Testing for genetic assimilation with phylogenetic comparative analysis: Conceptual, methodological, and statistical considerations

Alex R. Gunderson[1*] and Liam J. Revell[2,3]

1. Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA, USA

2. Department of Biology, University of Massachusetts Boston, Boston, MA, USA

3. Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile

Running header: Evolution of phenotypic plasticity

* corresponding author; email: agunderson@tulane.edu

Abstract

Genetic assimilation is a process that leads to reduced phenotypic plasticity during adaptation to novel conditions, a potentially important phenomenon under global environmental change. Null expectations when testing for genetic assimilation, however, are not always clear. For instance, the statistical artifact of regression to the mean could bias us towards detecting genetic assimilation when it has not occurred. Likewise, the specific mechanism underlying plasticity expression may affect null expectations under neutral evolution. We used macroevolutionary numerical simulations to examine both of these important issues and their interaction, varying whether or not plasticity evolves, the evolutionary mechanism, trait measurement error, and experimental design. We also modified an existing reaction norm correction method to account for phylogenetic non-independence. We found: 1) regression to the mean is pervasive and can generate spurious support for genetic assimilation; 2) experimental design and post-hoc correction can minimize this spurious effect; and 3) neutral evolution can produce patterns consistent with genetic assimilation without constraint or selection, depending on the mechanism of plasticity expression. Additionally, we re-analyzed published macroevolutionary data supporting genetic assimilation, and found that support was lost after proper correction. Considerable caution is thus required whenever investigating genetic assimilation and reaction norm evolution at macroevolutionary scales.

Keywords: plasticity, acclimation, global change, evolution, reaction norm

Introduction

Phenotypic plasticity is an important phenomenon in evolutionary biology with far-reaching implications (Ghalambor et al. 2007; Murren et al. 2015). For example, plasticity in a given trait can facilitate evolutionary radiations (Pfennig et al. 2010), contribute to local adaptation (Price et al. 2003), and dampen selection on underlying genetic variation (Price et al. 2003; Crispo 2008). Interest in the expression and evolution of phenotypic plasticity has increased in recent years due to the intense selection pressure that human activity has imposed on the natural world (Chevin et al. 2010; Kingsolver and Buckley 2017). Taxa with greater plasticity in traits that are important for tolerating human-induced changes to the environment are predicted to fare better under the novel conditions created by global warming, urbanization, and species introductions (Somero 2010; Gunderson and Stillman 2015; Seebacher et al. 2015; Gunderson et al. 2017). For instance, plasticity in thermal physiology is predicted to decrease the negative impact of global warming on ectotherms (Gunderson et al. 2017; Riddell et al. 2018; Rohr et al. 2018; Morley et al. 2019). Greater knowledge of the forces that shape and constrain phenotypic plasticity is crucial both for our fundamental understanding of ecology and evolution and our ability to predict and mitigate the consequences of global change.

Adaptation to novel conditions is sometimes associated with a reduction or loss of plasticity in traits under selection, a process that is known as genetic assimilation (Waddington 1953; Schlichting and Pigliucci 1998). Some have argued that genetic assimilation is a key mechanism of adaptive evolutionary change (West-Eberhard 2003; Crispo 2007). For example, the canalization of formerly induced phenotypes can provide a means of adaptive phenotypic divergence during local adaptation and adaptive radiation (Ehrenreich and Pfennig 2016;

67 Martin et al. 2016; Gunter et al. 2017; Schneider and Meyer 2017). In addition, genetic
68 assimilation may be an outcome of population responses to anthropogenic global change.
69 Lande (2009) found that the evolution of increased phenotypic plasticity followed by genetic
70 assimilation should result from selection due to dramatic changes in the environment.

71 The intersection of adaptation, genetic assimilation, and global change is highlighted by
72 the Baseline Tolerance/Tolerance Plasticity Trade-off Hypothesis (hereafter referred to as the
73 Trade-off Hypothesis). Emerging from the field of evolutionary physiology, the hypothesis
74 states that, as organisms evolve greater baseline tolerance to extreme temperatures (heat or
75 cold), plasticity in their thermal tolerance should actually decrease (van Heerwaarden and
76 Kellermann 2020). In other words, high levels of constitutive thermal tolerance evolve by
77 genetic assimilation, resulting in thermal tolerance phenotypes that are less sensitive to
78 environmental temperature variation (Sikkink et al. 2014). Support for this hypothesis comes
79 from a variety of organisms, including marine and aquatic vertebrates and invertebrates at both
80 the intraspecific and interspecific levels (e.g., Stillman 2003; Esperk et al. 2016; Comte and
81 Olden 2017; Armstrong et al. 2019). One of several implications of the Trade-off Hypothesis is
82 that the genetic mechanisms associated with the expression of physiological plasticity serve as
83 the raw material for thermal adaptation. Therefore, as organisms evolve greater baseline
84 tolerance in response to global change, they may lose both the ability to evolve even greater
85 tolerance and their capacity to track shorter-term environmental changes via a plastic response
86 (van Heerwaarden and Kellermann 2020).

87 Hypotheses that make predictions about trait change over time among individuals
88 and/or taxa, such as the Trade-off Hypothesis, are prone to spurious support due to the well-

89 known statistical phenomenon called *regression to the mean*. A consequence of this
90 phenomenon is that when a particular trait is measured repeatedly, the experimental units
91 (e.g., species, populations, or individuals) that start with particularly high or low values relative
92 to the population mean are also likely to record the greatest change in value over time as
93 subsequent measures regress towards the mean (Barnett et al. 2005). This is not a real
94 biological effect: rather, it is an artifact created by sampling and measurement error.
95 Regression to the mean is a common problem in repeated-measures analyses, and statistical
96 methods have been developed to overcome the spurious results that it can create. For
97 example, data resampling techniques have been applied to generate null distributions of trait
98 change (Jackson and Somers 1991), and methods to statistically remove regression to the mean
99 effects prior to statistical analysis have also been proposed (Kelly and Price 2005). These
100 approaches are most often employed at the intraspecific level (Ghalambor et al. 2015; Deery et
101 al. 2021), but accounting for regression to the mean is also necessary, though more often
102 neglected, in interspecific macroevolutionary analyses (Baker et al. 2015).
103 Motivated by current debate surrounding the Trade-off Hypothesis and increasing
104 interest in the evolution of phenotypic plasticity and genetic assimilation under global change
105 (Gunderson and Stillman 2015; Kelly 2019; Sasaki and Dam 2019; van Heerwaarden and
106 Kellermann 2020; Barley et al. 2021; Preston et al. 2021; Sasaki and Dam 2021), we used
107 simulations to explore the more general consequences of regression to the mean for tests of
108 genetic assimilation at the macroevolutionary level. We conducted simulations that assumed
109 mechanistically different models for plasticity evolution, as well as different experimental
110 sampling designs for measuring species phenotypic traits. In addition, we conducted

simulations with and without the assumption that phenotypes are measured with error. We found that measurement error alone can be sufficient to recover a macroevolutionary pattern that is consistent with genetic assimilation due entirely to the phenomenon of regression to the mean. This problem can be avoided, however, by using an appropriate experimental design. In addition, we found that macroevolutionary patterns consistent with genetic assimilation are the null expectation under random Brownian motion evolution for some models by which plasticity evolves. Finally, we applied a statistical method designed to remove the regression to the mean effect from compromised data (Kelly and Price 2005), that we modified to take non-independence due to phylogeny into account. We found that our approach can reduce the effect of regression to the mean in some, but not all, circumstances. We applied this modified approach to published phylogenetic comparative data that was thought to be consistent with the Trade-off Hypothesis, and showed that the relationship is not significant when properly corrected for regression to the mean.

Methods

In our simulations, we assumed that there were two different environments that induce different phenotypes. The difference between trait values in each environment represented the magnitude of phenotypic plasticity and, in a two-environment system, is equivalent to the slope of the reaction norm. In any given simulation, we first generated a random phylogenetic tree with 50 terminal nodes using the *pbtree* function in the phytools R package (Revell 2012; R Development Core Team 2021). Depending on the simulation, we either modeled the evolution of plasticity across the phylogeny using the *fastBM* function or assumed no phenotypic

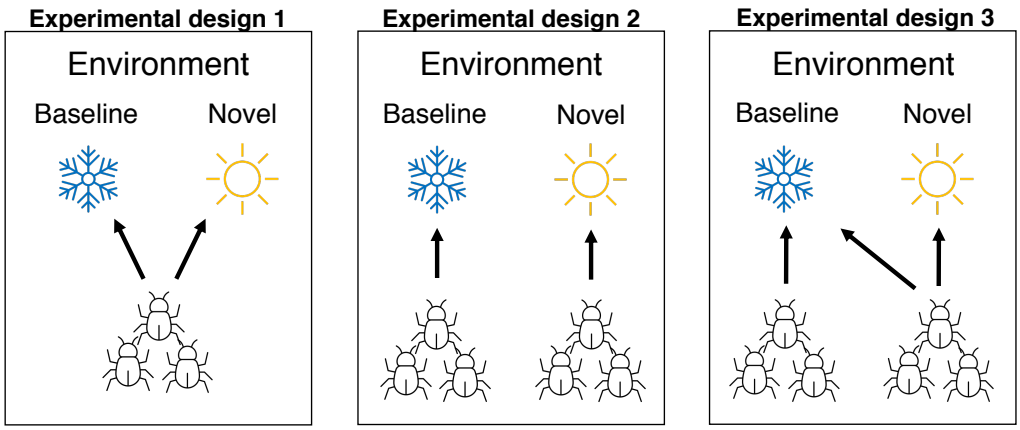

Figure 1. Three experimental designs that we simulated for estimating baseline phenotype and phenotypic plasticity of species. In Experimental Design 1, baseline phenotype and plasticity are estimated from a single group of individuals that have their phenotype measured twice: once after exposure to each environment. In Experimental Design 2, there are two groups of individuals that have their phenotype measured only once, after exposure to either the baseline or novel environment. In Experimental Design 3, there are also two groups of individuals. One group has their phenotype measured only after exposure to the baseline condition, and are used only to estimate the baseline phenotype. The other group of individuals have their phenotypes measured twice (as in Experimental Design 1), but they are only used to estimate phenotypic plasticity. Thermal environments are shown simply as an example.

evolution at all (see below for details). In addition, we assumed a certain degree of sampling error in the measurement of phenotypes in each environment. We did so by drawing phenotypes for individuals randomly from a Gaussian distribution with a mean value at the true phenotypic population mean. In practice, this type of error can emerge both from the random sampling of individuals with a population that varies, and by methodological or instrument imprecision. Since both types of error can contribute to the phenomenon of regression to the mean, we effectively combined them in our simulation. For each simulation run, we calculated the phylogenetic correlation coefficient for the association between the baseline phenotype and the reaction norm slope using the *phyl.vcv* function of phytools. Starting conditions assumed a baseline phenotypic trait value of 0 and a phenotypic value in the novel environment of 1. In simulations where evolution occurred, the Brownian motion rate

144 parameter (σ^2) was set at 1, 0.1, or 0.01 for respective traits. Phenotypic measurement error
145 was set to have variances (s^2) of 1, 0.1, or 0.01.
146 We predicted that experimental design would influence whether or not regression to
147 the mean biased our results. As such, we modeled three different experimental designs for
148 measuring the phenotypes of each taxon in each environment (Figure 1). In Design 1, we
149 assumed that each individual has their phenotype measured twice in total: once after
150 acclimation to each environment. Therefore, in this case the same individuals are used to
151 estimate both the baseline phenotype and the reaction norm. In Design 2, we assumed that
152 different individuals are acclimated to each environment prior to phenotypic measurement. In
153 this case, individuals exposed to the baseline environment are used to estimate the baseline
154 phenotype, and the difference in phenotype between the two sets of individuals is used to
155 estimate the reaction norm. Finally, in Design 3, we assumed that one set of individuals is
156 acclimated only to the baseline environment prior to phenotypic trait measurement, and that a
157 second set of individuals have their phenotypes measured twice: once after acclimation to each
158 of the two environments. The former individuals are used only to estimate baseline
159 phenotypes, and the latter individuals are used only to estimate the slope of the reaction norm.
160 We also tested whether regression to the mean could emerge as a significant problem
161 when phenotypic plasticity evolves, and did so assuming a neutral evolutionary process (in this
162 case, Brownian motion). We used two different models for the evolution of phenotypic
163 plasticity, that we refer to as Linked Phenotypes and Unlinked Phenotypes. In the Linked
164 Phenotypes model, we assumed that the slope and intercept of the reaction norms evolve
165 directly, with the phenotype that is expressed in each environment evolving indirectly as a

166 byproduct. This is one of the most common ways in which the evolution of phenotypic plasticity
167 is conceptualized and studied theoretically (Lande 2009; Chevin et al. 2010; Chevin et al. 2013).
168 We refer to this as the Linked Phenotypes model because the phenotypes that are expressed in
169 each environment are each determined by the slope and the intercept, linking them under a
170 neutral evolutionary process. In the Unlinked Phenotypes model, the phenotype that is
171 expressed in a given environment evolves directly and independently of the phenotype
172 expressed in the other environment. Therefore, the phenotypes that are expressed in each
173 environment lack genetic covariance (Via and Lande 1985), and the slope and intercept of the
174 reaction norm emerge merely as a byproduct of the trait values expressed in each environment.
175 This model is based on the concept that parts of a reaction norm can evolve independently of
176 one another as outlined in Ghalambor et al. (2007). Using each model, we conducted
177 simulations in which the evolving parameters diverge via Brownian motion, and then estimated
178 the phylogenetic correlation coefficient for an association between baseline phenotype and the
179 reaction norm slope, as described above.
180 When possible, we applied a statistical correction developed by Kelly and Price (2005) to
181 remove the regression to the mean effect from reaction norm values. The method was not
182 originally designed for use with phylogenetic comparative data, but we modified it to that
183 purpose here. We use the correction only with data simulated under Experimental Designs 1
184 and 2, as the method cannot be applied to Experimental Design 3. The approach is based on the
185 premise that if the plastic response is constrained by the baseline phenotype, then phenotypic
186 variation should be reduced under novel conditions and that phenotypes in the baseline and
187 novel conditions should be correlated within experimental units (Kelly and Price 2005). The

188 approach is implemented as follows. First, we estimate the phylogenetic variance in

189 phenotypes in each environment (s_i) using the *phyl.vcv* function, then we conduct a Pittman's

190 test of homogeneity of variance. If there is no difference in variance between the phenotypes,

191 then corrected reaction norms ($\widehat{D}^*$) are calculated using the following function:

192

193 Equation 1: $\widehat{D}^* = r(X_1 - \bar{X}_1) - (X_2 - \bar{X}_2)$

194

195 where *r* is the phylogenetic correlation coefficient for the relationship between phenotypes in

196 the baseline and novel environments, X_i is the vector of phenotypes in a given environment,

197 and $\bar{X}_i$ is the phylogenetic phenotypic mean in a given environment. If the phenotypic variance

198 differs between environments, then an adjusted correlation coefficient ($\hat{r}$) is used in Equation

199 1, calculated as follows:

200

201 Equation 2: $\hat{r} = \frac{2rs_1s_2}{s_1^2 + s_2^2}$

202

203 Where s_i^2 is the estimated phylogenetic variance in the environment *i*. $\widehat{D}^*$ is then used as the

204 reaction norm value for subsequent analyses as described above.

205 Using our modified Kelly and Price (2005) approach, we reanalyzed previously published

206 data on heat tolerance plasticity across Nudibranch species (Armstrong et al. 2019). The

207 authors of that study used what we refer to as Experimental Design 2 (Figure 1), and originally

208 found support for the Trade-off Hypothesis, as Nudibranch species with greater baseline heat

tolerance had reduced heat tolerance plasticity. However, they did not adjust values for potential regression to the mean as we do here.

Results

Results without evolution but with measurement error

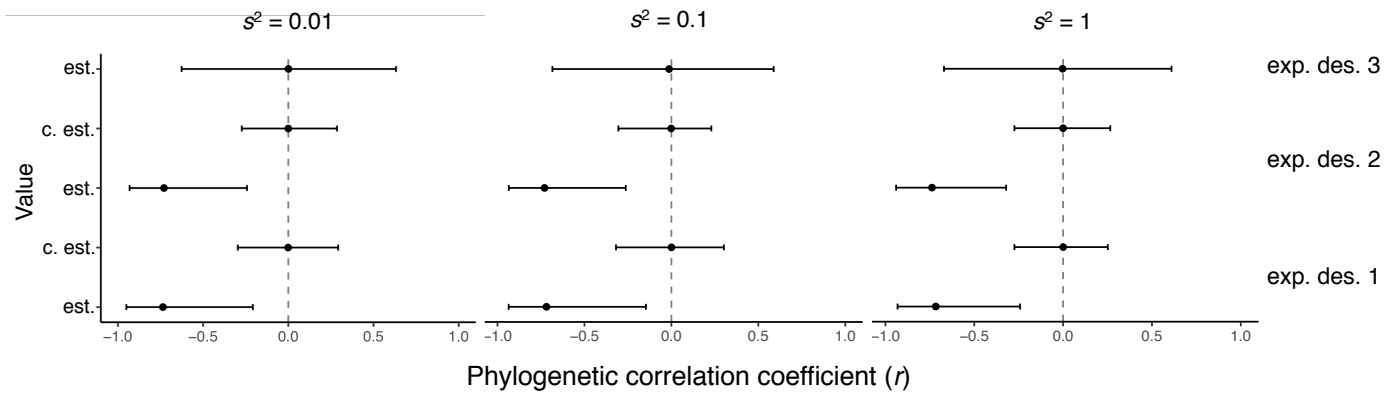

Figure 2. Summary of phylogenetic correlation coefficients between baseline phenotype and reaction norm slope from simulations in which plasticity is present but does not diverge among taxa, and phenotypes in each environment are measured with error. Points indicate the median and black lines the central 95% of correlation coefficients from 1000 simulations. The error variance of simulations is shown above each column. Alternating shading differentiates different simulated experimental designs, which are written on the righthand side of the figure. "est." denotes estimated correlation coefficients given measurement error, while "c. est" denotes those same correlation coefficients adjusted to remove the effect of regression to the mean following our modified Kelly and Price (2005) correction. Adjusted results are not given for Experimental Design 3 because the adjustment cannot be applied to data collected with that design.

We first present results of simulations in which phenotypic plasticity is present but does not diverge among taxa. In this case, the only source of variation among taxa is measurement error. If regression to the mean is not an issue, we would expect the correlation coefficients generated in our simulations to be centered on 0.

Instead, we found that measurement error generated a pattern consistent with genetic assimilation (that is, a negative correlation between baseline phenotype and reaction norm

220 slope) due to regression to the mean, but only for Experimental Designs 1 and 2 (Figure 2). By
221 contrast, regression to the mean does not bias results using Experimental Design 3 (Figure 2).
222 This finding does not dependent on the level of plasticity organisms exhibit. For example, even
223 if there is no plasticity in the trait, a negative relationship between baseline phenotype and
224 reaction norm slope will emerge with Experimental Designs 1 and 2 if the phenotypes in each
225 environment are measured with error (see Supplemental Figure S1; see also Deery et al. 2001
226 for an individual-level example). The bias induced by Experimental Designs 1 and 2 is largely
227 removed, however, when the simulated data are adjusted using our phylogenetic modification
228 of the Kelly and Price (2005) correction (Figure 2 "corrected estimates").
229
230 *Results with neutral evolution of phenotypic plasticity*
231 A pattern consistent with genetic assimilation is unlikely to evolve under neutral processes if
232 phenotypic plasticity evolves via the Linked Phenotype mechanism (i.e., if reaction norm slope
233 and intercept evolve directly) (Figure 3 "true" values). However, this assumes that there is no
234 measurement error when estimating phenotypes. When measurement error is considered,
235 regression to the mean can produce spurious support for genetic assimilation with
236 Experimental Designs 1 and 2, though this effect depends on the magnitude of measurement
237 error relative to the rate of evolution (Figure 3 "estimates" values). The greater the
238 measurement error, the more likely that regression to the mean can become a problem.
239 Adjusting data generated by Experimental Designs 1 and 2 using our modified Kelly and Price
240 correction removes the regression to the mean effect (Figure 3 "corrected estimates"). These
241 results, however, also highlight the importance of judicious application of the Kelly and Price

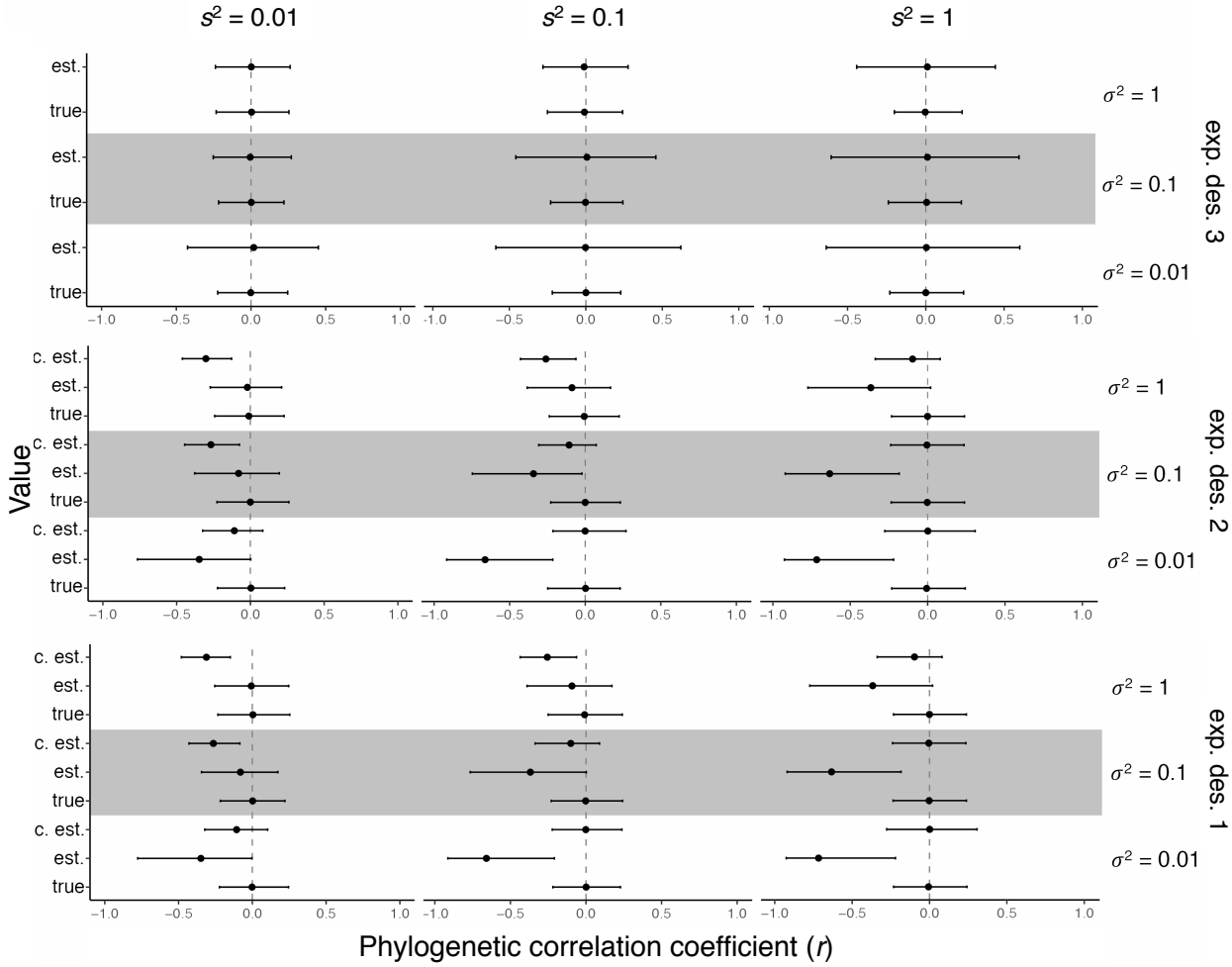

Figure 3. Summary of phylogenetic correlation coefficients between baseline phenotype and reaction norm slope from simulations in which plasticity evolved by Brownian motion via the Linked Phenotypes mechanism. Each panel shows the result of 1000 simulations. Each row shows the results of simulations with a different experimental design (Figure 1): *Bottom*, Design 1. *Middle*: Design 2. *Top*: Design 3. Alternating shaded areas differentiate simulations with different Brownian motion rate parameters (σ^2). Each column contains simulations with a different error variance for phenotypic estimation (s^2). "true" denotes the true phylogenetic correlation values from the simulations. "est." denotes estimates of the true correlation coefficients given measurement error. "c. est" denotes estimated correlation coefficients corrected to remove the effect of regression to the mean following our modified Kelly and Price (2005) adjustment. Adjusted results are not given for Experimental Design 3 because the method cannot be applied to data collected with that design.

correction. This is because if there is a weakly negative to positive correlation between baseline phenotype and plasticity, the Kelly and Price correction can *create* a spurious negative

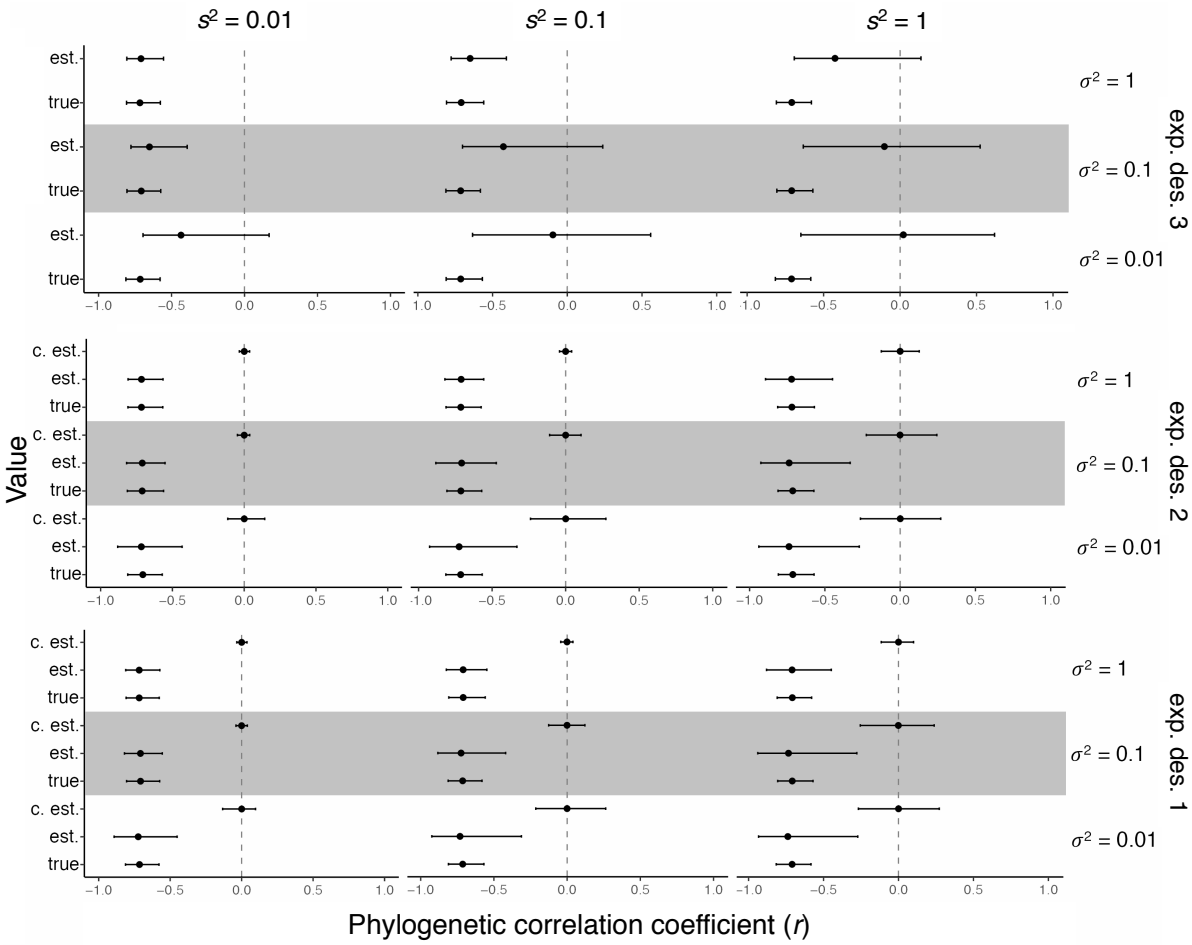

Figure 4. Summary of phylogenetic correlation coefficients between baseline phenotype and reaction norm slope from simulations in which plasticity evolved by Brownian motion via the Unlinked Phenotypes mechanism. Each panel shows the result of 1000 simulations. Each row shows the results of simulations with a different experimental design (Figure 1): *Bottom*, Design 1. *Middle*: Design 2. *Top*: Design 3. Alternating shaded areas differentiate simulations with different Brownian motion rate parameters (σ^2). Each column contains simulations with a different error variance for phenotypic estimation (s^2). "true" denotes the true phylogenetic correlation values from the simulations. "est." denotes estimates of the true correlation coefficients given measurement error. "c. est" denotes estimated correlation coefficients adjusted to remove the effect of regression to the mean following our modified Kelly and Price (2005) correction. Adjusted results are not given for Experimental Design 3 because the method cannot be applied to data collected with that design.

relationship between the two variables (for example, see Figure 3 results with Experimental Design 2, σ^2 = 1 and s^2 = 0.01).

If phenotypic plasticity evolves via the Unlinked Phenotypes mechanism (phenotypes in each environment evolve directly and independently of one another), neutral evolution will yield a negative relationship between baseline phenotype and reaction norm slope (Figure 4 "true" values). Furthermore, the negative relationship between these traits is maintained in the presence of measurement error (Figure 4 "estimated" values). In this case, however, adjusting the data using the Kelly and Price correction mostly eliminates the true negative relationships between the traits (Figure 4 "corrected" estimates). Conversely, if data are collected using Experimental Design 3, the true negative relationship between the traits is largely maintained. Nonetheless, as measurement error is increased the capacity for Experimental Design 3 to yield the true negative relationship decreases (Figure 4).

Reanalysis of Nudibranch heat tolerance data

We re-analyzed data on heat tolerance plasticity in Nudibranchs (Armstrong et al. 2019) by applying our phylogenetic adaptation of the Kelly and Price (2005) correction. In the original analysis, the authors found a significant negative relationship between independent contrasts for baseline heat tolerance and heat tolerance plasticity ($P < 0.005$, Armstrong et al. 2019). In our re-analysis, we found that a negative relationship remained, but was no longer significant at the $\alpha = 0.05$ level ($P = 0.069$; Figure 5). In addition, the variance explained dropped from 88% to 51%. The maintenance of a negative relationship and relatively high explanatory power suggests that there may be a real biological relationship between baseline tolerance and plasticity in this group, but the need to adjust for regression to the mean raises the bar for the evidence required.

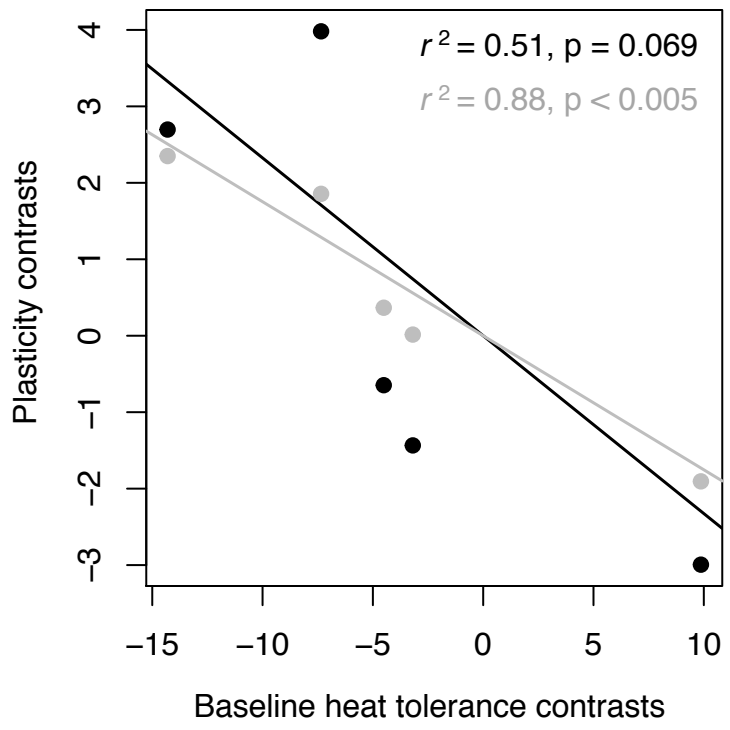

Figure 5. Nudibranch heat tolerance plasticity data from Armstrong et al. (2019). Phenotypes were collected using Experimental Design 2 (Figure 1). Original data and analysis in gray. Re-analysis applying the modified Kelly and Price (2005) method to remove regression to the mean effects in black.

Discussion

Herein, we have identified three general patterns important for testing and understanding the evolution of genetic assimilation at the macroevolutionary level: 1) regression to the mean is a pervasive problem; 2) experimental design and data correction can minimize the effect of regression to the mean; and 3) neutral evolutionary processes can produce patterns consistent with genetic assimilation, depending on the linkage between trait expression in different environments. Below, we discuss each of these findings in greater detail.

Regression to the mean when plasticity does not diverge

When phenotypic plasticity was present but did not diverge among taxa, measurement error alone generated spurious correlations between baseline phenotype and plasticity under Experimental Designs 1 and 2 (Figure 2 “estimated” values). Furthermore, the magnitude of the correlation coefficient was affected very little by the magnitude of measurement error simulated (Figure 2). This is because error variation cannot be swamped by biological variation when there are no phenotypic differences among taxa. These results clearly demonstrate that regression to the mean can be a significant problem when testing for genetic assimilation at the macroevolutionary level. The choice of experimental design, however, also matters considerably. In particular, when data are collected using Experimental Design 3, regression to the mean is not a persistent problem (Figure 2).

Why are Experimental Designs 1 and 2 susceptible to regression to the mean while Experimental Design 3 is not? The crux of the matter is that with Designs 1 and 2, the reaction norm slope depends on the estimated baseline phenotype. Designs 1 and 2 have only one estimate of baseline phenotype. That single baseline phenotype is also used to estimate plasticity. If measurement error causes an estimate of baseline phenotype to be above the true value, that will drive down the estimated reaction norm slope. Conversely, if error causes a baseline estimate to be below its true value, this will in turn drive up the estimated reaction norm slope. With Design 3, this cannot happen because the baseline phenotype and reaction norm are estimated using separate samples.

Design 3 demands two independent estimates of phenotype under baseline conditions, which we will call e1 and e2. One of those estimates (let’s say e1) is used only as the baseline

phenotype, but does not contribute to the calculation of plasticity. The other estimate (e2) contributes only to the calculation of plasticity. Under this design, if measurement error causes e1 to be above or below the true value it will have no effect on the plasticity estimate, which depends only on e2.

All hope is not lost with Experimental Designs 1 and 2, however. Using our modified Kelly and Price correction, phenotypic values can be adjusted such that the bias towards a spurious negative relationship is removed (Figure 2 “corrected” values; though see below for caveats). Next, we discuss the implications of these different approaches when plasticity itself evolves.

Regression to the mean under neutral evolution of plasticity (Linked Phenotype model)

Under the Linked Phenotypes model, the intercept (essentially, the baseline phenotype) and reaction norm slope evolve via Brownian motion. Under this model, neutral evolution does not produce a negative relationship between baseline phenotype and plasticity under any experimental design (Figure 3 “true” values). The introduction of measurement error, however, can again lead to regression to the mean under Experimental Designs 1 and 2 (Figure 3 “estimated” values). The degree to which regression to the mean is a problem depends on the magnitude of measurement error relative to the variance in true trait values among taxa as denoted by the Brownian motion rate parameter. If measurement error is large and/or phenotypic variation among taxa is low, regression to the mean emerges consistently for Experimental Designs 1 and 2. By contrast, if measurement error is small and phenotypic

320 variation among taxa is large, regression to the mean is much less of an issue (Figure 3
321 "estimated" values).
322 Applying the Kelly and Price correction to data collected under Experimental Designs 1
323 and 2 can remove the effect of regression to the mean when it is present (Figure 3 "corrected"
324 values). Our results, however, also highlights the caveat that a Kelly and Price correction should
325 only be applied if the initial analysis yields a significant negative relationship. When initial
326 analyses yield no significant negative relationship, applying the Kelly and Price correction can
327 itself create a spurious negative relationship between baseline phenotype and plasticity.
328 Experimental Design 3 again avoids the problem of regression to the mean due to the
329 estimation of two baseline phenotypes (see above).
330
331 *Regression to the mean under neutral evolution of plasticity (Unlinked Phenotypes model)*
332 Under the Unlinked Phenotypes model, the phenotypes in each environment evolve
333 independently of one another, and the magnitude of plasticity emerges as a byproduct of the
334 difference between the two phenotypic values. Under these conditions, a true negative
335 association between baseline phenotype and plasticity consistently arises under Brownian
336 motion evolution (Figure 4 "true" values). In other words, there are conditions in which neutral
337 evolution alone can produce a pattern consistent with genetic assimilation. This is in contrast to
338 the most common explanations for genetic assimilation, which invoke some combination of
339 selection and/or genetic constraint to explain why a reduction in plasticity evolves (West-
340 Eberhard 2003; Pigliucci et al. 2006; Ehrenreich and Pfennig 2016).

341 Unlike the Linked Phenotypes model, experimental error has little to effect on the
342 correlation coefficients calculated with the Unlinked Phenotypes model using Experimental
343 Designs 1 and 2 (Figure 4 “estimate” values). The reason for this is that phenotypic evolution via
344 Brownian motion in the Unlinked Phenotypes model amounts to random independent
345 increases or decreases in the phenotypes expressed in each environment, while experimental
346 error leads to spurious random increases or decreases in the phenotypes measured in each
347 environment, which from a statistical standpoint is the same process.

348 When data from the Unlinked Phenotypes model are adjusted using the Kelly and Price
349 correction, the negative relationship between baseline phenotype and plasticity disappears
350 (Figure 4, “corrected” values). In this case, however, the correction does not remove a spurious
351 relationship due to regression to the mean. Instead, it removes a true relationship that resulted
352 from a neutral evolutionary process. This occurs because of the conceptual foundation upon
353 which the Kelly and Price correction is predicated: it assumes that a hallmark of regression to
354 the mean is that phenotypic values between treatments (in this case, environments) within
355 experimental units are uncorrelated with one another. This, of course, is the precise pattern
356 that neutral evolution will produce if the phenotypes induced by plasticity are unlinked. Our
357 finding highlights that biological variation and error variation can be difficult to separate (Kelly
358 and Price 2004; Hansen and Bartoszek 2012), and statistical correction for regression to the
359 mean is not a panacea.

360 Correlation coefficients estimated using Experimental Design 3 can maintain the true
361 negative relationship between baseline phenotype and plasticity, but only if the measurement
362 error is not too large. As measurement error grows relative to the true phenotypic variance

among taxa, estimated correlation coefficients move away from the true values and towards 0 (Figure 4).

Implications

Our results make clear that regression to the mean is a serious statistical problem that can confound tests of genetic assimilation at macroevolutionary scales. Whether or not regression to the mean is an issue in any particular study will depend on the experimental design used. Unfortunately, the problematic experimental designs (our Designs 1 and 2) are those most often used to test for plasticity and genetic assimilation. This is certainly true within the field of thermal physiology, where tests of the Trade-Off Hypothesis, which states that greater thermal tolerance evolves via genetic assimilation (Heerwaarden and Kellerman 2020), are almost invariably based on data collected using Experimental Designs 1 and 2 (e.g., Armstrong et al. 2019). These studies typically find support for the Trade-off Hypothesis but do not account for regression to the mean.

Some evolutionary physiologists have used an alternative approach to test the Trade-off Hypothesis, in which thermal tolerance plasticity is correlated with habitat temperature (van Heerwaarden and Kellermann 2020). This approach avoids the regression to the mean problem, but it is important to recognize that modeling the association between plasticity and habitat tests a different hypothesis than modeling the association between plasticity and baseline phenotype. The latter directly pertains to the evolution of within-individual trait associations, trade-offs, and genetic assimilation. The former does not.

384 Fortunately, data collected using Experimental Designs 1 and 2 can be successfully
385 adjusted to remove regression to the mean using the Kelly and Price method. For example,
386 when plasticity is invariant among taxa, adjusting the data removes spurious relationships
387 between baseline phenotype and plasticity due to measurement error (Figure 3). The same
388 holds under neutral evolution when there is linkage between the traits expressed in each
389 environment (Figure 3). The caveat is that adjusting data using the Kelly and Price method can
390 incorrectly obfuscate a true negative relationship between baseline phenotype and plasticity if
391 the traits expressed in each environment are unlinked (Figure 4).
392 The need to adjust data can be avoided by using Experimental Design 3. This approach
393 does not lead to spurious negative relationships between baseline phenotype and plasticity
394 (Figures 3 and 4). It also recovers true negative relationships when they occur (Figure 4).
395 Unfortunately, however, this experimental design is significantly more labor intensive that
396 Designs 1 and 2. It requires that we estimate phenotypes on more groups of individuals (3
397 versus 2), and its reliability depends on measurement error in the estimation of phenotypic trait
398 means for each environment, making large sample sizes crucial. There are issues with
399 Experimental Design 3 as well. For example, when individuals have their phenotypes measured
400 more than once, carryover effects of previous experience can influence results (O'Connor et al.
401 2014). Furthermore, for some taxa and traits, measuring the same individual twice will simply
402 not be possible. One solution in such cases is to have three groups of individuals per taxa: two
403 that have their phenotype measured under baseline conditions, and one that has phenotype
404 measured under novel conditions. One baseline group is used to calculate plasticity, the other is

405 used as the baseline phenotype estimate: the key is that different baseline measurements are
406 used for each.

407 Whether or not induced phenotypes are genetically linked influences null expectations
408 for reaction norm evolution. Specifically, a macroevolutionary pattern consistent with genetic
409 assimilation can arise under a neutral evolutionary process in the absence of selection and
410 genetic linkage between phenotypes. The importance of genetic linkage for reaction norm
411 evolution has long been appreciated. For example, the classic work of Via and Lande (1985)
412 demonstrated that genetic covariance between induced phenotypes influences the trajectory
413 of reaction norm evolution under selection. Unfortunately, genetic linkage between
414 environmentally-induced phenotypes is almost certainly trait, population, and environment
415 dependent, and in many instances knowledge about linkage is completely lacking. Our results
416 highlight the need for greater scrutiny of the genetic architecture of induced phenotypes.

417 Overall, we show that great care is necessary both in the design of macroevolutionary
418 tests for genetic assimilation, as well as in the interpretation of our results. A lack of correlation
419 between plasticity and baseline phenotype is not the correct null hypothesis in many cases due
420 to the statistical phenomenon of regression to the mean. Once this is recognized, the potential
421 solutions to the regression to the mean problem must be applied with care. Additionally, the
422 expected null hypothesis under neutral evolution depends on the mechanism that underlies the
423 expression of phenotypic plasticity. We hope that this analysis will be a helpful reference for
424 others in the design and analysis of future studies of hypothesized genetic assimilation in
425 phenotypically plastic traits.

426

Acknowledgements

Data availability:

Data and code to reproduce all the analyses of this study available at:

GitHub: https://github.com/liamrevell/Gunderson-and-Revell-2021.

References

Armstrong, E. J., R. L. Tanner, and J. H. Stillman. 2019. High heat tolerance is negatively correlated with heat tolerance plasticity in nudibranch mollusks. Physiological and Biochemical Zoology 92:430-444.

Baker, J., A. Meade, M. Pagel, and C. Venditti. 2015. Adaptive evolution toward larger size in mammals. Proceedings of the National Academy of Sciences 112:5093-5098.

Barley, J. M., B. S. Cheng, M. Sasaki, S. Gignoux-Wolfsohn, C. G. Hays, A. B. Putnam, S. Sheth et al. 2021. Limited plasticity in thermally tolerant ectotherm populations: evidence for a trade-off. Proceedings of the Royal Society B 288:20210765.

Barnett, A. G., J. C. Van Der Pols, and A. J. Dobson. 2005. Regression to the mean: what it is and how to deal with it. International Journal of Epidemiology 34:215-220.

Chevin, L.-M., R. Gallet, R. Gomulkiewicz, R. D. Holt, and S. Fellous. 2013. Phenotypic plasticity in evolutionary rescue experiments. Philosophical Transactions of the Royal Society B: Biological Sciences 368:20120089.

Chevin, L.-M., R. Lande, and G. M. Mace. 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biology 8:e1000357.

Comte, L., and J. D. Olden. 2017. Evolutionary and environmental determinants of freshwater fish thermal tolerance and plasticity. Global change biology 23:728-736.

Crispo, E. 2007. The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. Evolution: International Journal of Organic Evolution 61:2469-2479.

—. 2008. Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow. Journal of Evolutionary Biology 21:1460-1469.

Deery, S. W., J. E. Rej, D. Haro, and A. R. Gunderson. 2021. Heat hardening in a pair of *Anolis* lizards: constraints, dynamics and ecological consequences. Journal of Experimental Biology 224:jeb240994.

Ehrenreich, I. M., and D. W. Pfennig. 2016. Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. Annals of Botany 117:769-779.

462 Esperk, T., A. Kjaersgaard, R. J. Walters, D. Berger, and W. U. Blanckenhorn. 2016. Plastic and
463 evolutionary responses to heat stress in a temperate dung fly: negative correlation
464 between basal and induced heat tolerance? Journal of Evolutionary Biology 29:900-915.
465 Ghalambor, C. K., K. L. Hoke, E. W. Ruell, E. K. Fischer, D. N. Reznick, and K. A. Hughes. 2015.
466 Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in
467 nature. Nature 525:372-375.
468 Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus non-
469 adaptive phenotypic plasticity and the potential for contemporary adaptation in new
470 environments. Functional Ecology 21:394-407.
471 Gunderson, A. R., M. E. Dillon, and J. H. Stillman. 2017. Estimating the benefits of plasticity in
472 ectotherm heat tolerance under natural thermal variability. Functional Ecology 31:1529-
473 1539.
474 Gunderson, A. R., and J. H. Stillman. 2015. Plasticity in thermal tolerance has limited potential
475 to buffer ectotherms from global warming. Proceedings of the Royal Society of London
476 B: Biological Sciences 282:20150401.
477 Gunter, H. M., R. F. Schneider, I. Karner, C. Sturmbauer, and A. Meyer. 2017. Molecular
478 investigation of genetic assimilation during the rapid adaptive radiations of East African
479 cichlid fishes. Molecular Ecology 26:6634-6653.
480 Hansen, T. F., and K. Bartoszek. 2012. Interpreting the evolutionary regression: the interplay
481 between observational and biological errors in phylogenetic comparative studies.
482 Systematic biology 61:413-425.
483 Jackson, D. A., and K. M. Somers. 1991. The spectre of 'spurious' correlations. Oecologia
484 86:147-151.
485 Kelly, C., and T. D. Price. 2004. Comparative methods based on species mean values.
486 Mathematical Biosciences 187:135-154.
487 —. 2005. Correcting for regression to the mean in behavior and ecology. The American
488 Naturalist 166:700-707.
489 Kelly, M. 2019. Adaptation to climate change through genetic accommodation and assimilation
490 of plastic phenotypes. Philosophical Transactions of the Royal Society B 374:20180176.
491 Kingsolver, J. G., and L. B. Buckley. 2017. Evolution of plasticity and adaptive responses to
492 climate change along climate gradients. Proceedings of the Royal Society B: Biological
493 Sciences 284:20170386.
494 Lande, R. 2009. Adaptation to an extraordinary environment by evolution of phenotypic
495 plasticity and genetic assimilation. Journal of Evolutionary Biology 22:1435-1446.
496 Martin, C. H., J. E. Crawford, B. J. Turner, and L. H. Simons. 2016. Diabolical survival in Death
497 Valley: recent pupfish colonization, gene flow and genetic assimilation in the smallest
498 species range on earth. Proceedings of the Royal Society B: Biological Sciences
499 283:20152334.
500 Morley, S., L. Peck, J. Sunday, S. Heiser, and A. Bates. 2019. Physiological acclimation and
501 persistence of ectothermic species under extreme heat events. Global Ecology and
502 Biogeography 28:1018-1037.
503 Murren, C. J., J. R. Auld, H. Callahan, C. K. Ghalambor, C. A. Handelsman, M. A. Heskel, J.
504 Kingsolver et al. 2015. Constraints on the evolution of phenotypic plasticity: limits and
505 costs of phenotype and plasticity. Heredity 115:293-301.

O'Connor, C. M., D. R. Norris, G. T. Crossin, and S. J. Cooke. 2014. Biological carryover effects: linking common concepts and mechanisms in ecology and evolution. Ecosphere 5:1-11.

Pfennig, D. W., M. A. Wund, E. C. Snell-Rood, T. Cruickshank, C. D. Schlichting, and A. P. Moczek. 2010. Phenotypic plasticity's impacts on diversification and speciation. Trends in Ecology & Evolution 25:459-467.

Pigliucci, M., C. J. Murren, and C. D. Schlichting. 2006. Phenotypic plasticity and evolution by genetic assimilation. Journal of Experimental Biology 209:2362-2367.

Preston, J. C., R. Wooliver, H. Driscoll, A. Coughlin, and S. N. Sheth. 2021. Spatial variation in high temperature-regulated gene expression predicts evolution of plasticity with climate change in the scarlet monkeyflower. Molecular Ecology in press.

Price, T. D., A. Qvarnström, and D. E. Irwin. 2003. The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:1433-1440.

R Development Core Team. 2021. R: a Language and Environment for Statistical Computing.The R Foundation for Statistical Computing, Vienna.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217-223.

Riddell, E. A., J. P. Odom, J. D. Damm, and M. W. Sears. 2018. Plasticity reveals hidden resistance to extinction under climate change in the global hotspot of salamander diversity. Science Advances 4:eaar5471.

Rohr, J. R., D. J. Civitello, J. M. Cohen, E. A. Roznik, B. Sinervo, and A. I. Dell. 2018. The complex drivers of thermal acclimation and breadth in ectotherms. Ecology Letters 21:1425-1439.

Sasaki, M. C., and H. G. Dam. 2019. Integrating patterns of thermal tolerance and phenotypic plasticity with population genetics to improve understanding of vulnerability to warming in a widespread copepod. Global Change Biology 25:4147-4164.

—. 2021. Negative relationship between thermal tolerance and plasticity in tolerance emerges during experimental evolution in a widespread marine invertebrate. Evolutionary Applications 14:2114-2123.

Schlichting, C. D., and M. Pigliucci. 1998, Phenotypic evolution: a reaction norm perspective, Sinauer Associates Incorporated.

Schneider, R. F., and A. Meyer. 2017. How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. Molecular Ecology 26:330-350.

Seebacher, F., C. R. White, and C. E. Franklin. 2015. Physiological plasticity increases resilience of ectothermic animals to climate change. Nature Climate Change 5:61-66.

Sikkink, K. L., R. M. Reynolds, C. M. Ituarte, W. A. Cresko, and P. C. Phillips. 2014. Rapid evolution of phenotypic plasticity and shifting thresholds of genetic assimilation in the nematode *Caenorhabditis remanei*. G3: Genes, Genomes, Genetics 4:1103-1112.

Somero, G. 2010. The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine ‘winners’ and ‘losers’. The Journal of Experimental Biology 213:912-920.

Stillman, J. H. 2003. Acclimation capacity underlies susceptibility to climate change. Science 301:65-65.

549 van Heerwaarden, B., and V. Kellermann. 2020. Does plasticity trade off with basal heat
550 tolerance? Trends in Ecology & Evolution 35:874-885.
551 Via, S., and R. Lande. 1985. Genotype-environment interaction and the evolution of phenotypic
552 plasticity. Evolution 39:505-522.
553 Waddington, C. H. 1953. Genetic assimilation of an acquired character. Evolution 7:118-126.
554 West-Eberhard, M. J. 2003, Developmental plasticity and evolution, Oxford University Press.
555