Title: How phenotypic convergence arises in experimental evolution

Running Title: Experimental Evolution of Convergence

Authors: Pedro Simões[1#*] & Inês Fragata[1#*], Josiane Santos[1], Marta A. Santos[1], Mauro Santos[3], Michael R. Rose[2] & Margarida Matos[1]

Affiliations: [1] cE3c – Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal, Tel: +351 21 75 000 00**;** [2] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA.; [3] Departament de Genètica i de Microbiologia, Grup de Genòmica, Bioinformàtica i Biologia Evolutiva (GGBE), Universitat Autonòma de Barcelona, Spain.

[*]Corresponding authors: PS: pmsimoes@fc.ul.pt, IF: irfragata@gmail.com

# both authors contributed equally to this work.

Author Contributions: P.S. and I.F. participated in the conception and design of the study, collected and analysed the data and drafted the manuscript. J.S. participated in the conception and design of the study, data collection and revised critically the manuscript. M.A.S. participated in data collection and revised critically the manuscript. M.S. and M.R.R. participated in the conception and design of the study, as well as substantively revising the manuscript. M.M. conceived and designed the study, coordinated the study and helped drafting the manuscript. All authors gave final approval for publication.

Acknowledgments: The authors thank Laurence D. Mueller for the first version of the script that was later adapted for the analyses of the present study. Two anonymous reviewers provided helpful comments on an earlier draft. This study was partially financed by "Fundação para a Ciência e a Tecnologia" (FCT) projects POCTI/BSE/33673/2000, POCI-PPCDT/BIA-BDE/55853/2004 and PTDC/BIA-BDE/65733/2006 and by the cE3c FCT Unit UID/BIA/00329/2013. PS and IF have postdoctoral FCT grants (SFRH/BPD/86186/2012 and within the project JPIAMR/0001/2016 respectively). MS is funded by grants CGL2017-89160P from Ministerio de Economía y Competitividad (Spain), and 2017SGR 1379 from Generalitat de Catalunya.

Data Accessibility Statement: Raw data will be available at the Dryad Digital Repository upon acceptance.

Keywords: Convergent Evolution; Experimental Evolution; *Drosophila subobscura*; Life-History traits; Long-term Evolution; Adaptation.

Abstract

Evolutionary convergence is a core issue in the study of adaptive evolution, as well as a highly debated topic at present. Few studies have analyzed this issue using a "real-time" or evolutionary trajectory approach. Do populations that are initially differentiated converge to a similar adaptive state when experiencing a common novel environment? *Drosophila subobscura* populations founded from different locations and years showed initial differences and variation in evolutionary rates in several traits during short-term (~20 generations) laboratory adaptation. Here we extend that analysis to 40 more generations to analyze (1) how differences in evolutionary dynamics between populations change between shorter and longer time spans, and (2) whether evolutionary convergence occurs after sixty generations of evolution in a common environment. We found substantial variation in longer-term evolutionary trajectories and differences between short and longer-term evolutionary dynamics. Though we observed pervasive patterns of convergence towards the character values of long-established populations, populations still remain differentiated for several traits at the final generations analyzed. This pattern might involve transient divergence, as we report in some cases, indicating that more generations should lead to final convergence. These findings highlight the importance of longer-term studies for understanding convergent evolution.

Introduction

Understanding how populations adapt to environmental challenges is becoming increasingly important in both evolutionary biology and conservation (Botero et al. 2015; Franks and Hoffmann 2012). However, we are still unsure how predictable adaptation to novel environments is (Lachapelle et al. 2015; Lässig et al. 2017; Lenski et al. 2015; Orgogozo 2015; Wiser et al. 2013). Unpredictability in evolution can be caused by different genetic backgrounds due to prior evolutionary history (see Barton and Keightley 2002; Barrett and Schluter 2008; Hansen 2013), and stochastic events such as founder events, genetic drift, bottlenecks, etc. (see Lenormand et al. 2009). Furthermore, interactions between selection and genetic drift may also increase variation in evolutionary responses (e.g. Cohan 1984; Cohan and Hoffmann 1986; Santos et al. 2012).

An important question when different populations adapt to new environmental challenges is whether they will diverge or converge through time. Convergent evolution is expected to arise through the action of natural selection, erasing differences between populations (Endler 1986; Losos 2011; Stern 2013). Alternatively, differentiated populations could conceivably evolve increased differentiation when placed under similar selective regimes (Wright 1931; Cohan 1984; Whitlock et al. 1995). Discovering the constraints that produce either evolutionary convergence or evolutionary divergence is fundamental to ultimately understanding the foundations of adaptive evolution.

Experimental evolution is a powerful tool with which to address this problem, especially by studying the real-time evolutionary trajectories of different populations subjected to the same selective challenge. Several studies have observed convergent evolutionary responses in a new common environment (e.g. Travisano et al. 1995;

Teotónio and Rose 2000; 2002; Joshi et al. 2003; Simões et al. 2007, 2008; Teotónio et al. 2009; Santos et al. 2012; Fragata et al. 2014; Burke et al. 2016; Rebolleda-Gómez and Travisano 2019). Nevertheless, divergent evolutionary responses have also been observed (e.g. Cohan 1984; Cohan and Hoffmann 1986; Melnyk and Kassen 2011). Furthermore, several studies support the notion that the impact of evolutionary contingencies varies between traits closely or loosely related to fitness (Travisano et al. 1995; Teotónio et al. 2002; Joshi et al. 2003; Simões et al. 2008, 2017). It is thus clear from experimental evidence that evolutionary contingencies have a role in shaping evolutionary responses.

An important question, seldom addressed in the literature (but see Burke et al. 2016), is the effect of initial differentiation between populations on their long-term evolution. In particular, it is expected that different initial genetic backgrounds will have a higher impact during short-term evolution in a constant environment (Joshi et al. 2003; Fragata et al. 2014; Burke et al. 2016). On the other hand, at longer evolutionary scales, the cumulative effects of genetic drift and other stochastic events acting on the evolving populations will likely have a higher impact on the evolutionary trajectories observed (e.g. see Brito et al. 2005; Lenormand et al. 2009). Furthermore, different levels of standing genetic variation and/or epistatic interactions can have an important impact on long-term evolution (Barrett and Schluter 2008; Goodnight 2015; Paixão and Barton 2016; see empirical examples in Barton and Keightley 2002; Hansen 2013; Wiser et al. 2013; Good and Desai 2015). This might produce differences between populations, even in populations subject to similar selective pressures, possibly through different timings in the deceleration of the evolutionary response over time, for example (Teotónio and Rose 2000; Gilligan and Frankham 2003; Simões et al. 2007; Khan et al. 2011; Schoustra et al. 2012).

Long-term evolutionary dynamics have been mostly studied in microbial experimental evolution systems rather than in sexual organisms, due to the shorter generation time of the former. In the *E. coli* long-term evolution experiment performed in Lenski's lab, recent evidence indicates a deceleration of the evolutionary rate over 50000 generations (Wiser et al. 2013; Lenski et al. 2015). Furthermore, and perhaps surprisingly, heterogeneity in evolutionary trajectories is still present after so many generations, in part due to differences in mutation rates (Lenski et al. 2015). Several studies with sexual organisms, though involving fewer generations, have also observed the slowing down of evolutionary responses to newly imposed selection regimes (e.g. Gilligan and Frankham 2003; Rose et al. 2004; Simões et al. 2007, see below). The expectation of a deceleration of laboratory evolutionary trajectories in sexual organisms is sometimes justified in terms of temporal exhaustion of additive genetic variance, although genomic scans in experimentally evolved *Drosophila* populations have found only limited evidence of fixed alleles following selection (Burke et al. 2010; Burke and Long 2012; Orozco-Terwengel et al. 2012; Long et al. 2015; Phillips et al. 2016; Seabra et al. 2018). In a previous study by our team, we found evidence for a deceleration in the evolutionary trajectory of fecundity in populations of *Drosophila subobscura* evolving for more than 80 generations in the lab environment (Simões et al. 2007). Teotónio and Rose (2000) also found this pattern of response in several *D. melanogaster* lines undergoing reverse selection in their ancestral environment. Gilligan and Frankham (2003) also reported a slowing down of the rate of adaptation to captivity after 87 generations in the lab by comparing *Drosophila* populations in different stages of adaptation. However, this pattern is not universal, as other experimental studies have not found such deceleration of the evolutionary response, even after a higher number of generations. One example of this is the work of Chippindale et al. (1997), who imposed

123 selection for accelerated development time in *D. melanogaster*. Nevertheless, studies of
124 long-term experimental evolution in sexual species are scarce and have not specifically
125 addressed the variation in evolutionary dynamics that might occur during evolution over
126 the short term, relative to longer evolutionary time periods.

127 We have previously shown variation in the evolutionary response of several
128 populations of *D. subobscura* during the first 20 generations of evolution in a new
129 environment, the lab (Simões et al. 2008). These populations were founded from
130 different nearby locations over several years. We observed higher variation in the
131 evolutionary response for female starvation resistance, a trait likely more loosely related
132 to fitness in our experimental setting. By contrast, patterns for fecundity traits, which
133 are expected to be closer to fitness, were more repeatable. Importantly, the different
134 starvation resistance patterns led in fact to convergence between populations. In this
135 study we extend the earlier analysis to cover around forty additional generations. We
136 address the following questions: (1) How much do evolutionary rates vary between
137 short-term and longer-term evolution? (2) Do differences in evolutionary dynamics
138 between populations change in the transition from earlier to later generations? (3) Is
139 convergence observed after sixty generations of evolution in a common environment?

140
141 We expect that, during short-term evolution, variation in the initial genetic
142 backgrounds will lead to disparate rates of adaptation to the new environment. Over the
143 longer term, as the evolutionary response decelerates, differences between populations
144 of contrasting initial genetic composition are likely to be reduced relative to those
145 observed during short-term evolution, particularly if populations are evolving towards
146 the same phenotypic optimum.

147
148
149

Materials and Methods

Founding and Maintenance of the Laboratory populations

Five sets of wild-caught samples of *Drosophila subobscura* were analyzed in this study. These populations were founded in 1998 (NW populations; see Matos et al. 2002), 2001 (AR and TW populations; see (Simões et al. 2007), and 2005 (FWA and NARA; see (Simões et al. 2008). NW, TW and FWA populations were collected from a pinewood near Sintra (Portugal), whereas AR and NARA populations were collected from a pinewood in Arrábida (also from Portugal, some 50 Km from Sintra, on the other margin of the Tagus river; see Simões et al. 2007, 2008). All populations were three-fold replicated two generations after founding (e.g., FWA_{1-3} designating the three populations of FWA). A set of long-established laboratory populations (called "NB", founded in 1990 from Sintra) was used as a control for all the experimental populations. NB populations were at their 90^{th}, 136^{th} and 181^{st} laboratory generations at the time of foundation of the 1998, 2001 and 2005 collections, respectively.

All populations were maintained under the same laboratory environment with discrete generations of 28 days, reproduction close to peak fecundity, controlled temperature of 18ºC, with a 12-h L: 12-h D photoperiod. Flies were kept in vials, with controlled densities for both adult (around 50 individuals per vial) and larval stages (around 80 per vial). At each generation, emergences from the several vials of each replicate population were randomized using CO_2 anesthesia. Census population sizes ranged between 600 and 1200 adults. To study the evolutionary trajectories during laboratory adaptation, all experimental populations and the controls were periodically assayed for several phenotypic traits (see below).

177 *Phenotypic Assays and Generations analyzed*

178 For the phenotypic assays, mated pairs of flies were transferred daily to fresh
179 medium and the number of eggs laid per female was counted during the first 12 days
180 since emergence. After the fecundity assay, each pair of flies was transferred to a vial
181 containing plain agar medium to measure starvation resistance (with deaths checked
182 every 6 h). Five characters were analyzed: age of first reproduction (number of days
183 between emergence and the day of first egg laying), early fecundity (total number of
184 eggs laid during the first week), peak fecundity (total number of eggs laid between days
185 8 and 12), and female and male starvation resistance. Sample sizes ranged between 14
186 and 24 pairs per replicate population and assay. All assays involved synchronous
187 analyses with NB populations.

188 Periodical phenotypic assays were performed starting at generation 3 or 4 up to
189 generation 58-60. All generations assayed for the several populations are presented in
190 Table S1. We analyze here both short-term - ~20 generations - and a longer-term period
191 - between ~20 and ~60 generations, here designated “long-term” - of laboratory
192 evolution of these populations. We also analyzed the entire evolutionary trajectory,
193 spanning the complete data set. The short-term data was studied in Simões et al. (2008)
194 for a larger number of populations, the five sets of populations referred to above and an
195 extra set of populations in each of the 2005 locations (details in Simões et al. 2008).
196 Moreover, for NW there were five replicate populations with data on short term, but
197 here we only analyze three replicate populations, for both short and long-term, as only
198 these have data for more advanced generations. Finally, we expand our analyses to
199 include male starvation resistance data, which was not analyzed in Simões et al. (2008).

200 In order to calculate the initial or final state for each replicate population, we
201 calculated the mean value of the 2 (or 3) first (or last) generations by choosing 15

202 random individual data points (with replacement) of each generation involved. The
203 initial generations used were the following: 4, 6 and 7 for AR and TW; 4 and 8 for NW;
204 3, 6 and 10 for NARA and FWA. The final generations analyzed were: 48, 55 and 60
205 for AR and TW; 52, 53 and 58 for NW; 49 and 58 for NARA and FWA.
206
207 *Statistical Methods*
208 To estimate the evolutionary trajectories for each population, in each assayed
209 generation, we used the differences between individual data and the mean of the same-
210 numbered NB replicate population (assayed synchronously with experimental
211 populations; e.g. AR1-average NB1), see (Simões et al. 2008). This was done to remove
212 the effect of possible temporal changes not related to laboratory adaptation such as
213 trends due to environmental variation or to inadvertent evolutionary changes not
214 intended in the study (e.g. due to slight changes of conditions in lab). This procedure
215 also minimizes the effects of environmental heterogeneity between non-synchronous
216 assays (see also Matos et al. 2002; Simões et al. 2007, 2008). Temporal performance of
217 the control populations was generally quite stable across traits, allowing us to rule out
218 undesirable sources of variation such as those due to further laboratory adaptation or
219 inbreeding (see Fig S1).
220 Linear and linear-log models were tested for both periods separately and over the
221 whole evolutionary trajectory of the populations (around 60 generations). Models were
222 chosen according to their fit to the data based on R^2 values (see Table S2). For the
223 separate analyses of short-term and long-term periods, we chose the linear over the
224 linear-log model as a compromise across populations and periods, since the same model
225 had to be applied to allow for direct comparisons between periods (e.g. for the tests in

Table 1 and 2). For the analysis of overall trajectories, the linear-log model was chosen, as it generally presented a better fit than the linear model (see Table S2).

Bootstrap Techniques

Variation in the slope of evolutionary response between sets of populations and periods was studied using bootstrap techniques as in Simões et al. (2008). Briefly, for each replicate population we estimated the intercept ($\hat{\beta}_0$), evolutionary slope ($\hat{\beta}_1$) and the residuals of each point (ε) using a simple linear regression. In each iteration of the bootstrap, a new vector of phenotypic data was created by resampling the residuals, with replacement (ε^*) and employing the following formula to calculate a new phenotypic value for each data point used:

1) $$y^* = \hat{\beta}_0 + \hat{\beta}_1 x \text{ Generation} + \varepsilon^*$$

After this, a new slope (β_1^*) and intercept (β_0^*) were estimated through a linear regression. For the linear-log model the same analysis was applied using the natural logarithm of the generation. A total of 10000 slopes were generated for each replicate population. All analyses testing differences between slopes were done using these values.

To compare two sets of populations from the same location in different years, we calculated the mean of each set involved in the comparison (by randomly sampling one slope from each replicate population) and the difference between them (e.g. comparison Arrábida 2001 vs. Arrábida 2005: $((AR_1\beta_1^* + AR_2\beta_1^* + AR_3\beta_1^*)/3) - ((NARA_1\beta_1^* + NARA_2\beta_1^* + NARA_3\beta_1^*)/3)$. This process was repeated 10000 times. Statistical significance was assessed by estimating the fraction of these 10000 differences that were greater than zero. Two times this fraction or 1 minus two times this fraction (whichever is less) corresponds to the *P*-value. To compare differences

between the two locations we used all 2001 and 2005 replicate populations from each location. NW data was not included as there were no corresponding populations founded in Arrábida in 1998. We calculated the location means using data of six replicate populations (e.g. $FWA_{1\text{-}3}$ and $TW_{1\text{-}3}$ for the Sintra slopes), again using random samples of slopes from each replicate population, as above. Differences between the short and long-term evolutionary response for each set of populations were also assessed (e.g. comparison of $TW_{1\text{-}3}$ short-term slopes vs $TW_{1\text{-}3}$ long-term slopes). We further analyzed whether differences between periods varied between populations founded from distinct years or locations. These comparisons followed the same rationale as above (e.g. comparison short vs long-term for Arrábida 2001 vs Arrábida 2005: $((AR_1\beta_1{*}_{\mathbf{S}} + AR_2\beta_1{*}_{\mathbf{S}} + AR_3\beta_1{*}_{\mathbf{S}})/3 - (NARA_1\beta_1{*}_{\mathbf{S}} + NARA_2\beta_1{*}_{\mathbf{S}} + NARA_2\beta_1{*}_{\mathbf{S}})/3)$- $((AR_1\beta_1{*}_{\mathbf{L}} + AR_2\beta_1{*}_{\mathbf{L}} + AR_3\beta_1{*}_{\mathbf{L}})/3 - (NARA_1\beta_1{*}_{\mathbf{L}} + NARA_2\beta_1{*}_{\mathbf{L}} + NARA_2\beta_1{*}_{\mathbf{L}})/3)$. This analysis was performed with 10000 random samples and tested as described above.

To test whether populations differed in the initial or final performance, 10000 comparisons between years and locations were assessed using the same rationale as above.

When testing for differences between populations statistical significance is presented both with and without False Discovery Rate (FDR) correction for five tests (theorem 1.3 Benjamini and Yekutieli 2001). Marginally significant results after FDR correction will also be considered when the general reading justifies, i.e. if there are consistent patterns across populations. This is a compromise, as being too conservative also has drawbacks, given that the focus of this study is not on single tests but rather to analyze patterns across comparisons.

All analyses were performed using R version 3.3.1 (R Core Team 2018), package reshape2 (Wickham 2007) and visualization was done using ggplot2 package (Wickham 2009).

Results

Initial differences between populations

The experimental populations were clearly differentiated from the control populations in the initial performance of fecundity traits, though less so for starvation resistance (Fig 1 and 2). NW populations performed significantly better than the other Sintra populations, both in age of first reproduction and early fecundity, whereas they performed worse for male starvation resistance (see Table S3 and Figs 1 and 2). Most populations from different years showed significant differences in the initial performance for peak fecundity and female starvation resistance. On the other hand, no significant differences were found between locations for any trait (see Table S3).

Short-term Evolutionary Dynamics

In general, fecundity-related traits, particularly early fecundity, show a clear evolutionary increase in performance during short-term evolution across populations, with a tendency to converge to control values, although at different rates (see Fig 1 and below; see also Fig S2, for data on the mean and variation of slopes of replicate populations). In contrast, patterns for starvation resistance are less consistent. In fact, male starvation resistance does not show a noticeable evolutionary response, although there is a suggestion of increased starvation across generations for all sets of populations except AR (see Fig 2 and Fig S2). The evolutionary response of female starvation resistance varies greatly among sets of populations with patterns of stasis

(AR and TW), decreased (FWA and NARA) and increased performance (for NW). In spite of these differences, the patterns are again of convergence to control values (see Fig 2, S2 and below).

When comparing the evolutionary response among populations, we observe significant differences of slopes between years (see Table 1). This variation is particularly evident for female starvation resistance, in agreement with our previous analysis (Simões et al. 2008). On the other hand, for male starvation resistance no significant variation in the evolutionary response was found. Significant differences between locations were only observed for peak fecundity (see Table 1).

Interestingly, of the eight comparisons showing significant (or at least marginally significant, after FDR correction) differences between populations in short-term dynamics across all assayed traits (Table 1), six of these showed also significant variation in initial performance (cf. Table 1 and Table S3). This concordance corresponded to a reduction of differences between populations through time for age of first reproduction and female starvation resistance. In contrast, for early fecundity, a higher initial performance of NW relative to TW or FWA was followed by faster improvement through time increasing the initial differences, leading at least to transient divergence between populations (Table 1 and Table S3, Fig 1 and 2).

Long-term Evolutionary Dynamics

In each set of populations there was a clear variation of evolutionary rates (slopes) between the short-term and the long-term period for age of first reproduction, early fecundity and female starvation resistance (Table S4). This corresponded to a general slowing down of the evolutionary response as populations tended to converge to the

control values (see Figs 1, 2 and S2; Fig S3 shows the same pattern in the evolutionary trajectories using all generations).

Differences in evolutionary dynamics between sets of populations were more evident in the long-term than in the short-term evolutionary response for several traits, particularly for early fecundity (see Table 1; Figs 1, 2 and S2). For this trait, a significant effect of location was due to a higher evolutionary rate in Sintra populations. Also, several comparisons showed significant effects of year, due in part to a lower slowing down of the response of the 2001 populations. On the other hand, differences between populations in the evolutionary response of female starvation resistance decreased in this period with only two significant effects in five comparisons– see Table 1. These significant effects involved comparisons with NW, which showed a clear drop in performance during this later period (see Fig 2).

When comparing the variation in evolutionary rates of the different sets of populations between the two periods (short vs. long-term evolution), early fecundity and female starvation resistance showed the greatest differences between populations, due to the above mentioned differential slowing down of response for early fecundity during long-term evolution and to the reported high variation in evolutionary rates seen in the short term evolution of female starvation resistance (Figs 1 and 2, Table 2). Importantly, for early fecundity, populations with higher short-term evolutionary rates (NW and the two 2005 populations) were also those with a stronger slowing down in the long-term period (Fig 1), which is expected under convergent evolution (see below).

Final differences between populations

In more advanced generations, there was a loss of the initial differences between populations for several comparisons, as expected if full convergence occurs (see Table

S3). This was observed between NW and TW for all fecundity traits, between NW and FWA for age of first reproduction, and between the 2001 and 2005 populations for female starvation resistance (Table S3 and Figs. 1 and 2). Nevertheless, significant differences in final performance were also found for several comparisons (see Table S3). In some cases, differentiation was also present at the start. Three temporal patterns were observed taking into account initial, intermediate, and final values (see Table S5): 1- continuous reduction of differences (NW versus TW for male starvation resistance); 2- increased differences through time (TW versus FWA for peak fecundity); 3- differentiation at the initial and final generations but with intermediate loss of differentiation (NW versus FWA for early fecundity and female starvation resistance). Finally, in other comparisons there was a significant (or at least marginally significant after FDR correction) differentiation between populations at the later stage of adaptation, not present at the start (Table S3). In this case two temporal patterns were observed (see Fig 1 and Table 1, S3 and S5): 1- higher differences at the end than at intermediate or initial generations (Arrábida versus Sintra populations for early fecundity and NW versus FWA for peak fecundity); 2 - - higher differences at intermediate generations than at the end of the study due to a differential slowing down of the evolutionary rate (between the two sets of Arrábida populations for age of first reproduction and early fecundity; in both cases differences are marginally significant after FDR correction).

Overall Evolutionary Dynamics

Evolutionary trajectories across the entire time span confirm a general deceleration of the evolutionary response through time, as populations evolved towards the control values (see Fig S3). This led to a generally better fit of the overall evolutionary

trajectory to a linear-log model relative to a linear one, particularly for fecundity-related data (see Table S2). Differences between sets of populations in the overall evolutionary response were due to variable changes between short and long periods, leading to pervasive contrasts, particularly for early fecundity (see Table 1 and S6).

Discussion

Evolutionary convergence is a core expectation for adaptive evolution in a similar environment (Losos 2011; Stern 2013). With a smooth fitness landscape, that lacks multiple peaks, populations will tend to evolve to the same outcome (Wright 1931). In such cases, the outcome of evolution will be predictable. The predictability of evolution is an issue of much interest at present (e.g. de Visser and Krug 2014; Orgogozo 2015). Experimental evolution is a great tool for testing whether adaptive evolution involves smooth or rugged landscapes, as it allows us to study the fate of populations initially differentiated when subject to similar selective pressures, especially whether they evolve towards similar or different fitness values (Fragata et al. 2018; Matos et al. 2015; Orgogozo 2015; Rebolleda-Gómez and Travisano 2019). Here we add to the previous Simões et al. (2008) study the analysis of c. 40 more generations of laboratory adaptation, in order to determine whether: 1) longer-term evolution leads to similar outcomes as short-term evolution; 2) populations will ultimately tend to converge or show more complex evolutionary patterns.

In this study we found a general pattern of convergent evolution, with clear changes in the evolutionary rates between the short-term (~20 generations) and longer-term (~60 generations) periods. We observed a slowing down of the evolutionary response through time for several traits as populations approached the evolutionary equilibria of long-established populations. Empirical evidence for deceleration of evolutionary rate

399 has been observed in other experimental studies using both asexual (Wiser et al. 2013;
400 Lenski et al. 2015) and sexual organisms (Gilligan and Frankham 2003; Simões et al.
401 2007).

402 We also observed that the differences between short-term and longer-term dynamics
403 were trait and population specific. Whereas differences in the early-fecundity response
404 between sets of populations increased from short- to long-term evolution, the inverse
405 pattern was observed for female starvation resistance. The source of differences
406 between populations also varied between traits. In the case of early fecundity, trajectory
407 variation was due to a continuous increase in performance of the 2001 populations, even
408 during long-term evolution, contrasting with the 1998 and 2005 populations, where
409 quicker short-term evolution was followed by a slowing of the evolutionary response
410 after generation 20. These differences are consistent with convergent evolution, as faster
411 evolution in an earlier period is followed by a plateauing, while slower evolution
412 corresponds to a steadier evolutionary rate throughout generations. Such contrasting
413 evolutionary dynamics led to an interesting pattern: an intermediate phase of transient
414 divergence was followed in the long-term by a partial convergence among evolving
415 populations. In contrast, for female starvation resistance there were striking differences
416 in the evolutionary trajectories during short-term evolution, with increase, decrease, or
417 stasis contingent on the degree of initial differentiation from controls (see also Simões
418 et al. 2008). For this trait, convergence was fast between all populations. These patterns
419 were followed in general by a reduction of differences between evolutionary trajectories
420 over the longer time period analyzed. The exception was the NW populations, which
421 presented an initial positive trend, unique across populations (see also Matos et al.
422 2004), followed by a negative long-term trend. Nevertheless, despite the different

423 underlying evolutionary dynamics, both early fecundity and female starvation resistance
424 show a general pattern that suggests convergence in longer-term periods.

425 It is an inherent expectation of convergent evolution that there will be a negative
426 association between initial state and subsequent evolutionary rates of populations
427 adapting to a new environment (Simões et al. 2007). This expectation was confirmed
428 for *D. subobscura* populations with clear initial historical differentiation, founded from
429 contrasting latitudes of the European cline (Fragata et al. 2014). In that study fast
430 convergence was observed after only 14 generations in a common environment. In our
431 study, evidence of such an association was only found for age of first reproduction and
432 female starvation resistance for the short-term dynamics. Even so, for female starvation
433 resistance the overall trend was not of convergence in the case of NW populations (see
434 above). The relative lack of such overall and rapid convergence in our study might be
435 due to the smaller degree of initial differentiation of these populations, with greater
436 sampling effects (Santos et al. 2012).

437 If full convergence occurs, an obvious corollary is that populations will not be
438 differentiated as an outcome of evolution in a common environment. This expectation
439 was not entirely met in our study, as several populations remained differentiated for
440 some traits after sixty generations of evolution. In this context, several patterns emerged
441 when comparing dynamics between different populations: (1) continuous reduction of
442 differences indicating partial convergence (for male starvation resistance); (2)
443 continuous divergence between populations (for early and peak fecundity); (3) transient
444 divergence followed by partial convergence (for age of first reproduction and early
445 fecundity) or (4) transient convergence followed by later divergence (for early fecundity
446 and female starvation resistance). Teotónio and his collaborators (Teotónio et al. 2002;
447 Teotónio and Rose 2000) performed a reverse evolution study during 50 generations

448 involving many genetically differentiated *Drosophila melanogaster* populations. They
449 found that populations converged to ancestral values, but this trend was not general as it
450 varied with the previous history and the trait studied. They concluded that populations
451 converged to similar fitness values to a larger extent than other characters did. In
452 contrast, in our study we did not see any clear relation between the extent of
453 convergence and how the traits analyzed were presumed to determine fitness. In fact,
454 several populations remained differentiated for early fecundity, a trait that is under
455 strong selection in our environment with clear and consistent improvement across many
456 independent studies (Fragata et al. 2014; Matos et al. 2002; Matos et al. 2004; Simões et
457 al. 2007; Simões et al. 2008). Given our interpretation of transient divergence and
458 partial convergence in some of these populations, it is possible that the time span of the
459 study was not sufficient to allow for full convergence in some cases, convergence that
460 might ultimately occur over more generations of evolution.

461 We observed considerable differences between short-term and longer-term dynamics
462 in all our populations, which raises questions about predicting long-term evolution from
463 short-term evolution. This contrasts with the study of Burke et al. (2016), which
464 suggests that short-term evolution is predictive of longer evolutionary time periods. In
465 that study recently selected *D. melanogaster* populations converged to the trait values of
466 other independently derived populations evolving in a similar selection regime for a
467 longer time scale, regardless of the evolutionary history of the populations studied.
468 However, different time scales were involved, as the shorter-term evolutionary
469 responses of that study were sometimes more than 100 generations in duration, with
470 long-term evolution approaching 1,000 generations. In general, the fact that our study
471 showed such differentiated outcomes and complex evolutionary patterns highlights the

472 importance of characterizing extended periods of experimental evolution and the
473 possible pitfalls of predicting evolution from short-term adaptive patterns.
474
475 **Conclusions**
476 We here showed that after 60 generations of evolution in a common environment,
477 *Drosophila subobscura* populations remain differentiated for several traits. Noticeably,
478 this was observed even for life-history traits that are clearly under selection in our lab.
479 In this context, we found evidence for transient divergence, as a result of heterogeneity
480 in evolutionary rates through time, occurring under a general scenario of convergence.
481 Ultimately, we conclude that extrapolating from short-term evolutionary patterns to
482 longer evolutionary periods might be risky, particularly if one is interested in predicting
483 the outcomes of evolution.
484
485 ## Literature Cited

486 Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation.
487 Trends Ecol. Evol. 23:38–44.

488 Barton, N. H., and P. D. Keightley. 2002. Understanding quantitative genetic variation.
489 Nat. Rev. Genet. 3:11–21.

490 Benjamini, Y., and D. Yekutieli. 2001. The control of the false discovery rate in
491 multiple testing under dependency. Ann. Stat. 29:1165–1188.

492 Botero, C. A., F. J. Weissing, J. Wright, and D. R. Rubenstein. 2015. Evolutionary
493 tipping points in the capacity to adapt to environmental change. Proc. Natl. Acad.
494 Sci. USA 112:184–189.

Brito, R. A., L. S. Pletscher, and J. M. Cheverudi. 2005. the Evolution of Genetic Architecture. I. Diversification of Genetic Backgrounds By Genetic Drift. Evolution 59:2333–2342.

Burke, M., J. Dunham, P. Shahrestani, K. Thornton, M. Rose, and A. Long. 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature 467:587–590.

Burke, M. K., T. T. Barter, L. G. Cabral, J. N. Kezos, M. A. Phillips, G. A. Rutledge, K. H. Phung, R. H. Chen, H. D. Nguyen, L. D. Mueller, and M. R. Rose. 2016. Rapid divergence and convergence of life-history in experimentally evolved *Drosophila melanogaster*. Evolution 70:2085–2098.

Burke, M. K., and A. D. Long. 2012. What paths do advantageous alleles take during short-term evolutionary change? Mol. Ecol. 21:4913–4916.

Chippindale, A. K., J. A. Alipaz, H.-W. Chen, and M. R. Rose. 1997. Experimental Evolution of Accelerated Development in Drosophila. 1. Developmental Speed and Larval Survival. Evolution 51:1536–1551.

Cohan, F. M. 1984. Can Uniform Selection Retard Random Genetic Divergence Between Isolated Conspecific Populations? Evolution 38:495–504.

Cohan, F. M., and A. A. Hoffmann. 1986. Genetic divergence under uniform selection. II. Different responses to selection for knockdown resistance to ethanol among *Drosophila melanogaster* populations and their replicate lines. Genetics 114:145–164.

de Visser, J. A. G. M., and J. Krug. 2014. Empirical fitness landscapes and the predictability of evolution. Nat. Rev. Genet. 15:480–490.

518 Endler, J. A. 1986. Natural Selection in the Wild. Princeton University Press.

519 Fox, C. W., J. D. Wagner, S. Cline, F. A. Thomas, and F. J. Messina. 2011. Rapid
520 Evolution of Lifespan in a Novel Environment: Sex-Specific Responses and
521 Underlying Genetic Architecture. Evol. Biol. 38:182–196.

522 Fragata, I., P. Simões, M. Lopes-Cunha, M. Lima, B. Kellen, M. Bárbaro, J. Santos, M.
523 R. Rose, M. Santos, and M. Matos. 2014. Laboratory selection quickly erases
524 historical differentiation. PLoS One 9:e96227.

525 Fragata, I., P. Simões, M. Matos, E. Szathmáry, and M. Santos. 2018. Playing evolution
526 in the laboratory: From the first major evolutionary transition to global warming.
527 Epl 122:3.

528 Franks, S. J., and A. A. Hoffmann. 2012. Genetics of Climate Change Adaptation.
529 Annu. Rev. Genet 46:185–208.

530 Garland, T., and M. R. Rose (eds). 2009. Experimental Evolution: Concepts, Methods,
531 and Applications of Selection Experiments. University of California Press, Los
532 Angeles, USA.

533 Gilligan, D. M., and R. Frankham. 2003. Dynamics of genetic adaption to captivity.
534 Conserv. Genet. 4:189–197.

535 Good, B. H., and M. M. Desai. 2015. The impact of macroscopic epistasis on long-term
536 evolutionary dynamics. Genetics 199:177–190.

537 Goodnight, C. J. 2015. Long-term selection experiments: Epistasis and the response to
538 selection. Pp. 1–18 *in* J. H. Moore and S. M. Williams, eds. Epistasis: Methods and
539 protocols. Springer Science+Business Media New York.

Hansen, T. F. 2013. Why epistasis is important for selection and adaptation. Evolution 67:3501–3511.

Joshi, A., R. B. Castillo, and L. D. Mueller. 2003. The contribution of ancestry, chance, and past and ongoing selection to adaptive evolution. J. Genet. 82:147–162.

Kawecki, T. J., and F. Mery. 2006. Genetically idiosyncratic responses of *Drosophila melanogaster* populations to selection for improved learning ability. J. Evol. Biol. 19:1265–1274.

Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. Science 332:1193–1196.

Lachapelle, J., J. Reid, and N. Colegrave. 2015. Repeatability of adaptation in experimental populations of different sizes. Proc. R. Soc. B 282:20143033.

Lässig, M., V. Mustonen, and A. M. Walczak. 2017. Predicting evolution. Nat. Ecol. Evol. 1:0077.

Lenormand, T., D. Roze, and F. Rousset. 2009. Stochasticity in evolution. Trends Ecol. Evol. 24:157–165.

Lenski, R. E., M. J. Wiser, N. Ribeck, Z. D. Blount, R. Maddamsetti, A. R. Burmeister, E. J. Baird, and J. Bundy. 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. Proc. R. Soc. B 282:20152292.

Long, A., G. Liti, A. Luptak, and O. Tenaillon. 2015. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat. Rev. Genet.

16:567–582.

Losos, J. B. 2011. Convergence, adaptation, and constraint. Evolution 65:1827–1840.

Matos, M., T. Avelar, and M. R. Rose. 2002. Variation in the rate of convergent evolution: adaptation to a laboratory environment in *Drosophila subobscura*. J. Evol. Biol. 15:673–682.

Matos, M., P. Simões, A. Duarte, C. Rego, T. Avelar, and M. R. Rose. 2004. Convergence to a novel environment: comparative method versus experimental evolution. Evolution 58:1503–1510.

Matos, M., P. Simões, M. A. Santos, S. G. Seabra, G. S. Faria, and F. Vala. 2015. History, chance and selection during phenotypic and genomic experimental evolution: replaying the tape of life at different levels. Front. Genet. 6:71.

Melnyk, A. H., and R. Kassen. 2011. Adaptive landscapes in evolving populations of *Pseudomonas fluorescens*. Evolution 65:3048–3059.

Orgogozo, V. 2015. Replaying the tape of life in the twenty-first century. Interface Focus 5:20150057.

Orozco-Terwengel, P., M. Kapun, V. Nolte, R. Kofler, T. Flatt, and C. Schlãtterer. 2012. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Mol. Ecol. 21:4931–4941.

Paixão, T., and N. H. Barton. 2016. The effect of gene interactions on the long-term response to selection. Proc. Natl. Acad. Sci. 113:4422–4427.

Phillips, M. A., A. D. Long, Z. S. Greenspan, L. F. Greer, M. K. Burke, B.

584 Villeponteau, K. C. Matsagas, C. L. Rizza, L. D. Mueller, and M. R. Rose. 2016.
585 Genome-wide analysis of long-term evolutionary domestication in *Drosophila*
586 *melanogaster*. Sci. Rep. 6:39281.

587 R Core Team. 2018. R: A Language and Environment for Statistical Computing.

588 Rebolleda-Gómez, M., and M. Travisano. 2019. Adaptation, chance, and history in
589 experimental evolution reversals to unicellularity. Evolution 73:73–83.

590 Rose, M. R., H. B. Passananti, and M. Matos. 2004. Methuselah Flies: A Case Study in
591 the Evolution of Aging. World Scientific, Singapore.

592 Santos, J., M. Pascual, P. Simões, I. Fragata, M. Lima, B. Kellen, M. Santos, A.
593 Marques, M. R. Rose, and M. Matos. 2012. From nature to the laboratory: the
594 impact of founder effects on adaptation. J. Evol. Biol. 25:2607–2622.

595 Schoustra, S. E., D. Punzalan, R. Dali, H. D. Rundle, and R. Kassen. 2012. Multivariate
596 Phenotypic Divergence Due to the Fixation of Beneficial Mutations in
597 Experimentally Evolved Lineages of a Filamentous Fungus. PLoS One 7(11):
598 e50305.

599 Seabra, S. G., I. Fragata, M. A. Antunes, G. S. Faria, M. A. Santos, V. C. Sousa, P.
600 Simões, and M. Matos. 2018. Different genomic changes underlie adaptive
601 evolution in populations of contrasting history. Mol. Biol. Evol. 35:549–563.

602 Simões, P., I. Fragata, S. Seabra, G. S. Faria, M. Santos, M. R. Rose, M. Santos, and M.
603 Matos. 2017. Predictable phenotypic, but not karyotypic, evolution of historically
604 differentiated populations. Sci. Rep. 7:913.

605 Simões, P., M. R. Rose, A. Duarte, R. Gonçalves, and M. Matos. 2007. Evolutionary

606 domestication in *Drosophila subobscura*. J. Evol. Biol. 20:758–766.

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

610 Stern, D. L. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet.
611 14:751–764.

612 Teotónio, H., I. M. Chelo, M. Bradić, M. R. Rose, and A. D. Long. 2009. Experimental
613 evolution reveals natural selection on standing genetic variation. Nat. Genet.
614 41:251–257.

615 Teotónio, H., M. Matos, and M. Rose. 2002. Reverse evolution of fitness in *Drosophila*
616 *melanogaster*. J. Evol. Biol. 15:608–617.

617 Teotónio, H., and M. R. Rose. 2000. Variation in the reversibility of evolution. Nature
618 408:463–466.

619 Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental
620 Tests of the Roles of Adaptation, Chance, and History in Evolution. Science
621 267:87–90.

622 Whitlock, M. C., P. C. Phillips, F. B. Moore, and S. J. Tonsor. 1995. Multiple Fitness
623 Peaks and Epistasis. Annu. Rev. Ecol. Syst. 26:601–629.

624 Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New
625 York.

626 Wickham, H. 2007. Reshaping Data with the reshape Package. J. Stat. Softw. 21:1–20.

627 Wiser, M. J., N. Ribeck, R. E. Lenski, J. S. Littell, C. J. Muller, K. A. Dunne, A. V
628 Vecchia, L. R. Leung, Y. Qian, Y. Zhang, J. L. Betancourt, and G. J. Mccabe.
629 2013. Long-Term Dynamics of Adaptation in Asexual Populations. Science
630 342:1364–1367.

631 Wright, S. 1931. Evolution in Mendelian Populations. Genetics 16:97–159.

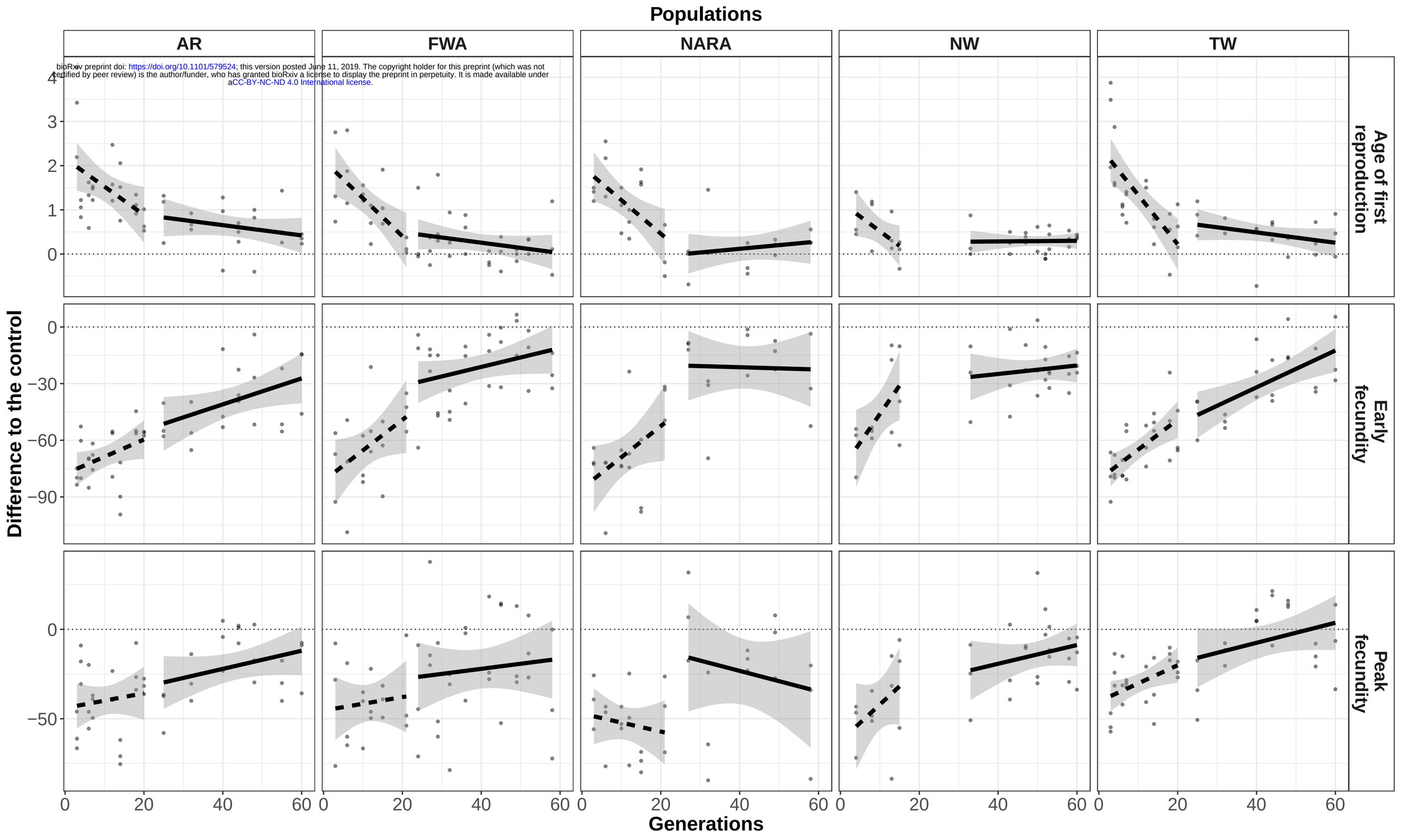
632

633

Figure Legends

Figure 1 - Short and long-term evolutionary trajectories for fecundity related traits for the 5 sets of populations studied. Age of first reproduction (number of days), Early fecundity (number of eggs), Peak fecundity (number of eggs) are represented. Points represent mean values for each replicate at each generation. Dashed lines indicate short term period and full line indicates long-term period. Shaded area represents 95% confidence intervals estimated from the regression, using mean replicate population values.

Figure 2 - Short and long-term evolutionary trajectories for female and male starvation resistance for the 5 sets of populations studied. Male starvation resistance (in hours) and Female starvation resistance (in hours) are represented. Points represent mean value for each replicate at each generation. Dashed lines indicate short term period and full line indicates long-term period. Shaded area represents 95% confidence intervals estimated from the regression, using mean replicate population values.



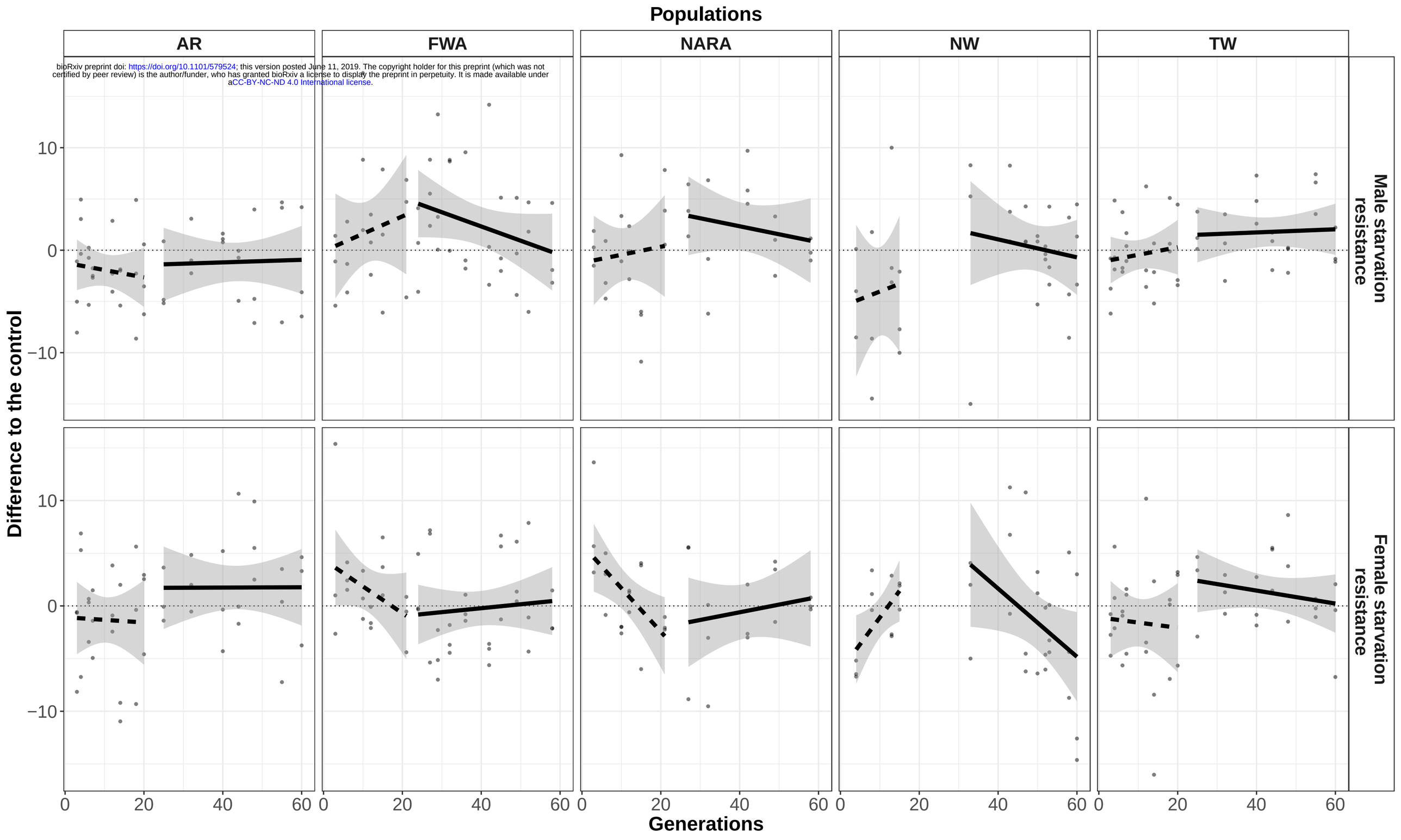


Table 1 - Comparison of evolutionary rates between different years or locations for short or longer periods.

	Comparison	Age First Reprod	Early Fecundity	Peak Fecundity	Fem Starv Resist	Male Starv Resist
Short	Arrábida 2001 vs 2005	0.7534	0.1462	0.05	**0.0004** ***	0.3682
	Sintra 1998 vs 2001	**0.0408** m.s.	**0.0066** *	0.185	**0.0014** **	0.6638
	Sintra 1998 vs 2005	0.3634	**0.0292** m.s.	**0.0392** m.s.	**0** ***	0.9202
	Sintra 2001 vs 2005	0.1888	0.6282	0.189	**0.0294** m.s.	0.4228
	Arrábida vs Sintra	0.063	0.2862	**0.0496** n.s	0.4724	0.1286
Long	Arrábida 2001 vs 2005	**0.0084** **	**0.002** **	**0.0004** ***	0.2442	0.191
	Sintra 1998 vs 2001	0.069	**0.0014** **	0.1594	**0.0004** ***	**0.0478** n.s.
	Sintra 1998 vs 2005	0.1188	0.0938	0.7108	**0.0002** ***	0.9632
	Sintra 2001 vs 2005	0.894	**0.0396** m.s.	0.2322	0.0862	**0.0106** *
	Arrábida vs Sintra	**0.0382** m.s.	**0.0036** **	**0.0114** *	0.3736	0.6238

Note: P-values were obtained by residual bootstraping of 10000 samples and estimated the fraction of these samples that were greater than 0 (see Material and Methods for more details). When $p<0.05$ (indicated in bold) significance levels after FDR correction are also presented (in superscript): *** $p<0.00044$ ($\alpha=0.001$); ** $0.00044<p<0.0044$ ($\alpha=0.01$); * $0.0044<p<0.022$ ($\alpha=0.05$); m.s. $0.022<p<0.044$ ($\alpha=0.1$); n.s. $p>0.044$ ($\alpha=0.1$)

Table 2 - Comparison of short and long term evolutionary rates between years and locations.

Comparison	Age First Reprod	Early Fecundity	Peak Fecundity	Fem Starv Resist	Male Starv Resist
Arrábida 2001 vs 2005	0.2376	**0.0042** **	0.7132	**0** ***	0.1444
Sintra 1998 vs 2001	0.1402	**0** ***	0.069	**0** ***	0.2288
Sintra 1998 vs 2005	0.6716	**0.006** *	**0.048** n.s.	**0** ***	0.9166
Sintra 2001 vs 2005	0.238	0.1714	0.5928	**0.0078** *	0.0702
Arrábida vs Sintra	0.2628	0.6254	0.7414	0.2986	0.1144

Note: P-values were obtained by residual bootstraping of 10000 samples and estimated the fraction of these samples that were greater than 0 (see Material and Methods for more details). Significant results are indicated in bold. When $p<0.05$ (indicated in bold) significance levels after FDR correction are also presented (in superscript): *** $p<0.00044$ ($\alpha=0.001$); ** $0.00044<p<0.0044$ ($\alpha=0.01$); * $0.0044<p<0.022$ ($\alpha=0.05$); m.s. $0.022<p<0.044$ ($\alpha=0.1$); n.s. $p>0.044$ ($\alpha=0.1$)