Received: 31 July 2021 | Accepted: 19 September 2021

DOI: 10.1111/1365-2656.13598

Journal of Animal Ecology | BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Understanding climate change response in the age of genomics

Selection on growth rate and local adaptation drive genomic adaptation during experimental range expansions in the protist *Tetrahymena thermophila*

Felix Moerman[1,2,3,4] | **Emanuel A. Fronhofer**[4] | **Florian Altermatt**[1,2] | **Andreas Wagner**[1,3,5,6]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland

[2]Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

[3]Swiss Institute of Bioinformatics, Quartier Sorge—Bâtiment Génopode, Lausanne, Switzerland

[4]ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

[5]The Santa Fe Institute, Santa Fe, NM, USA

[6]Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch, South Africa

Correspondence
Felix Moerman
Email: felix.moerman@ieu.uzh.ch
Andreas Wagner
Email: andreas.wagner@ieu.uzh.ch

Present address
Felix Moerman, Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh, UK

Funding information
H2020 European Research Council, Grant/Award Number: 739874; University of Zurich URPP Evolution in Action; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 31003A_172887 and PP00P3_179089

Handling Editor: David Berger

Abstract

1. Populations that expand their range can undergo rapid evolutionary adaptation of life-history traits, dispersal behaviour and adaptation to the local environment. Such adaptation may be aided or hindered by sexual reproduction, depending on the context.
2. However, few empirical and experimental studies have investigated the genetic basis of adaptive evolution during range expansions. Even less attention has been given to the question how sexual reproduction may modulate such adaptive evolution during range expansions.
3. We here studied genomic adaptation during experimental range expansions of the protist *Tetrahymena thermophila* in landscapes with a uniform environment or a pH gradient. Specifically, we investigated two aspects of genomic adaptation during range expansion. First, we investigated adaptive genetic change in terms of the underlying numbers of allele frequency changes from standing genetic variation and *de novo* variants. We focused on how sexual reproduction may alter this adaptive genetic change. Second, we identified genes subject to selection caused by the expanding range itself, and directional selection due to the presence or absence of the pH gradient. We focused this analysis on alleles with large frequency changes that occurred in parallel in more than one population to identify the most likely candidate targets of selection.
4. We found that sexual reproduction altered adaptive genetic change both in terms of *de novo* variants and standing genetic variation. However, sexual reproduction affected allele frequency changes in standing genetic variation only in the absence of long-distance gene flow. Adaptation to the range expansion affected genes involved in cell divisions and DNA repair, whereas adaptation to the pH gradient additionally affected genes involved in ion balance and oxidoreductase reactions. These genetic changes may result from selection on growth and adaptation to low pH.

5. In the absence of gene flow, sexual reproduction may have aided genetic adaptation. Gene flow may have swamped expanding populations with maladapted alleles, thus reducing the extent of evolutionary adaptation during range expansion. Sexual reproduction also altered the genetic basis of adaptation in our evolving populations via *de novo* variants, possibly by purging deleterious mutations or by revealing fitness benefits of rare genetic variants.

KEYWORDS
gene ontology, life history, pH, range expansions, Tetrahymena, whole genome resequencing

1 | INTRODUCTION

Range expansions and biological invasions are increasingly common in many species. They often result from anthropogenic disturbances or species introductions (Chen et al., 2011; Kowarik, 2003; Parmesan et al., 1999). Because range expansions can have dramatic consequences on species distributions and ecosystems (e.g. Grosholz, 2002; Phillips et al., 2009), it is important to understand their ecological and evolutionary consequences. Theory predicts that expanding populations can undergo adaptation through two major mechanisms. First, spatial sorting of individuals is expected to lead to increased dispersal behaviour at the range edge (Phillips & Perkins, 2019; Shine et al., 2011). Second, natural selection at the low population density range edge leads to selection for an *r*-selected life-history strategy (Burton et al., 2010; Phillips, 2009; Phillips et al., 2010). Recent studies have demonstrated that expanding populations can indeed rapidly evolve their dispersal behaviour (Brown et al., 2007; Ochocki & Miller, 2017; Williams et al., 2016) and life-history strategy (Fronhofer & Altermatt, 2015; Phillips, 2009). Additionally, they can evolve specific adaptations to local conditions, such as temperature and novel food sources (Szücs et al., 2017; Van Petegem et al., 2016). Because selection pressures during range expansion can be strong, populations at a range edge may adapt quickly (Sakai et al., 2001). However, evolution during range expansion does not necessarily affect both dispersal behaviour and life-history traits, especially when these traits trade-off with each other (Burton et al., 2010; Clarke et al., 2019).

A population's potential to adapt may be constrained if the adapting traits have a complex genetic architecture (Hansen, 2006; Mackay, 2001). For example, when multiple traits under selection are genetically not independent from each other, selection may act less efficiently on each individual trait (Walsh & Blows, 2009). One important mechanism that can help overcome evolutionary constraints imposed by genetic architecture is the reshuffling of existing genetic variation through sexual reproduction (Bell, 1982; Maynard-Smith, 1978; Otto & Lenormand, 2002). This benefit of sexual reproduction has been demonstrated repeatedly (Colegrave, 2002; Luijckx et al., 2017; McDonald et al., 2016; Petkovic & Colegrave, 2019), and is especially pronounced in the presence of high genetic diversity (Lachapelle & Bell, 2012). Increased genetic variation or an influx of locally adapted genes can also increase the success of invasions and range expansions (Currat et al., 2008; Lavergne & Molofsky, 2007; Schmeller et al., 2005). However, sexual reproduction during range expansions can also hinder adaptation, for example in the process of 'gene swamping'. Gene swamping occurs when populations expand their range against an environmental gradient, and when genes ill-adapted to the range edge flow from a high-density core population to a low-density population at the range edge. In asexual populations, such maladapted individuals may simply disappear, as they fail to survive. In contrast, sexual populations may experience a general decline in fitness, when enough individuals move to the range edge. According to the gene swamping hypothesis, gene flow can decelerate adaptation of a population at the range edge (García-Ramos & Kirkpatrick, 1997; Haldane & Ford, 1956; Kirkpatrick & Barton, 1997; Polechová & Barton, 2015; Polechová, 2018), as recently demonstrated both in an experimental study (Moerman, Fronhofer, et al., 2020b) and a field study (Bachmann et al., 2020).

However, little is known of the genetic mechanisms behind range expansions. With few exceptions (Bosshard et al., 2020), studies on the evolutionary genetics of range expansions have focused on genetic population structure and diversity (e.g. Bors et al., 2019; Swaegers et al., 2013. This work has demonstrated that, first, populations at the range edge typically harbour less genetic variation than the core populations (Excoffier et al., 2009). Second, genetic drift can strongly affect populations at a range edge, which can lead to 'allele surfing', the increase in frequency and potential fixation of neutral or maladaptive variants at a range edge (Excoffier et al., 2009; Klopfstein et al., 2006). Both these mechanisms can constrain the adaptive potential at a range edge. Despite such work, the genetic basis of phenotypic evolution during range expansions remains poorly understood. For example, we know little about the kinds of genes that are involved in phenotypic evolution during range expansion. We know even less about how gene swamping may alter such evolution.

We here studied the population genomics of previously published range expansion experiment with the protist *Tetrahymena thermophila* (Moerman, Fronhofer, et al., 2020b). In this experiment, we had experimentally investigated the gene swamping hypothesis, by assessing how sexual reproduction and gene flow

affected evolution during range expansion either into a uniform environment or into a pH gradient. We chose a pH gradient for two reasons. First, pH is an important environmental stressor linked to ocean acidification (Caldeira & Wickett, 2003; Raven et al., 2005; Zeebe et al., 2008). Second, pH can easily be manipulated experimentally. The experiment showed that sexual reproduction aided adaptation during range expansion, but only in the absence of gene flow. We here performed whole genome sequencing of pooled *T. thermophila* populations from this experiment. We identified the populations with the strongest phenotypic changes, pooled individuals from these populations and sequenced the genomic DNA of the resulting pools. We then investigated how adaptive genetic change (i.e. genetic changes from standing variation and *de novo* variants) was affected by sexual reproduction, gene flow and a pH gradient. We predicted that the amount of genetic change reflects the amount of phenotypic change, and that sexual reproduction could facilitate genetic change by bringing together adaptive variants, but only in the absence of gene swamping. Furthermore, we identified genes involved in adaptation to the range expansion itself and to the pH gradient, focusing on selective sweeps of large effect alleles that occurred in parallel in multiple populations. Here we predicted that adaptation during range expansion may act on life-history traits or dispersal behaviour, whereas adaptation to the pH gradient may act predominantly on the ability to cope with acidity, either through ion maintenance or cell membrane function.

2 | MATERIALS AND METHODS

2.1 | Study organism and experimental evolution

Tetrahymena thermophila is a freshwater ciliate commonly used in ecological and evolutionary experiments (Altermatt et al., 2015; Cairns et al., 2020; Collins, 2012; Moerman, Fronhofer, et al., 2020b), including in studies of range expansions (Fronhofer & Altermatt, 2015; Giometto et al., 2014, 2017). *Tetrahymena* is characterized by nuclear dualism. That is, cells carry a polyploid ($n = 45$; 225 chromosomes) macronucleus and a diploid micronucleus (five chromosomes) (Lynn & Doerder, 2012). *Tetrahymena* usually reproduces asexually, but can be induced to reproduce sexually through starvation (Cassidy-Hanley, 2012). During asexual reproduction, both nuclei are copied, but the chromosome copies of the macronucleus are partitioned randomly among daughter cells (Ruehle et al., 2016). We used ancestral and evolved *Tetrahymena* populations from a range expansion experiment by Moerman, Fronhofer, et al. (2020b) to investigate the genomic basis of adaptation during range expansions. We explain the most pertinent details of this experiment here and refer the reader for other details to Moerman, Fronhofer, et al. (2020b).

Briefly, in this experiment, we mixed four phenotypically divergent (Moerman, Arquint, et al., 2020a) clonal strains of *T. thermophila* to create a genetically diverse ancestral population. For details on strain identities and the mixing of the ancestral population, see Section S1.1 in Supporting Information. Subsequently, we allowed 40 replicates of this ancestral population to expand their range for 10 weeks (250 generations). We kept all stock cultures and populations during experimental evolution in a climate room at 20°.

During experimental range expansion, we used an established system of two-patch landscapes (Fronhofer & Altermatt, 2015) to emulate an expanding range edge. These two-patch landscapes consist of two Sarstedt tubes (25 ml) that contain 15 ml of modified Neff medium (Cassidy-Hanley, 2012), and are connected by an 8-cm silicone tube. A plastic clamp is used to close this tube. When open, the tube allows cells to actively disperse between the two patches (see Figure 1), during several distinct dispersal events. If during such dispersal events cells moved from the first patch (home patch) to the second patch (target patch), we propagated the experiment by transferring the contents of the target patch to a new two-patch landscape. Otherwise, we transferred the contents of the home patch to a new two-patch landscape. To emulate the range core, we maintained the ancestor clones in separate glass tubes, in which they experienced continuously high population densities and slow division rates. These conditions experienced by the 'range core' populations are identical to the conditions experienced by these clones while being maintained in the laboratory. We initiated the experiment by inoculating one patch of the two-patch landscape with 200μL of the ancestral population. During range expansion, we controlled three factors. First, we controlled the abiotic environment, where populations either experienced a 'uniform' environment, with a pH equal to 6.5, or a 'pH-gradient' environment, where the pH gradually decreased from 6.5 of 4.0 during range expansion, to emulate an environmental pH gradient. We determined this pH range as the pH of fresh Neff medium (pH 6.5), and the lowest pH allowing cells to grow (pH 4.0). Second, we controlled the reproductive mode, with 'asexual' populations experiencing only asexual reproduction during the experiment, whereas 'sexual' populations also reproduced mainly asexually, but experienced four sexual reproduction events during the experiment. We induced sex by transferring populations to starvation medium, but placed 'asexual' populations on a shaker, as this prevents cells from conjugating and therefore from mating. Lastly, we controlled long-distance gene flow from the range core to the edge, where either long-distance gene flow was 'absent', meaning we did not impose long-distance gene flow, or gene flow was 'present', meaning that we imposed four long-distance gene flow events, by replacing part of the population at the range edge with individuals from the range core. Altogether, our experiment thus used eight treatments. For each treatment, we evolved five replicate populations, resulting in a total of 40 populations.

We then subjected the populations to 10 weeks of experimental evolution, during which we repeated the same cyclical procedure every 14 days. Specifically, we subjected each population to three dispersal events, which took place on days 1, 3 and 5 of the 14-day cycle. After the third dispersal event, we subjected populations of the appropriate treatment groups to a long-distance gene flow and/

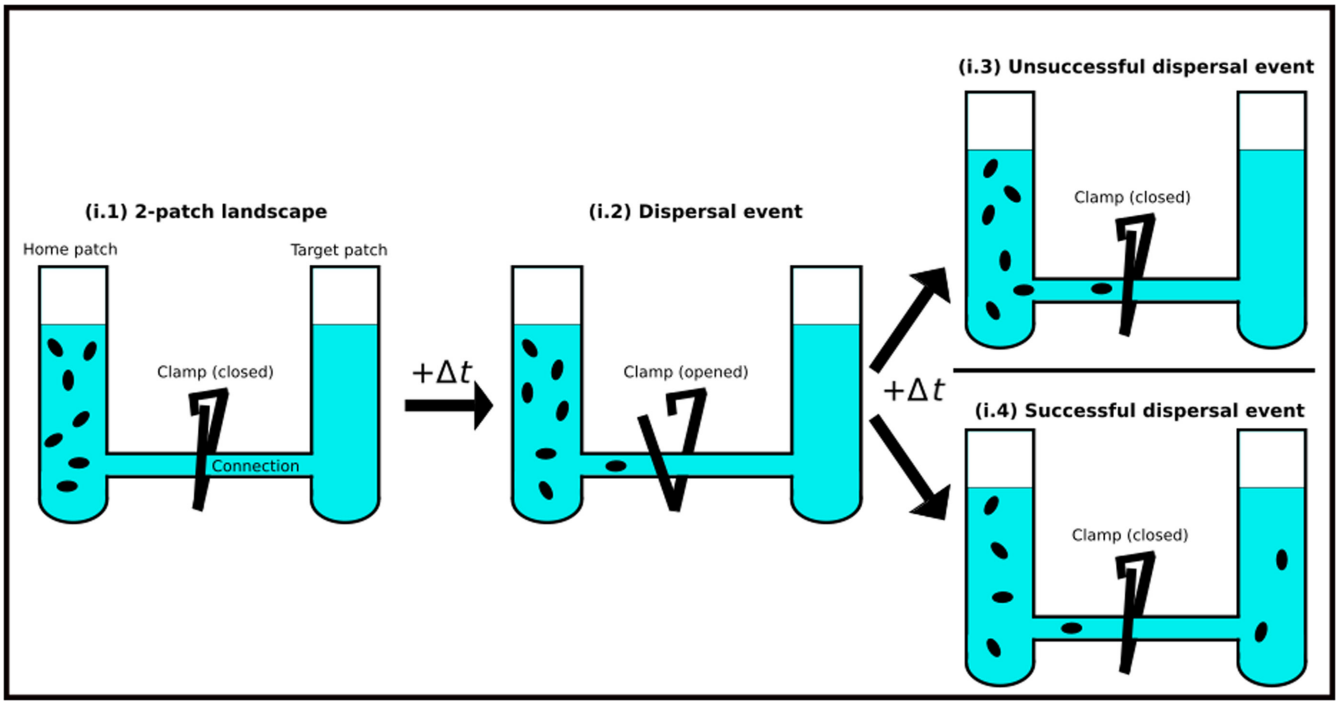

FIGURE 1 Two-patch landscapes. i.1: Two-patch landscapes consist of connected tubes. i.2: To initiate dispersal, the connection is opened for 1 hour, allowing cells to actively swim from the home patch to the target patch. i.3: If we found no cells in the target patch (unsuccessful dispersal), we transferred the contents of the home patch to a new two-patch landscape. i.4: If we did detect cells in the target patch (successful dispersal), we transferred the content of the target patch to a new two-patch landscape

or sexual reproduction event on day 8. Lastly, we subjected the populations to two additional dispersal events on days 10 and 12. Consequently, our experiment included four distinct gene flow and sexual reproduction events, in weeks 2, 4, 6 and 8 of the experiment. Over the entire experiment, population densities remained high ($10^4 - 10^6$ individuals). During the dispersal events, between 1% and 20% of our populations typically dispersed such that dispersal never led to extreme bottlenecks. Therefore, genetic drift likely did not affect the evolution of large effect genetic variants (Hartl & Clark, 2006), which were the focus of this study, but may have affected the evolutionary trajectories of small effect alleles. In consequence, genetic changes are likely due to adaptive evolution, or hitchhiking. For a detailed description and a visual representation of the experimental design, see Supporting Information Sections S1.3 and S1.4.

2.2 | Common garden and growth assessment

After a common garden phase of 18 generations to reduce epigenetic and maternal effects, we assessed the fitness change of the evolved populations compared to the ancestral population. To do so, we performed growth assays for the evolved and ancestral populations under multiple pH values (pH 6.5, 6.0, 5.5, 5.0, 4.5, 4.0, 3.5 and 3.0). We then fit a Beverton–Holt population growth model (Beverton & Holt, 1993) to infer the intrinsic population growth rate r_0 (see Supporting Information Section S1.5). We then calculated the change in fitness that occurred in each evolved population as the change in the population's intrinsic growth rate r_0 relative to the ancestral population. More specifically, we calculated the log-ratio response as

$$ln\left(\frac{r_0,e}{r_0,a}\right), \quad (1)$$

where r_0,e and r_0,a are the intrinsic population growth rate of the evolved population and the ancestral population, respectively.

2.3 | Sequencing and identification of variants

To identify candidate loci for genetic adaptation, we performed population-level sequencing ('PoolSeq'; Schlötterer et al., 2014).

We only wished to sequence the macronuclear DNA, because only the expressed macronuclear DNA affects the phenotype. We thus first separated macronuclei from micronuclei. Following nuclear separation, we extracted DNA from each population. To obtain statistically sound results, the recommended sequencing depth for whole genome population resequencing is between 50X and 200X (Kofler & Schlötterer, 2014; Schlötterer et al., 2014). Because of the large genome size of *T. thermophila* (105 Mb; Eisen et al., 2006), it was only possible to sequence a limited number of populations at this depth. We therefore selected two populations for genomic sequencing from each treatment group, based on the fitness change these populations experienced. That is, we

sequenced the populations with the largest increase in fitness, because they were the most likely candidates to show adaptive genetic changes. In this way, we obtained sequencing data for 16 (= 2 × 8 treatments) populations. Additionally, we sequenced the genome of the four ancestor clones that we used to create the ancestral population. Next, we mapped the obtained sequences to the *T. thermophila* reference genome, and performed variant calling to identify genetic variants. In analysing the resulting data, we studied standing genetic variation and its allele frequency change, as well as variants that had occurred *de novo*. In the next sections, we describe each of these steps in detail.

2.3.1 | Nuclear separation and DNA extraction

We separated macronuclei and micronuclei of the ancestral clones and of all 34 surviving evolved populations using a differential centrifugation protocol adapted from Sweet and Allis 2006; see Supporting Information Section S2). We obtained for every population three pellets containing macronuclei (pellets P1, P2 and P3). We extracted DNA from each of the pellets using a Qiagen® DNeasy Blood and Tissue kit (Cat No./ID: 69506), following the manufacturer's protocol (https://www.qiagen.com/us/resources/resourcedetail?id=68f29296-5a9f-40fa-8b3d-1c148d0b3030&lang=en). After DNA extraction, we determined the quality and quantity of the DNA using a Nanodrop ND-1000 spectrophotometer and selected for each population the pellet with the highest quality (see Supporting Information Section S3). We stored DNA samples in a − 80° freezer until sequencing. Library preparation and sequencing was performed by the Functional Genomics Center Zürich, which prepared 2 × 150 bp paired-end libraries and sequenced them using the Illumina Novaseq 6000 platform. We aimed to obtain population-level DNA sequences at approximately 100× genome coverage (see Supporting Information Section S4.2 for coverage statistics).

2.3.2 | Bioinformatic pipeline

We trimmed the reads of all fastq files using Trimmomatic version 0.39 (Bolger et al., 2014) to remove Illumina adapter sequences (ILLUMINACLIP option) and low-quality segments (SLIDINGWINDOW:4:15). Additionally, we removed any reads falling below a length of 36 bp (MINLEN:36). After trimming, we mapped the reads for each of the 20 populations to the *Tetrahymena thermophila* macronuclear reference genome (Eisen et al., 2006) using the Burrows–Wheeler aligner (Li & Durbin, 2009) version 0.7.17-r1188. On average, 97.7% of the reads aligned to the reference genome (see Supporting Information Section S4.1 for mapping rates of all populations).

We then called genetic variants using the BCFtools multiallelic caller (Danecek et al., 2016). We first called variants to identify sites where the 4 ancestral and 16 evolved populations differed from the reference genome, and stored the resulting allele counts.

We mapped all variants to genes using bedtools (v2.27.1; Quinlan & Hall, 2010). To this end, we first obtained the genome annotation file from the *T. thermophila* reference genome (Eisen et al., 2006), and filtered this file to only keep entries corresponding to protein-coding genes. We then used the intersect function in bedtools to map the variants to the gene entries.

2.3.3 | Quantifying allele frequencies

To detect changes in allele frequencies, we first calculated the frequency of each allele present in the ancestral population (expected allele frequency). Because we had sequenced the ancestor clones individually, we started by identifying all positions along the genome where at least one of the ancestral clones differed from the reference genome. We then calculated for these positions the allele frequency in the ancestral population (see Supporting Information Section S5). Subsequently, we filtered these positions to exclude positions with low sequencing quality (see Supporting Information Section S6.1 for details). We then calculated the allele frequencies for the 16 evolved populations at all positions along the genome that had not been eliminated by filtering. For each of the evolved populations, this resulted in a list of positions where genetic variation had initially been present in the ancestral population.

2.3.4 | Identifying *de novo* variants

To detect *de novo* variants, we kept only those variants that differed from the reference genome in an evolved population, but not so in any of the four ancestral clones. We then removed any positions with low sequencing quality (see Supporting Information Section S6.2).

2.4 | Statistical analyses

We performed statistical analyses, unless otherwise specified, using the R statistical language version 4.0.2 (R Core Team, 2020).

2.4.1 | Analysing fitness changes

We applied the fitness assay described in Moerman, Fronhofer, et al. (2020b) to the populations included in the genomic analysis. Briefly, we determined the change in the intrinsic population growth rate r_0 of each evolved population relative to the ancestral population in the pH conditions experienced during range expansion (pH 6.5 for populations expanding into a uniform environment; pH 4.0 for populations expanding into a pH gradient). For details, see Section 11 in Supporting Information.

2.4.2 | Quantifying frequency changes in standing genetic variation

We determined for all polymorphic sites in the ancestral population if the alleles at the site had experienced a significant change in

frequency. To do so, we performed a binomial test comparing the number of reference and non-reference reads in evolved populations with the expected allele frequency of the ancestral population. Because we aimed at identifying the most likely candidates of selection, we used a *p*-value of 0.001 in this test, to which we applied a Bonferroni correction to avoid falsely detecting allele frequency changes due to multiple testing. We therefore considered only variants with a *p*-value smaller than 10^{-10} as significantly different from the expectation. Such a stringent cut-off implies a low false-positive rate, but it also implies that we may not detect a large number of weakly selected alleles (Kofler & Schlötterer, 2014).

In a preliminary analysis of standing genetic variation, we observed that the ancestor clone SB3539 showed a number of anomalies that made its exclusion from further analysis desirable. First, this clone contained approximately 30 times more non-reference alleles than the other ancestral clones and than each of the evolved populations (Table S7). Second, clone SB3539 deviated strongly in its transition/transversion ratio (clone SB3539: 0.9344; other populations 0.6544–0.7611; Table S7). Furthermore, selection acted very strongly against this clone as a whole, as evidenced by post-evolution allele frequencies that were dramatically different from those in all other populations (Figure S4). Relatedly, all evolved populations showed an extremely high number of variants (approx. 1.1e4) that changed in their frequency by approximately 0.25, a value that corresponds closely to the initial frequency of clone SB3539 (27% of the ancestral population; Table S1). To avoid potential confounding effects due to this clonal selection, we limited our analyses to those alleles whose frequency changed by an amount that is larger than that could possibly be caused by selection against this clone. Specifically, we only analysed alleles whose frequency changed by a value of 0.3 or higher (rounded up from 0.27, the starting frequency of clone SB3539).

Next, we assessed whether the number of alleles whose frequency changed significantly was affected by reproductive mode, gene flow and the pH gradient. To do so, we counted for every evolved population the number of variants that significantly changed in frequency. In this analysis, we applied various cut-offs (0.3, 0.4, 0.5, 0.6, 0.7 and 0.8) for the magnitude of the allele frequency change, that is, the minimum absolute difference between the expected and the observed allele frequency, required for variants to be included. We fit a Bayesian generalized linear mixed model with a Poisson distribution to the resulting data, using the 'Rstan' and 'statistical rethinking' packages (McElreath, 2015; Stan Development Team, 2020). We created a model in which reproduction, gene flow, abiotic conditions and the cut-off value are all allowed to interact as fixed effects, and used the identity of each replicate evolved population as a random effect. We report posterior distributions (means and 95% confidence intervals) for the parameter estimates. Additionally, we plotted the population's fitness change against the number of alleles that changed significantly in frequency. We calculated the Pearson correlation between the two metrics (paired sample correlation test), and did so for different cut-off values of the magnitude of allele frequency change.

2.4.3 | Quantifying *de novo* variants

To assess how reproductive mode, gene flow and the pH gradient affect the evolutionary dynamics of *de novo* variants in our populations, we counted the number of novel variants in each of the evolved populations. To this end, we used eight cut-off values of increasing allele frequencies, ranging from 0.1 to 0.9. For each of these values, we counted the number of *de novo* variants that had reached an allele frequency larger than the cut-off. We then fit a Bayesian linear mixed model with 'Rstan' and 'statistical rethinking' package (McElreath, 2015; Stan Development Team, 2020) using reproduction, gene flow, presence of a pH gradient and cut-off as fixed effects, and replicate population as a random effect. We report posterior distributions (means and 95% confidence intervals) for the parameter estimates.

2.4.4 | General adaptations during range expansion

To identify allele frequency changes due to general adaptation—they were subject to selection regardless of reproductive mode, gene flow or the pH gradient—we focused on alleles whose frequency increased or decreased consistently across all or most populations. Specifically, we identified allelic variants that were present in the ancestral population, that changed their frequency not merely as a result of the global selection we observed against clone SB3539 (allele frequency change > 0.3) and that changed significantly in frequency (*p*-value< 10^{-10}). Among the remaining alleles, we then considered only those whose frequency changed in the same direction for 75% (12/16) of all evolved populations. We call the resulting dataset the 'general adaptation dataset'.

2.4.5 | Gradient-specific adaptations: Cochran–Mantel–Haenszel test

To identify directional allelic changes due to the presence/absence of the pH gradient, we performed a Cochran–Mantel–Haenszel test (PoPoolation2-package; Kofler et al., 2011). This test compares pairs of populations and it can identify alleles that show consistently different frequencies. To account for the effect of reproduction and gene flow, we always compared pairs of populations with the same reproduction or gene flow treatment. For example, we compared an asexual population without gene flow expanding into a pH gradient with another asexual population without gene flow but expanding into a uniform environment (see Supporting Information Section S8.1). With this approach, we aimed at detecting loci that display consistent differences between populations expanding into a pH gradient and populations expanding into a uniform environment, independent of gene flow and reproductive mode. To account for multiple testing, we applied a Bonferroni correction (Bland & Altman, 1995). Because our genomes harboured approximately 10^5 genetically variable positions, and we wished to

use a conservative p-value of 0.001, we only considered positions with a p-value smaller than 10^{-8}. We refer to this dataset as the 'gradient-specific adaptation' dataset. We again focused on alleles within protein-coding genes and examined the functional annotations of these genes.

2.4.6 | Gene ontology enrichment

To assess whether allele frequency changes occur in genes associated with specific functions, we performed a gene ontology (GO) term enrichment analysis using GOWINDA (version 1.12; Kofler & Schlötterer, 2012). This analysis identifies terms describing the molecular or biological function and cellular component associated with each gene, and then assesses which of such terms are over-represented in a dataset. We first queried the Uniprot Knowledgebase to obtain all entries associated with the *T. thermophila* reference genome (TaxID: 312017). This resulted in a list of genes encoded in the Tetrahymena genome together with GO terms associated with each gene. We then performed a GOWINDA analysis using this dataset as a reference database, to identify enriched gene ontology terms. We performed this analysis for two sets of genes: the genes associated with general adaptations and the genes associated with gradient-specific adaptations.

We kept only those GO terms with an FDR (false discovery rate) value < 0.05, following a Benjamini–Hochberg correction (Benjamini & Hochberg, 1995). Next, we grouped the enriched GO terms into 13 categories of cellular mechanisms (see Supporting Information Table S15). We then compared the proportion of all GO terms associated with the 13 categories between general adaptation genes and gradient-specific genes, using a χ^2-test. Finally, we performed a post-hoc test to assess which categories are differently enriched in the gradient-specific genes and the general adaptation genes (Ebbert, 2019).

3 | RESULTS AND DISCUSSION

3.1 | Genetic basis of adaptation

To investigate how gene swamping may affect genetic adaptation during range expansions, we here assessed how sexual reproduction altered phenotypic change and genetic change through standing variation and *de novo* variants, in landscapes with/without a pH gradient and in the presence/absence of gene flow.

3.1.1 | Gene swamping affects fitness change

We repeated the analysis described in Moerman, Fronhofer, et al. (2020b) using only those 16 populations that we included in the genetic analysis.

We observed a positive change in fitness for all populations (Figure 2). There was a significant interaction between sexual reproduction and the presence of gene flow ($\chi^2_{1,16} = 14.36$, $p = 0.003$), suggesting that sexual reproduction only provided benefits if gene flow was absent. These results are congruent with those described in Moerman, Fronhofer, et al. (2020b).

3.1.2 | Frequencies of *de novo* variants change most strongly in sexually reproducing populations

In the second analysis, we investigated the presence of *de novo* variants that rose to detectable allele frequencies. We found a total of 27,766 *de novo* variants across all 16 evolved populations. They reached frequencies between 10% and 100%. Only 0.02% of these *de novo* variants went to fixation. We found that 31.73% of *de novo* variants occurred in protein-coding genes. Given that 47.87% of the *Tetrahymena* genome is protein coding (Eisen et al., 2006), *de novo* variants are less likely to occur in protein-coding regions than expected by chance (exact binomial test, $p < 0.001$). Any one non-coding *de novo* variants may have hitch-hiked to high allele frequency due to its physical proximity to an adaptive coding variant (Charlesworth et al., 2000). Alternatively, it may itself have been subject to positive selection, for example by beneficially altering gene regulation through DNA-binding sites of transcription factors (Latchman, 1997; Spitz & Furlong, 2012). Most (82.30%) *de novo* variants were indels. Significantly fewer indels (29.14%) than SNPs (46.36%) occurred in protein-coding regions (Pearson Chi-squared test: $\chi^2_1 = 130.89$; $p < 0.001$). Only *de novo* indels but not SNPs were less likely to occur in protein-coding regions than expected by chance (indels: 29.14%; $p < 0.001$; SNPs; 46.36%, $p = 0.066$, exact binomial test). These observations are consistent with the notion that indels inside protein-coding regions are more likely to be deleterious than SNPs, by disrupting open reading frames.

We then studied the role of reproductive mode, gene flow and the pH gradient as explanatory variables for the number of *de novo* variants (response variable). We observed more variants in sexually reproducing populations (Figure 3, blue lines and circles) than in asexual populations (Figure 3, yellow lines and circles), but only at low values for the cut-off (Reproduction × Cut-off). This effects appeared to be most pronounced in populations expanding into a pH gradient and in the presence of gene flow; however, such differences were too small to detect statistically. See Supporting Information Section S13 for posterior distributions.

Because most new mutations are neutral or deleterious (Loewe & Hill, 2010), the accumulation of mutations in a population often leads to a fitness decrease. Sexual reproduction can help avert this decrease by purging deleterious mutations and separating them from rare beneficial mutations (Fisher, 2000; Johnson & Barton, 2002; Maynard-Smith, 1978; McDonald et al., 2016; Muller, 1932; Peck, 1994). This may explain why more *de novo* variants rose to detectable allele frequency in sexually reproducing populations. Because the chromosomes of the polyploid macronucleus are partitioned randomly

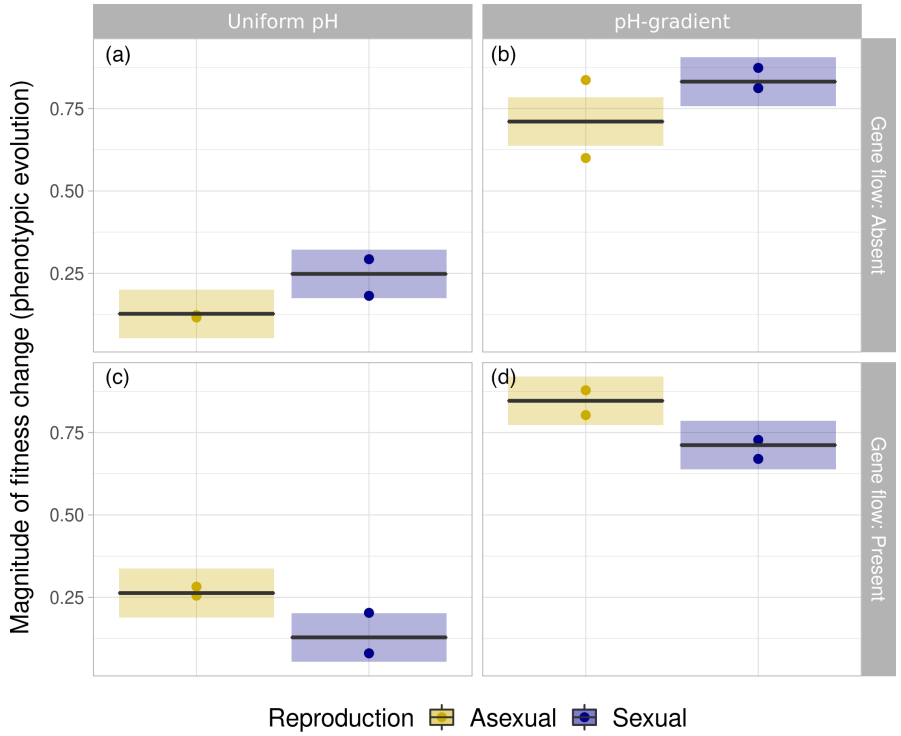

FIGURE 2 Gene swamping affects fitness changes during range expansion. Fitness change calculated as the ratio of the logarithm (base 2) of the intrinsic population growth rate (r_0) of an evolved population and the intrinsic population growth rate (r_0) of the ancestral population. Circles represent data for individual populations. Black lines and shaded areas represent mean predictions and 95% confidence intervals of the best model. Colours represent mode of reproduction (yellow = asexual, blue = sexual). Left subplots (panels a and c) show data and model predictions for populations expanding into a uniform environment, right subplots (panels b and d) for populations expanding into a pH gradient. Upper subplots (panels a and b) show data and model predictions for populations where gene flow is absent, and lower subplots (panels c and d) for populations where gene flow is present. Adapted from Moerman, Fronhofer, et al. (2020b)

among the offspring of asexually reproducing *T. thermophila*, different offspring of the same individual may vary in the copy numbers of specific variants (Ruehle et al., 2016). Therefore, in asexual *T. thermophila* populations, selection likely acts against individuals carrying deleterious variants at a high copy number. In contrast, sexual recombination may reduce this mutational load by separating maladaptive mutations from neutral or beneficial mutations (Maynard-Smith, 1978). Given the large number of variants we identified, an alternative explanation could be that some genetic variants we had classified as *de novo* were actually present at the start of our experiment, but at an undetectably low frequency. As discussed in Hansen (2006), some such variants may be beneficial, but these benefits may be hidden, for example if a variant co-occurs with a deleterious variant that offsets its advantage. Such benefits can be revealed by sexual reproduction. Therefore, some initially very rare variants may have risen to high frequency in our populations following sexual reproduction.

3.1.3 | Reproduction alters allele frequency change when gene flow is absent

Next, we assessed how reproductive mode, gene flow and the presence of a pH gradient affected standing genetic variation. After correcting for multiple testing and clonal selection on one of our ancestral clones (see Section 2.4), we observed significant allele frequency changes in 13,278 alleles. Only 29.39% of alleles occurred in protein-coding regions of the genome. The percentage of macronuclear DNA predicted to occur in protein-coding genes of *T. thermophila* equals 47.78% (Eisen et al., 2006). This means that we observed significantly fewer allele frequency changes in protein-coding genes than expected by chance alone (exact binomial test: $p < 0.001$). The abundance of non-coding alleles with significant frequency changes is consistent with the possibility that gene regulatory changes are important for evolutionary adaptation in our populations. Alternatively, non-coding variants may frequently hitch-hike to high frequency with protein-coding variants. More than half (57.1%) of all alleles were indels as opposed to single nucleotide polymorphisms (42.9%). More indels and SNPs than expected by chance occurred in non-coding regions (exact binomial test; $p < 0.001$ for both). However, there were significantly more SNPs (34.61%) than indels (25.46%; Pearson Chi-squared test: $\chi^2_1 = 130.89$; $p < 0.001$) in protein-coding regions of the genome, consistent with the expectation that indels are likely to be more deleterious than SNPs.

We then investigated how the mode of reproduction, gene flow and abiotic conditions affected the number of alleles which changed in frequency.

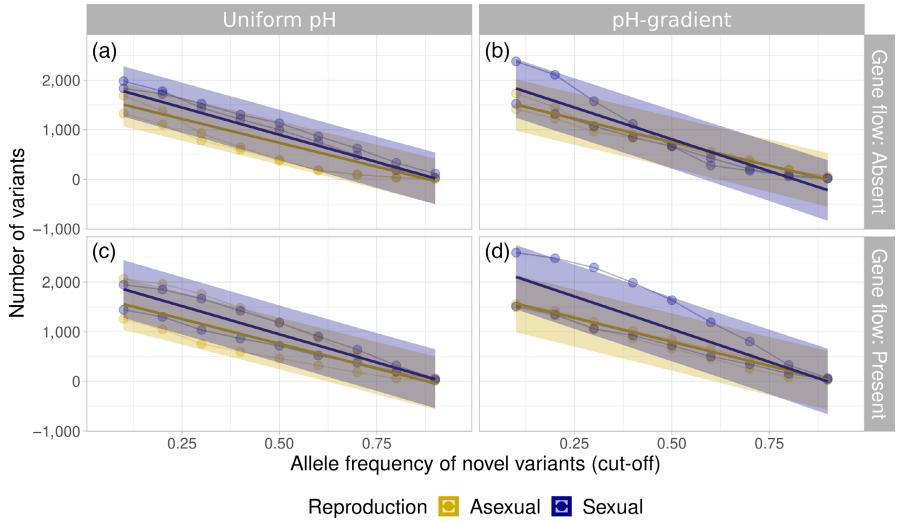

FIGURE 3 *De novo* variants increase to higher frequency in sexual populations. The figure shows the number of novel variants for the evolved populations expanding into a uniform environment (panels a and c) or into a pH gradient (panels b and d). The *x*-axis shows the cut-off value in minimum allele frequency. The *y*-axis shows the number of *de novo* variants whose frequency is above a given cut-off. Circles represent the data, with thin lines connecting data from the same replicate population. The thicker opaque lines and shaded areas show the predicted means and 95% confidence intervals for the best model. Colours represent the mode of reproduction

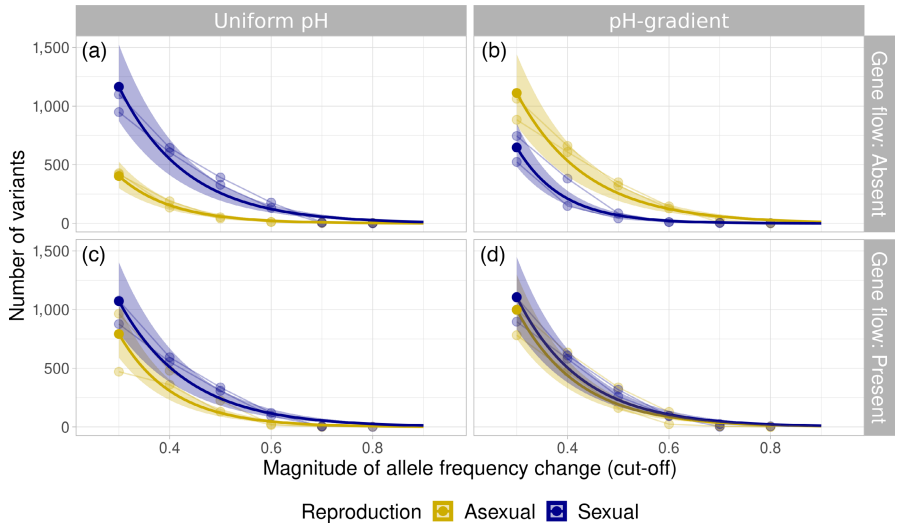

FIGURE 4 Reproduction alters genetic changes in standing genetic variation when gene flow is absent. Number of alleles that changed significantly in frequency during range expansion into a uniform environment (panels a and c) or into a pH gradient (panels b and d). The *x*-axis shows the cut-off value in the magnitude of allele frequency. The *y*-axis shows the number of variants analysed at each cut-off value. Circles represent allele frequency data, with thin lines connecting data from the same replicate population. Thick opaque lines show the model predictions for the best model. Shaded areas show 95%-confidence intervals from the posterior distribution. Colours represent mode of reproduction (yellow: asexual, blue:sexual)

The total number of alleles that changed significantly in frequency was similar for populations expanding into a uniform environment (Figure 4a,c) and into a pH gradient (Figure 4b,d). However, the effect of reproduction and gene flow differed between these populations. In the absence of gene flow, we observed more genetic changes for populations with a sexual mode of reproduction when populations expanded into a uniform environment (Figure 4a), but fewer genetic changes when populations expanded into a pH gradient (Figure 4b). When populations experienced gene flow during range expansion (Figure 4c,d), there was no clear difference between asexual and sexual populations. Full posterior distributions can be found in Supporting Information Section S12.1.

Next, to investigate whether the amount of genetic change from standing genetic variation can help explain the extent of fitness change, we compared the number of alleles that changed their frequency during evolution with the change in growth rate (fitness). To do so, we calculated the Pearson correlation coefficient between these two metrics (paired sample correlation test). We calculated

this Pearson correlation for each cut-off value for the magnitude of allele frequency change that we had used in the previous analysis of genetic change from standing variation. We did so separately for populations expanding into uniform conditions and populations expanding into a pH gradient, as well as for sexual and asexual populations (Figure 5). Thus, we calculated the correlation coefficient for four groups containing four evolved populations, combining populations with and without gene flow in a single group.

For asexual populations (yellow circles and triangles), the number of alleles changing their frequency and fitness change are not correlated. In contrast, for sexual populations (blue circles and triangles), the number of alleles changing their frequency and the extent of fitness change are correlated, depending on the abiotic conditions. Specifically, for populations expanding into a uniform environment, populations in which more variants changed their frequency also increased their fitness to a greater extent (Figure 5, left panels). In contrast for populations expanding into a pH gradient, these metrics were not correlated (Figure 5, right panels).

To explain why, in the absence of gene flow, sexual reproduction resulted in more allele frequency changes for populations expanding into a uniform environment, and fewer such changes for populations expanding into a pH gradient, we considered the evolutionary history of our populations. Because the conditions a population experiences when expanding into a uniform environment are similar to those experienced by the ancestral population, selection on standing genetic variation may suffice for adaptation. This possibility is supported by the positive correlation between the number of genetic changes and the fitness increase for populations expanding into a uniform environment (Figure 5). Sexual recombination could help such populations adapt by bringing together adaptive alleles from different ancestor clones (Maynard-Smith, 1978). In contrast, in the pH gradient (novel environment), new mutations or few variants from standing variation with a strong beneficial fitness effect may have been more important. In asexual populations, the lack of sexual recombination implies that any *de novo* variant or any initially rare beneficial variant from standing variation cannot be separated from its genetic background. Thus, if a beneficial variant is under strong positive selection, neutral or deleterious alleles that co-occur with it may hitch-hike to high allele frequencies. The importance of hitchhiking is supported by the observation that in asexual populations, the number of allelic changes does not correlate with fitness (see also Figure S6), suggesting that most allelic changes do not affect fitness strongly. In both cases, sexual reproduction may have aided adaptation, either by bringing together adaptive variants in the uniform environment or by separating adaptive mutations from their genetic background in the pH gradient.

When gene flow was present, we observed no clear difference in the number of variants whose frequency changed between asexual and sexual populations (Figure 4c,d). The reason may be the detrimental effect of gene flow on adaptation (Lenormand, 2002), as also evidenced by the finding that sexual reproduction only increased fitness in the absence of gene flow (Figure 2). Although we found that more *de novo* variants reached detectable frequencies in populations that experienced sexual reproduction (Figure 3), we found no clear link with fitness change (see Supporting Information Figure S6). This suggests that most of the detected mutations did not affect fitness strongly. Our finding that sexual reproduction hindered adaptation in the presence of gene flow not only in the pH gradient (Figure 2d), but also in the uniform environment (Figure 2c) is somewhat surprising within the framework of gene swamping. This could be explained if gene swamping does not only affect adaptation to the local environment, but also affects the evolution of life-history traits (Moerman, Fronhofer, et al., 2020b). Because our experiment linked gene flow with sexual reproduction, sexual reproduction may have led to the swamping of expanding populations with maladapted alleles (García-Ramos & Kirkpatrick, 1997; Haldane & Ford, 1956; Kirkpatrick & Barton, 1997), thus counteracting the adaptive effect of selection prior to sexual reproduction. Specifically, when populations expanded their range in the presence of gene flow, the repeated bouts of sexual reproduction may have reverted the adaptive benefits of previous selection. Therefore, these populations may have undergone adaptation primarily via selection of asexual clones during periods of asexual reproduction between bouts of sexual reproduction.

3.2 | Genes involved in adaptive evolution

In our next analysis, we wanted to identify specific classes of genes that may be involved in adaptive evolution. We distinguished between genes involved in general adaptation to conditions experienced by all populations, such as the growth medium and the expanding range, and genes that are specifically involved in adaptation to our pH gradient. For both classes of genes, we here focus on alleles that changed significantly in frequency during evolution, and report an analogous analysis of *de novo* variants in Supporting Information Section 15.

3.2.1 | Changes in standing variation associated with general adaptation involve transmembrane proteins and kinase domains

Of all polymorphic genomic loci at which significant allele frequency changes occurred, 242 loci fit our general adaptation criterion. Of these 242 loci, 53 occurred in protein-coding regions, which fell into 43 different genes (Table S16). Of these 43 genes, 12 encode transmembrane proteins, four encode kinase domain proteins, three encode cyclic nucleotide-binding domain proteins, three encode zinc fingers and two encode cation channel proteins. All other genes either encode uncharacterized proteins (10 out of 43 genes) or fell into unique functional categories (see also Table S16).

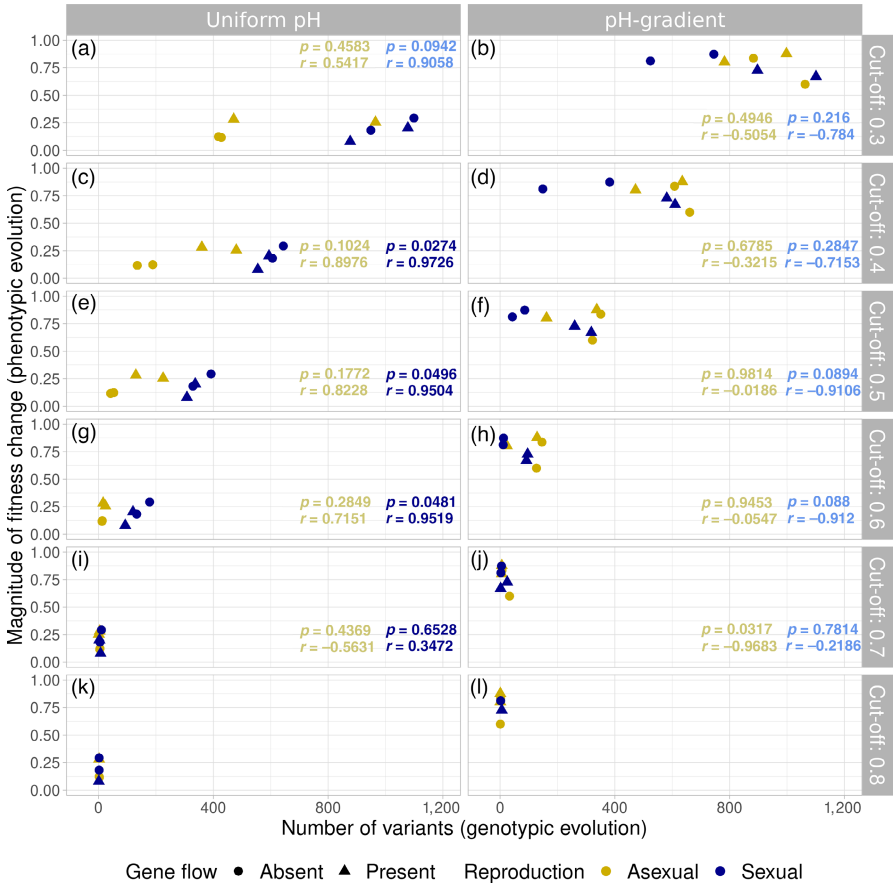

FIGURE 5 A positive correlation between fitness change and genetic change when sexual populations expand into a uniform environment. Statistical association between change in fitness and the number of allele frequency changes in standing genetic variation, for populations expanding into a uniform environment (left panels) and populations expanding into a pH gradient (right panels). The *x*-axis shows the number of standing genetic variants that changed significantly in allele frequency. The *y*-axis shows the change in growth rate (fitness) of evolved populations compared to the ancestral population. Horizontal subplots show the data for different cut-off values used for the minimum change in allele frequency at the end of evolution for inclusion in the analysis. Each geometric symbol within a subplot represents a single population, with colour representing reproductive mode (yellow = asexual, blue = sexual), and shape representing gene flow (circle = gene flow absent, triangle = gene flow present). Text insets show the Pearson correlation coefficient *r* and significance *p*, based on a paired sample correlation test for sexual populations (blue text) and asexual populations (yellow text). Non-significant correlations are shown in light colour (yellow or blue), whereas significant correlations are shown in dark colour. For the most stringent cut-off value (allele frequency change> 0.8), we did not calculate the correlation, because several populations did not harbour any variants that changed so dramatically in allele frequency

3.2.2 | Gradient-specific changes in allele frequency are associated with ion balance

We found 97,690 genomic positions for which we could quantify differences in allele frequencies through a Cochran–Mantel–Haenszel test (Methods). After Bonferroni correction, we retained 4,388 positions with significantly different allele frequencies between populations expanding into a pH gradient and populations expanding into a uniform environment. In all, 1,758 of these positions occurred in protein-coding regions, and fell into 636 different genes (Table S17 in Supporting Information). The largest subset of these genes with a functional annotation (136 genes) encode transmembrane proteins. Other highly represented groups include genes that encode kinase domain proteins (99 genes), cyclic-nucleotide-binding domain proteins (39 genes), cation channel proteins (18 genes) and zinc fingers (14 genes). The remaining genes either encode uncharacterized proteins (132 genes) or they fell into functional categories represented by only few genes. Notably, two genes, a gene encoding a cation channel family protein and a gene encoding an oxalate/formate antiporter, were involved in both gradient-specific changes in allele frequency and gradient-specific *de novo* variants (Tables S17 and S19 in Supporting Information).

3.2.3 | General adaptation preferentially affects genes involved in DNA repair and gene expression

To find out whether genes involved in general adaptation and in gradient-specific adaptation were associated with different functions, we studied the differential enrichment of gene ontology terms

in these two groups of genes, focusing on those gene ontology terms with clear differences between the groups.

Among 701 GO terms associated with the 43 general adaptation genes, 72 terms in 13 major categories were significantly enriched after multiple testing correction (Benjamini & Hochberg, 1995; Kofler et al., 2011). Likewise, among 1928 GO terms associated with 626 gradient-specific genes, 693 terms in 13 major categories were significantly enriched.

In Figure 6, the percentage of the 13 major categories is shown for the GO terms associated with general adaptation (panel a) and gradient-specific adaptation (panel b). In both datasets, we found many enriched GO terms related to metabolism, but the percentage of GO terms did not differ significantly between the two datasets. We found only two significantly differentially enriched GO-term categories. The first comprise transcription and translation functions, which were more likely to be associated with genes involved in general adaptation (8.97% of all GO-terms) than with genes involved in gradient-specific adaptation (1.3% of all GO terms; FDR< 0.001). The second comprises mitosis, DNA repair and chromosome division, which were also more likely to be associated with genes involved in general adaptation (16.67% of all GO terms) than in genes associated with genes gradient-specific adaptation (3.75% of all GO terms; FDR< 0.001). We thus found evidence that genes related to gene expression and genes related to cell division were under selection in all our populations, indicating that changes in such genes may be adaptive at a range edge. These observations are consistent with range expansion theory, which predicts strong selection on population growth rate at a range edge (Burton et al., 2010; Phillips et al., 2010; Shine et al., 2011).

Range expansion theory also predicts strong selection for increased dispersal (Burton et al., 2010; Shine et al., 2011). Although dispersal behaviour often has a genetic basis (Saastamoinen et al., 2018), we did not observe selection on genes obviously associated with dispersal in this experiment.

Genes involved in gradient-specific adaptation were associated more often with ion binding and transport, as well as with mitochondrial functioning and oxidoreductase reactions (see Figure 6). This suggests that genes coding for ion-related functions and mitochondrial-related functions may be adaptive depending on the pH environment. However, these differences were no longer significant after correcting for multiple testing. In general, genetic adaptations related to ion transport and binding may help counter osmotic stress under low pH conditions (Bremer & Krämer, 2019). Oxidoreductases play an important role in the production of reactive oxygen species (ROS), which are harmful to cells (Esterházy

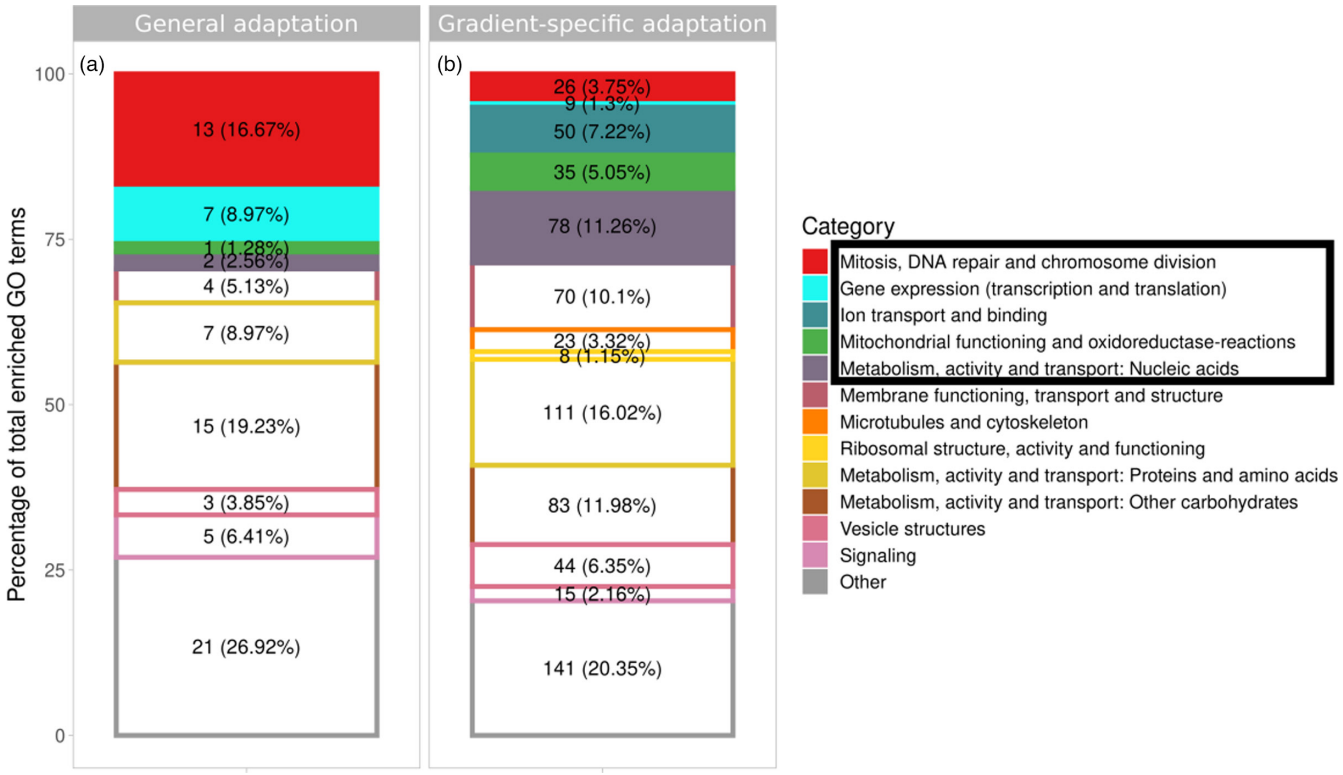

FIGURE 6 General and gradient-specific adaptations during range expansion. Partitioning of all enriched gene ontology terms (i.e. terms describing biological or molecular function of genes, as well as cellular structure) among 13 major functional categories for genes involved in general adaptations (panel a) and gradient-specific adaptations (panel b). Colours correspond to the 13 major GO categories we identified. Numbers in each rectangle represent the number and percentage of genes whose annotation fell in each category. Differentially enriched categories are displayed as filled rectangles and are boxed in the colour legend. The remaining categories are displayed as open rectangles with a coloured border

et al., 2008). Low pH can also contribute to elevated production of ROS (Lambert & Brand, 2004; Lemarie et al., 2011). A previous study found that selection for the capacity to metabolize ROS is an important aspect of evolution under pH stress in phytoplankton (Lindberg & Collins, 2020). Consequently, genes associated with mitochondrial functioning and oxidoreductase reactions may help cope with the harmful consequences of ROS.

3.3 | Limitations

While our experiments identified multiple genomic changes associated with range expansions, our ability to generalize from them is subject to several limitations. The first of them is the strong selection we observed against one ancestral clone. This phenomenon required us to apply stringent criteria to identify genes under selection, and may thus have masked some signatures of general adaptation. Specifically, it limited our ability to detect polygenic adaptation of multiple weakly selected variants, and complementary but non-parallel adaptation of different genes in different populations (Barghi et al., 2020; Stephan, 2016). Future range expansion experiments should be designed with the detection of these adaptation mechanisms in mind.

A second limitation is that we could not quantify the fitness effect of individual genetic changes. The reason is that we observed many genetic changes, and that it is difficult to genetically engineer *T. thermophila* to study the effect of any one such change. A different study organism may be needed to overcome this limitation.

Third, due to limitations caused by sequencing costs, we could only sequence pooled samples from a limited number of populations. One consequence is a loss of linkage information for specific variants, which limits our understanding of the role of sexual reproduction. Such information may be important for comparison with other species, as the large number of macronuclear chromosomes and random chromosome division in *Tetrahymena* may facilitate evolution in the absence of sexual reproduction, compared to species that lack such genetic characteristics. Another consequence is limited replication of sequencing across populations, which reduces our ability to identify signatures of selection based on parallel evolution across populations. The sequencing of more populations or individuals rather than pooled populations remains an important task for future work.

4 | CONCLUSIONS

We showed in this experiment that sexual reproduction can alter both phenotypic evolution and the underlying genetic architecture during range expansions, depending on both the presence of a pH gradient and the presence of long-distance gene flow. Our findings suggest that sexual reproduction may bring together adaptive variants from standing variation when populations expand in a uniform landscape. In addition, sexual reproduction may help separate adaptive from deleterious mutations and unlock the hidden benefits of rare genetic variants when populations expand their range into the novel environment of a pH gradient. Our previous observation that clonal populations also undergo strong adaptation to a low pH environment in the absence of sexual reproduction (Moerman, Arquint, et al., 2020a) suggests mutations rather than standing variation may play a major role for evolutionary adaptation to the novel pH-gradient environment.

Genetic adaptation in our experiment mostly affected genes linked to growth rate and the ability to grow in the abiotic environment. Unlike previous experiments, we found no indication for the evolution of dispersal behaviour itself, neither on a genetic level nor on a phenotypic level, indicating that in this experiment, the role of natural selection was more important than the role of spatial sorting.

ACKNOWLEDGEMENTS

We thank Samuel Hürlemann, Silvana Käser and Sarah Bratschi for help with laboratory work, Vanessa Weber De Melo for aid on the nuclear separation protocol, Carla Bello and Hélène Boulain for help with the bioinformatics, as well as two anonymous reviewers for their helpful comments on a previous version of the manuscript. Funding is from the University of Zurich URPP Evolution in Action and the Swiss National Science Foundation, Grant No PP00P3_179089. This is publication ISEM-2021-227 of the Institut des Sciences de l'Evolution—Montpellier. We would also like to acknowledge support by Swiss National Science Foundation grant 31003A_172887 and European Research Council Advanced Grant No. 739874. Open Access Funding provided by Universitat Zurich.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

AUTHORS' CONTRIBUTIONS

F.M., E.A.F., A.W. and F.A. designed the experiment; F.M. performed the experimental work and statistical analyses; F.M., F.A., A.W. and E.A.F. interpreted the results; F.M. and A.W. wrote the first version of the manuscript and all authors commented on and approved of the final version.

DATA AVAILABILITY STATEMENT

Phenotypic data and output files from variant calling, Cochran–Mantel–Haenszel test and GO-enrichment analyses are available on Dryad: https://doi.org/10.5061/dryad.w6m905qqh (Moerman et al., 2021). Raw sequence files are available from the European Sequence Archive (https://www.ebi.ac.uk/ena/browser/view/PRJEB47652). All analysis scripts and video analysis parameters are available on Github: https://doi.org/10.5281/zenodo.5521023 (Moerman, 2021).

ORCID

Felix Moerman http://orcid.org/0000-0002-5164-0978
Emanuel A. Fronhofer http://orcid.org/0000-0002-2219-784X
Florian Altermatt http://orcid.org/0000-0002-4831-6958
Andreas Wagner https://orcid.org/0000-0003-4299-3840

REFERENCES

Altermatt, F., Fronhofer, E. A., Garnier, A., Giometto, A., Hammes, F., Klecka, J., Legrand, D., Mächler, E., Massie, T. M., Pennekamp, F., Plebani, M., Pontarp, M., Schtickzelle, N., Thuillier, V., & Petchey, O. L. (2015). Big answers from small worlds: A user's guide for protist microcosms as a model system in ecology and evolution. *Methods in Ecology and Evolution*, *6*(2), 218–231.

Bachmann, J. C., Jansen van Rensburg, A., Cortazar-Chinarro, M., Laurila, A., & Van Buskirk, J. (2020). Gene flow limits adaptation along steep environmental gradients. *The American Naturalist*, *195*(3), E67–E86.

Barghi, N., Hermisson, J., & Schlötterer, C. (2020). Polygenic adaptation: A unifying framework to understand positive selection. *Nature Reviews Genetics*, *21*(12), 769–781.

Bell, G. (1982). *The masterpiece of nature: The evolution and genetics of sexuality*. University of California Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Ournal of the Royal Statistical Society Series B: Statistical Methodology*, *57*(1), 289–300.

Beverton, R., & Holt, S. (1993). *On the dynamics of exploited fish populations*. Springer.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ*, *310*(6973), 170.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bors, E. K., Herrera, S., Morris, J. A., & Shank, T. M. (2019). Population genomics of rapidly invading lionfish in the Caribbean reveals signals of range expansion in the absence of spatial population structure. *Ecology and Evolution*, *9*(6), 3306–3320. https://doi.org/10.1002/ece3.4952

Bosshard, L., Peischl, S., Ackermann, M., & Excoffier, L. (2020). Dissection of the mutation accumulation process during bacterial range expansions. *BMC Genomics*, *21*(1), 253. https://doi.org/10.1186/s12864-020-6676-z

Bremer, E., & Krämer, R. (2019). Responses of microorganisms to osmotic stress. *Annual Review of Microbiology*, *73*(1), 313–334. https://doi.org/10.1146/annurev-micro-020518-115504

Brown, G. P., Shilton, C., Phillips, B. L., & Shine, R. (2007). Invasion, stress, and spinal arthritis in cane toads. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(45), 17698–17700.

Burton, O. J., Phillips, B. L., & Travis, J. M. J. (2010). Trade-offs and the evolution of life-histories during range expansion. *Ecology Letters*, *13*(10), 1210–1220.

Cairns, J., Moerman, F., Fronhofer, E., Altermatt, F., & Hiltunen, T. (2020). Evolution in interacting species alters predator life history traits, behavior and morphology in experimental microbial communities. *Proceedings of the Royal Society B: Biological Sciences*, *287*. https://doi.org/10.1098/rspb.2020.0652

Caldeira, K., & Wickett, M. E. (2003). Anthropogenic carbon and ocean pH. *Nature*, *425*(6956), 365. https://doi.org/10.1038/425365a

Cassidy-Hanley, D. M. (2012). *Tetrahymena* in the laboratory: Strain resources, methods for culture, maintenance, and storage. *Methods in Cell Biology*, *109*, 237–276.

Charlesworth, B., Harvey, P. H., & Barton, N. H. (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society B*, *355*(1403), 1553–1562.

Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, *333*(6045), 1024–1026. https://doi.org/10.1126/science.1206432

Clarke, G. S., Shine, R., & Phillips, B. L. (2019). May the (selective) force be with you: Spatial sorting and natural selection exert opposing forces on limb length in an invasive amphibian. *Journal of Evolutionary Biology*, *32*(9), 994–1001. https://doi.org/10.1111/jeb.13504

Colegrave, N. (2002). Sex releases the speed limit on evolution. *Nature*, *420*(6916), 664–666. https://doi.org/10.1038/nature01191

Collins, K. (Ed.). (2012). *Tetrahymena thermophila* (1st ed.). Academic Press.

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution*, *62*(8), 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

Danecek, P., Schiffels, S., & Durbin, R. (2016). *Multiallelic calling model in bcftools (-m)*.

Ebbert, D. (2019). *chisq.posthoc.test: Perform post hoc analysis based on residuals of Pearson's Chi-Squared Test for Count Data*.

Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Smith, R. K., … Orias, E. (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biology*, *4*(9), e286.

Esterházy, D., King, M. S., Yakovlev, G., & Hirst, J. (2008). Production of reactive oxygen species by Complex I (NADH: Ubiquinone Oxidoreductase) from *Escherichia coli* and comparison to the enzyme from mitochondria. *Biochemistry*, *47*(12), 3964–3971. https://doi.org/10.1021/bi702243b

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics*, *40*(1), 481–501.

Fisher, R. A. (2000). *The genetical theory of natural selection*. Oxford University Press.

Fronhofer, E. A., & Altermatt, F. (2015). Eco-evolutionary feedbacks during experimental range expansions. *Nature Communications*, *6*, 6844. https://doi.org/10.1038/ncomms7844

García-Ramos, G., & Kirkpatrick, M. (1997). Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, *51*(1), 21–28. https://doi.org/10.1111/j.1558-5646.1997.tb02384.x

Giometto, A., Altermatt, F., & Rinaldo, A. (2017). Demographic stochasticity and resource autocorrelation control biological invasions in heterogeneous landscapes. *Oikos*, *126*(11), 1554–1563. https://doi.org/10.1111/oik.04330

Giometto, A., Rinaldo, A., Carrara, F., & Altermatt, F. (2014). Emerging predictable features of replicated biological invasion fronts. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 297–301.

Grosholz, E. (2002). Ecological and evolutionary consequences of coastal invasions. *Trends in Ecology & Evolution*, *17*(1), 22–27.

Haldane, J. B. S., & Ford, E. B. (1956). The relation between density regulation and natural selection. *Proceedings of the Royal Society B: Biological Sciences*, *145*(920), 306–308.

Hansen, T. F. (2006). The evolution of genetic architecture. *Annual Review of Ecology Evolution and Systematics*, *37*(1), 123–157.

Hartl, D. L., & Clark, A. G. (2006). *Principles of population genetics* (4th ed.). Sinauer Associates is an imprint of Oxford University Press.

Johnson, T., & Barton, N. H. (2002). The effect of deleterious alleles on adaptation in asexual populations. *Genetics*, *162*(1), 395–411. https://doi.org/10.1093/genetics/162.1.395

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a species' range. *The American Naturalist*, *150*(1), 1–23.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, *23*(3), 482–490.

Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, *27*(24), 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Kofler, R., & Schlötterer, C. (2012). Gowinda: Unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*, *28*(15), 2084–2085. https://doi.org/10.1093/bioinformatics/bts315

Kofler, R., & Schlötterer, C. (2014). A Guide for the design of evolve and resequencing studies. *Molecular Biology and Evolution*, *31*(2), 474–483.

Kowarik, I. (2003). Human agency in biological invasions: Secondary releases foster naturalisation and population expansion of alien plant species. *Biological Invasions*, *5*(4), 293–312. https://doi.org/10.1023/B:BINV.0000005574.15074.66

Lachapelle, J., & Bell, G. (2012). Evolutionary rescue of sexual and asexual populations in a deteriorating environment. *Evolution*, *66*(11), 3508–3518. https://doi.org/10.1111/j.1558-5646.2012.01697.x

Lambert, A. J., & Brand, M. D. (2004). Superoxide production by NADH:Ubiquinone oxidoreductase (complex I) depends on the pH gradient across the mitochondrial inner membrane. *The Biochemical Journal*, *382*(2), 511–517.

Latchman, D. S. (1997). Transcription factors: An overview. *International Journal of Biochemistry & Cell Biology*, *29*(12), 1305–1312.

Lavergne, S., & Molofsky, J. (2007). Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(10), 3883–3888.

Lemarie, A., Huc, L., Pazarentzos, E., Mahul-Mellier, A.-L., & Grimm, S. (2011). Specific disintegration of complex II succinate:Ubiquinone oxidoreductase links pH changes to oxidative stress for apoptosis induction. *Cell Death and Differentiation*, *18*(2), 338–349. https://doi.org/10.1038/cdd.2010.93

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183–189.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lindberg, R. T., & Collins, S. (2020). Quality–quantity trade-offs drive functional trait evolution in a model microalgal 'climate change winner'. *Ecology Letters*, *23*(5), 780–790.

Loewe, L., & Hill, W. G. (2010). The population genetics of mutations: Good, bad and indifferent. *Philosophical Transactions of the Royal Society B*, *365*(1544), 1153–1167.

Luijckx, P., Ho, E. K. H., Gasim, M., Chen, S., Stanic, A., Yanchus, C., Kim, Y. S., & Agrawal, A. F. (2017). Higher rates of sex evolve during adaptation to more complex environments. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(3), 534–539.

Lynn, D. H., & Doerder, F. P. (2012). The life and times of *Tetrahymena*. *Methods in Cell Biology*, *109*, 9–27.

Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics*, *35*(1), 303–339. https://doi.org/10.1146/annurev.genet.35.102401.090633

Maynard-Smith, J. (1978). *The evolution of sex*. Cambridge University Press.

McDonald, M. J., Rice, D. P., & Desai, M. M. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature*, *531*(7593), 233–236. https://doi.org/10.1038/nature17143

McElreath, R. (2015). *Statistical rethinking: A Bayesian course with examples in R and Stan* (1st ed.). Chapman and Hall/CRC.

Moerman, F. (2021). Data from: Accompanying data analysis scripts for: Selection on growth rate and local adaptation drive genomic adaptation during experimental range expansions in the protist *Tetrahymena thermophila*. *Zenodo*, https://doi.org/10.5281/zenodo.5521023.

Moerman, F., Altermatt, F., Fronhofer, E., & Wagner, A. (2021). Data from: Selection on Growth Rate and Local Adaptation Drive Genomic Adaptation during Experimental Range Expansions in the protist *Tetrahymena thermophila*. *Dryad*, https://doi.org/10.5061/dryad.w6m905qqh

Moerman, F., Arquint, A., Merkli, S., Wagner, A., Altermatt, F., & Fronhofer, E. A. (2020). Evolution under pH stress and high population densities leads to increased density-dependent fitness in the protest *Tetrahymena thermophila*. *Evolution*, *74*(3), 573–586.

Moerman, F., Fronhofer, E. A., Wagner, A., & Altermatt, F. (2020). Gene swamping alters evolution during range expansions in the protist *Tetrahymena thermophila*. *Biology Letters*, *16*(6), 20200244.

Muller, H. J. (1932). Some genetic aspects of sex. *The American Naturalist*, *66*(703), 118–138.

Ochocki, B. M., & Miller, T. E. X. (2017). Rapid evolution of dispersal ability makes biological invasions faster and more variable. *Nature Communications*, *8*, 14315. https://doi.org/10.1038/ncomms14315

Otto, S. P., & Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nature Reviews Genetics*, *3*(4), 252–261.

Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J. K., Thomas, C. D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tammaru, T., Tennent, W. J., Thomas, J. A., & Warren, M. (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*, *399*(6736), 579–583. https://doi.org/10.1038/21181

Peck, J. R. (1994). A ruby in the rubbish: Beneficial mutations, deleterious mutations and the evolution of sex. *Genetics*, *137*(2), 597–606. https://doi.org/10.1093/genetics/137.2.597

Petkovic, N., & Colegrave, N. (2019). Sex increases the probability of evolutionary rescue in the presence of a competitor. *Journal of Evolutionary Biology*, *32*(11), 1252–1261.

Phillips, B. (2009). The evolution of growth rates on an expanding range edge. *Biology Letters*, *5*(6), 802–804.

Phillips, B. L., Brown, G. P., & Shine, R. (2010). Life-history evolution in range-shifting populations. *Ecology*, *91*(6), 1617–1627. https://doi.org/10.1890/09-0910.1

Phillips, B. L., & Perkins, T. A. (2019). Spatial sorting as the spatial analogue of natural selection. *Theoretical Ecology*, *12*(2), 155–163.

Phillips, I. D., Vinebrooke, R. D., & Turner, M. A. (2009). Ecosystem consequences of potential range expansions of *Orconectes virilis* and *Orconectes rusticus* crayfish in Canada – A review. *Environmental Reviews*, *17*, 235–248.

Polechová, J. (2018). Is the sky the limit? On the expansion threshold of a species' range. *PLoS Biology*, *16*(6), e2005372. https://doi.org/10.1371/journal.pbio.2005372

Polechová, J., & Barton, N. H. (2015). Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), 6401–6406.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raven, J., Caldeira, K., Elderfield, H., Hoegh-Guldberg, O., Liss, P. S., Riebesell, U., Sheperd, J., Turley, C., & Watson, A. (2005). *Ocean acidification due to increasing atmospheric carbon dioxide*. Royal Society Policy Document.

Ruehle, M. D., Orias, E., & Pearson, C. G. (2016). *Tetrahymena* as a unicellular model Eukaryote: Genetic and genomic tools. *Genetics*, *203*(2), 649–665.

Saastamoinen, M., Bocedi, G., Cote, J., Legrand, D., Guillaume, F., Wheat, C. W., Fronhofer, E. A., Garcia, C., Henry, R., Husby, A., Baguette, M., Bonte, D., Coulon, A., Kokko, H., Matthysen, E., Niitepõld, K., Nonaka, E., Stevens, V. M., Travis, J. M. J., … Delgado, M. D. M. (2018). Genetics of dispersal. *Biological Reviews*, *93*(1), 574–599.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., Baughman, S., Cabin, R. J., Cohen, J. E., Ellstrand, N. C., McCauley, D. E., O'Neil, P., Parker, I. M., Thompson, J. N., & Weller, S. G. (2001). The population biology of invasive species.

Annual Review of Ecology and Systematics, *32*, 305–332. https://doi.org/10.1146/annurev.ecolsys.32.081501.114037

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals – Mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763.

Schmeller, D. S., Seitz, A., Crivelli, A., & Veith, M. (2005). Crossing species' range borders: Interspecies gene exchange mediated by hybridogenesis. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1572), 1625–1631.

Shine, R., Brown, G. P., & Phillips, B. L. (2011). An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5708–5711.

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Stan Development Team. (2020). *RStan: The R interface to Stan*.

Stephan, W. (2016). Signatures of positive selection: From selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, *25*(1), 79–88. https://doi.org/10.1111/mec.13288

Swaegers, J., Mergeay, J., Therry, L., Larmuseau, M. H. D., Bonte, D., & Stoks, R. (2013). Rapid range expansion increases genetic differentiation while causing limited reduction in genetic diversity in a damselfly. *Heredity*, *111*(5), 422–429. https://doi.org/10.1038/hdy.2013.64

Sweet, M. T., & Allis, C. D. (2006). Isolation and purification of *Tetrahymena* nuclei. *Cold Spring Harbor Protocols*, *2006*(4), pdb.prot4500. https://doi.org/10.1101/pdb.prot4500

Szücs, M., Vahsen, M. L., Melbourne, B. A., Hoover, C., Weiss-Lehman, C., & Hufbauer, R. A. (2017). Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(51), 13501–13506.

Van Petegem, K. H. P., Boeye, J., Stoks, R., & Bonte, D. (2016). Spatial selection and local adaptation jointly shape life-history evolution during range expansion. *The American Naturalist*, *188*(5), 485–498.

Walsh, B., & Blows, M. W. (2009). Abundant genetic variation + strong selection = Multivariate genetic constraints: A geometric view of adaptation. *Annual Review of Ecology Evolution and Systematics*, *40*(1), 41–59.

Williams, J. L., Kendall, B. E., & Levine, J. M. (2016). Rapid evolution accelerates plant population spread in fragmented experimental landscapes. *Science*, *353*(6298), 482–485. https://doi.org/10.1126/science.aaf6268

Zeebe, R. E., Zachos, J. C., Caldeira, K., & Tyrrell, T. (2008). Carbon emissions and acidification. *Science*, *321*(5885), 51–52.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Moerman, F., Fronhofer, E. A., Altermatt, F., & Wagner, A. (2022). Selection on growth rate and local adaptation drive genomic adaptation during experimental range expansions in the protist *Tetrahymena thermophila*. *Journal of Animal Ecology*, 91, 1088–1103. https://doi.org/10.1111/1365-2656.13598