experimental evolution. In *Evolution, origin of life, concepts and methods*. (ed. by P. Pontarotti). Springer International Publishing, pp. 57–83.

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution*, *65*, 1827–1840. https://doi.org/10.1111/j.1558-5646.2011.01289.x

Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*, *3*, e104. https://doi.org/10.7717/peerj-cs.104

Mahler, D. L., Weber, M. G., Wagner, C. E., & Ingram, T. (2017). Pattern and process in the comparative study of convergent evolution. *The American Naturalist*, *190*, S13–S28. https://doi.org/10.1086/692648

Milcu, A., Puga-Freitas, R., Ellison, A. M., Blouin, M., Scheu, S., Freschet, G. T., Rose, L., Barot, S., Cesarz, S., Eisenhauer, N., Girin, T., Assandri, D., Bonkowski, M., Buchmann , N., Butenschoen, O., Devidal, S., Gleixner G Gessler, A., Gigon, A., Greiner, A., ...Roy, J. (2018). Genotypic variability enhances the reproducibility of an ecological study. *Nature Ecology & Evolution*, *2*, 279–287. https://doi.org/10.1038/s41559-017-0434-x

National Academies of Sciences, E. and M. (2019). Reproducibility and replicability in science. *Reproducibility and Replicability in Science*, 1–234.

Noble, L. M., Chelo, I., Guzella, T., Afonso, B., Riccardi, D. D., Ammerman, P., Dayarian, A., Carvalho, S., Crist, A., Pino-Querido, A., Shraiman, B., Rockman, M. V., & Teotónio, H. (2017). Polygenicity and epistasis underlie fitness-proximal traits in the *Caenorhabditis elegans* multiparental experimental evolution (CeMEE) panel. *Genetics*, *207*, 1663–1685. https://doi.org/10.1534/genetics.117.300406

Nosil, P., de Carvalho, C. F., Villoutreix, R., Zamorano, L. S., Sinclair-Waters, M., Planidin, N. P., Parchman, T. L., Feder, J., & Gompert, Z. (2024). Evolution repeats itself in replicate long-term studies in the wild. *Science Advances*, *10, eadl3149*. https://doi.org/10.1126/sciadv.adl3149

Nosil, P., Feder, J. L., & Gompert, Z. (2021). Biodiversity, resilience and the stability of evolutionary systems. *Current Biology*, *31*, R1149–R1153. https://doi.org/10.1016/j.cub.2021.01.022

Nosil, P., Flaxman, S. M., Feder, J. L., & Gompert, Z. (2020). Increasing our ability to predict contemporary evolution. *Nature Communications*, *11*, 5592. https://doi.org/10.1038/s41467-020-19437-x

Ohta, T. (1972). Fixation probability of a mutant influenced by random fluctuation of selection intensity. *Genetical Research*, *19*, 33–38. https://doi.org/10.1017/S0016672300014245

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P.R. O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., ... Weedon, J. (2025). vegan: community ecology package. *R package version 2.6-10*, [WWW Document]. https://cran.r-project.org/package=vegan [accessed on 2025].

Otte, K. A., Nolte, V., Mallard, F., & Schlötterer, C. (2021). The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime. *Genome Biology*, *22*, 211. https://doi.org/10.1186/s13059-021-02425-9

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peterson, R. A., & Cavanaugh, J. E. (2020). Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, *47*, 2312–2327. https://doi.org/10.1080/02664763.2019.1630372

R Core Team. (2025). *R: a language and environment for statistical computing*. R foundation for statistical computing, .

Rozen, D.E., Habets, MG. J. L., Handel, A., & de Visser, J. A.G. M. (2008). Heterogeneous adaptive trajectories of small populations on complex fitness landscapes. *PLoS ONE*, *3*, e1715. https://doi.org/10.1371/journal.pone.0001715

Schlötterer, C. (2023). How predictable is adaptation from standing genetic variation? Experimental evolution in *Drosophila* highlights the central role of redundancy and linkage disequilibrium. *Philosophical Transactions of the Royal Society B*, *378*, 20220046. https://doi.org/10.1098/rstb.2022.0046

Seabra, S. G., Fragata, I., Antunes, M. A., Faria, G. S., Santos, M. A., Sousa, V.C., Simões, P., & Matos, M. (2018). Different genomic changes underlie adaptive evolution in populations of contrasting history. *Molecular Biology and Evolution*, *35*, 549–563. https://doi.org/10.1093/molbev/msx247

Shtonda, B. B., & Avery, L. (2006). Dietary choice behavior in *Caenorhabditis elegans*. *Journal of Experimental Biology*, *209*, 89–102. https://doi.org/10.1242/jeb.01955

Simões, P., Santos, J., Fragata, I., Mueller, L. D., Rose, M. R., & Matos, M. (2008). How repeatable is adaptive evolution? The role of geographical origin and founder effects in laboratory adaptation. *Evolution*, *62*, 1817–1829.

Stiernagle, T. (2006). Maintenance of *C. elegans* [WWW Document]. *WormBook*, http://www.wormbook.org [accessed on 2006].

Teotonio, H., Carvalho, S., Manoel, D., Roque, M., & Chelo, I. M. (2012). Evolution of outcrossing in experimental populations of *Caenorhabditis elegans*. *PLoS ONE*, *7*, e35811. https://doi.org/10.1371/journal.pone.0035811

Teotónio, H., Chelo, I. M., Bradić, M., Rose, M. R., & Long, A. D. (2009). Experimental evolution reveals natural selection on standing genetic variation. *Nature Genetics*, *41*, 251–257.

Teotónio, H., Estes, S., Phillips, P. C., & Baer, C. F. (2017). Experimental evolution with *Caenorhabditis* nematodes. *Genetics*, *206*, 691–716.

Theologidis, I., Chelo, I. M., Goy, C., & Teotónio, H. (2014). Reproductive assurance drives transitions to self-fertilization in experimental *Caenorhabditis elegans*. *BMC Medicine*, *12*, 93.

Travisano, M., Mongold, J. A., Bennett, A. F., & Lenski, R. E. (1995). Experimental tests of the roles of adaptation, chance, and history in evolution. *Science*, *267*, 87–90. https://doi.org/10.1126/science.7809610

Van den Bergh, B., Swings, T., Fauvart, M., & Michiels, J. (2018). Experimental design, population dynamics, and diversity in microbial experimental evolution. *Microbiology and Molecular Biology Reviews*, *82, 10–1128*. https://doi.org/10.1128/mmbr.00008-18

Voelkl, B., Altman, N.S., Forsman, A., Forstmeier, W., Gurevitch, J., Jaric, I., Karp, N.A., Kas, M. J., Schielzeth, H., Van de Casteele, T., & Würbel, H. (2020). Reproducibility of animal research in light of biological variation. *Nature Reviews Neuroscience*, *21*, 384–393. https://doi.org/10.1038/s41583-020-0313-3

Voelkl, B., & Würbel, H. (2021). A reaction norm perspective on reproducibility. *Theory in Biosciences*, *140*, 169–176. https://doi.org/10.1007/s12064-021-00340-y

Wahl, L. M., Gerrish, P. J., & Saika-Voivod, I. (2002). Evaluating the Impact of Population Bottlenecks in Experimental Evolution. *Genetics*, *162*, 961–971. https://doi.org/10.1093/genetics/162.2.961

Weber, K. E. (1990). Increased selection response in larger populations. I. Selection for wing-tip height in *Drosophila melanogaster* at three population. *Genetics*, *125*, 579–584. https://doi.org/10.1093/genetics/125.3.579

Wein, T., & Dagan, T. (2019). The effect of population bottleneck size and selective regime on genetic diversity and evolvability in bacteria. *Genome biology and evolution*, *11*, 3283–3290.

Weinreich, D. M., Lan, Y., Wylie, C. S., & Heckendorn, R. B. (2013). Should evolutionary geneticists worry about higher-order epistasis? *Current Opinion in Genetics & Development*, *23*, 700–707. https://doi.org/10.1016/j.gde.2013.10.007

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, *59*, 1165–1174.

Windels, E. M., Fox, R., Yerramsetty, K., Krouse, K., Wenseleers, T., Swinnen, J., Matthay, P., Verstraete, L., Wilmaerts, D., Van den Bergh, B., & Michiels, J. (2021). Population bottlenecks strongly affect the evolutionary dynamics of antibiotic persistence. *Molecular Biology and Evolution*, *38*, 3345–3357. https://doi.org/10.1093/molbev/msab107

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*, 1–13. https://doi.org/10.1186/s13059-019-1891-0

Wortel, M.T., Agashe, D., Bailey, S.F., Bank, C., Bisschop, K., Blankers, T., Cairns, J., Sandro Colizzi, E., Cusseddu, D., Desai, M.M., van Dijk, B., Egas, M., Ellers, J., Groot, A.T., Heckel, D. G., Johnson, M.L., Kraaijeveld, K., Krug, J., Laan, L.,...Pennings, P.S. (2023). Towards evolutionary predictions: current promises

and challenges. *Evolutionary Applications*, *16*, 3–21. https://doi.org/10.1111/eva.13513

Worthan, S. B., McCarthy, R. D. P., & Behringer, M. G. (2023). Case studies in the assessment of microbial fitness: seemingly subtle changes can have major effects on phenotypic outcomes. *Journal of Molecular Evolution*, *91*, 311–324. https://doi.org/10.1007/s00239-022-10087-9

Received: 18 November 2025. **Revised:** 18 February 2026. **Accepted:** 30 March 2026