Adaptive divergence and the evolution of hybrid phenotypes in threespine stickleback

November 26, 2020

Avneet K. Chhina*†, Ken A. Thompson*‡, Dolph Schluter

Department of Zoology & Biodiversity Research Centre, University of British Columbia, 2329 West Mall, Vancouver, BC V6T 1Z4

*equal contribution †chhinaavneet@gmail.com ‡kthomp1063@gmail.com

Keywords: adaptive divergence, dominance, hybridization, speciation, phenotypic mismatch

Abstract

The fitness of hybrids is a critical determinant of gene flow between hybridizing populations. If hybrid phenotypes change predictably as parental populations become increasingly divergent, this could provide insight into general mechanisms linking ecological divergence with reproductive isolation. In this study, we used threespine stickleback fish (*Gasterosteus aculeatus* L.) to examine how phenotypic divergence between populations drives the evolution of dominance, phenotypic variation, and trait 'mismatch' in hybrids. We generated F_1 and F_2 hybrids between 12 freshwater populations—which ranged from highly planktivorous to highly benthic-feeding—and an anadromous population that is highly planktivorous and resembles the ancestral state of derived freshwater populations. We measured 16 phenotypic traits in hybrids and pure parental individuals raised under common conditions. We found that dominance varied markedly among traits. By contrast, dominance for a given trait was typically consistent among populations except for two traits where dominance was predicted by the phenotype of the freshwater parent. We find that multivariate phenotypic variation is greater in hybrids between more divergent parents. Finally, we demonstrate that the extent to which parental traits are 'mismatched' in both F_1 and F_2 hybrids increases with the phenotypic distance between the parent populations. Critically, this relationship was clearer in F_1 hybrids than in F_2s—largely due to traits having different dominance coefficients and F_1s having relatively little phenotypic variation. Our results demonstrate that some aspects of hybrid phenotypes evolve predictably as parental populations diverge. We also find evidence for a possible general mechanistic link between ecological divergence and reproductive isolation—that more divergent parent populations tend to produce hybrids with novel and potentially deleterious multivariate phenotypes.

One of the central tenets of the 'ecological speciation' hypothesis is that adaptive divergence is causally linked in some way to reproductive isolation (Nosil 2012; Schluter 2000). Synthetic studies have found support for this link (Shafer and Wolf 2013), but the general mechanisms that might underlie this relationship are largely unknown. One critical determinant of reproductive isolation is the fitness of hybrids because introgression cannot occur if hybrids are removed from the population before they can back-cross with parents (Coyne and Orr 2004). Theory predicts that several aspects of hybrid morphology are linked to the magnitude of phenotypic divergence between parents. For example, the amount of phenotypic variation in characters under both divergent and stabilizing selection increases as populations diverge phenotypically (Barton 2001; Chevin et al. 2014; Slatkin and Lande 1994), although empirical evidence for this phenomenon is scarce (though see Thompson 2020). Experimentally quantifying how adaptive divergence affects the phenotypes of hybrids formed between divergent natural populations therefore has the potential to generate hypotheses about general mechanisms that might link ecological divergence to reproductive isolation (Edmands 1999).

One possible consequence of phenotypic variation in hybrids is that divergent traits might be expressed in 'mismatched' combinations, meaning that hybrids resemble one parent more for some traits and the other parent for other traits. Phenotypic mismatch has been linked to lower fitness in several systems. In threespine stickleback fish, (*Gasterosteus aculeatus* L.), individual F_2 hybrids that had mismatched jaw traits performed worse than all other individuals (Arnegard et al. 2014). In sunflowers (*Helianthus* spp.), individual backcross hybrids with a greater extent of mismatch across multiple pairs of traits had lower fitness than less mismatched individuals (Thompson et al. 2019). In these recombinant hybrids, mismatch readily results from the segregation of divergent alleles. Mismatch can manifest in F_1 hybrids as well if dominance acts in opposing directions for different traits (i.e., trait 1 is dominant toward parent 1 but trait 2 is dominant toward parent 2) (Matsubayashi et al. 2010; Thompson et al. 2019; Vinšálková and Gvoždík 2007). If hybrid trait mismatch were to increase with the degree of phenotypic divergence between parents, this might have the potential to directly link divergence to reproductive

isolation.

We hypothesized that mismatch in hybrids would be associated with phenotypic divergence between parental populations. To test our central prediction—that adaptive divergence (between parents) and phenotypic mismatch (in hybrids) would be positively correlated—we leveraged the unique biology of the threespine stickleback fish. Freshwater populations of stickleback span a continuum of trait-space along a primarily limnetic (i.e., planktivorous; more ancestral or marine-like) to benthic (i.e., consuming large macro-invertebrates in shallow water; more derived) axis (Bell and Foster 1994). These derived freshwater populations are recently (approx. 10 kya) and independently derived from the colonizing anadromous ancestors that remain abundant today and are readily crossed with derived forms.

The goal of this article is to test the hypothesis that trait mismatch observed in hybrids is associated with the magnitude of phenotypic divergence between parents. We examine the evolution of dominance, phenotypic variation, and mismatch in turn. We first ask whether dominance differs among traits and populations. Next, we compare patterns of trait variation among traits and populations. Finally, we compare patterns of trait mismatch—which results from dominance and trait variation—across populations. Our results hint at a possible general mechanistic basis for the evolution of hybrid fitness during ecological speciation.

Methods

Study system

The threespine stickleback is a teleost fish species distributed throughout the coastal areas of the northern hemisphere (Bell and Foster 1994). Sticklebacks are ideal for studying the consequences of adaptive divergence on phenotypic traits in hybrids because anadromous populations colonized an array of post-glacial lakes and have rapidly adapted to prevailing ecological conditions (Schluter 1996). Stickleback that live in large lakes with predators and other competitor fish species (e.g., prickly sculpin) have adapted to freshwater but largely retained the ancestral plank-

tivorous habit (Miller et al. 2019). By contrast, populations that have evolved in small pond-like lakes with few or no predators and competitors often have more derived phenotypes specialized for foraging on large benthic invertebrates.

Because this divergence occurred so rapidly, populations can be readily crossed with little to no evidence for inviability caused by 'intrinsic' incompatibilities (Hatfield and Schluter 1999; Lackey and Boughman 2016; Rogers et al. 2012). Extant anadromous populations, within a particular geographic location, are phenotypically similar to the ancestral populations that founded present-day freshwater populations (Morris et al. 2018). We leveraged this continuum of phenotypic variation using crosses to test our hypotheses about the effects of adaptive divergence on hybrid phenotypes.

Fish collection and husbandry

Wild fish were collected in British Columbia, Canada, in April–June of 2017 and 2018. We sampled twelve freshwater populations from nine lakes (Fig. 1; three lakes contain reproductively isolated benthic-limnetic 'species pairs' [McPhail 2008] and thus contributed two populations each). The anadromous population was collected from the Little Campbell River (Fig. 1). Fish were caught using minnow traps or dip nets. We crossed six gravid anadromous females with six males from each freshwater population to generate six unique F_1 hybrid families, and also generated four to six non-hybrid (i.e., 'pure' species) families for each parental population. All offspring were raised in the lab under common conditions (see **Supplementary Methods**). Because all hybrid crosses were made with the anadromous female (with the exception of one F_1 family that used a freshwater Bullock Lake dam but did not differ from others and was not used to make F_2s), we cannot rule out cross-direction specific patterns. The decision to make crosses in only one direction was made because obtaining wild gravid females for some populations is prohibitively difficult. This choice, however, does not limit our ability to compare hybrid phenotypes across populations.

When fish reached reproductive maturity, F_1 hybrids from unrelated families were crossed to

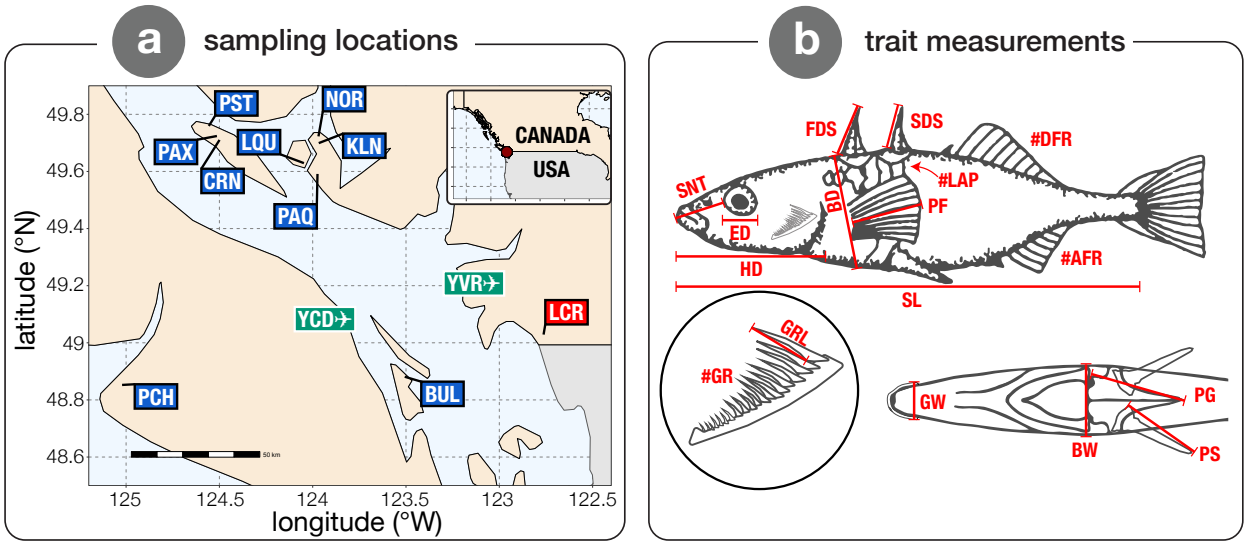


Figure 1: **Overview of sampling locations and trait measurements.** Panel (a) shows locations where the parents of our crosses were collected in British Columbia, Canada. Boxes show collection locations of the anadromous population (red box; LCR—Little Campbell River), freshwater populations (blue boxes; left to right: PCH—Pachena Lake; PAX—Paxton Lake; CRN—Cranby Lake; PST—Priest Lake; LQU—Little Quarry Lake; PAQ—Paq (Lily) Lake; NOR—North Lake; KLN—Klein Lake; BUL—Bullock Lake). Green labels give airport codes for major cities (YCD—Nanaimo; YVR—Vancouver). Panel (b) shows the measurements of all 16 traits in the dataset and standard length. The upper section of the panel shows the lateral view (traits left to right: SNT—snout length; ED—eye diameter; HD—head length; FDS—length of first dorsal spine; BD—body depth; SL—standard length; SDS—length of second dorsal spine; PF—pectoral fin length; #LAP—number of lateral armour plates; #DFR—number of dorsal fin rays; #AFR—number of anal fin rays). The bottom left section of the panel shows a zoomed in drawing of the upper arm of the outer gill raker arch (#GR—number of gill rakers; GRL—length of longest gill raker). The lower right section shows a ventral view of the anterior portion of the body (GW—gape width; BW—body width; PG—length of pelvic girdle; PS—length of pelvic spine). The upper drawing was originally published by Bell and Foster (1994) and re-used with permission from M. Bell.

make three F_2 families within each cross population (with the exception of Paxton Lake benthics which, due to space constraints in 2018, had only two F_2 families from the same two F_1 parent families).

Fish from each family were lethally sampled when individuals in the tank reached a mean standard length of approximately 40 mm. Fish had not reached reproductive maturity at the time of sampling and we therefore could not quantify sex. Fish were preserved in formalin, stained with alizarin red, and then stored permanently in 40% isopropyl alcohol. For F_1s, tanks were sub-sampled and remaining individuals were raised to produce F_2s. For F_2s, entire tanks were lethally sampled.

Phenotype measurements

We measured 16 traits and standard length on stained fish. For all traits, we measured at least 100 pure anadromous parents, 30 pure freshwater parents, 30 F_1 hybrids, and 60 F_2 hybrids (all lab-raised; see summary dataset [**to be archived on Dryad**] for trait means, standard deviations (SDs), and sample sizes for all populations). We used a dissecting microscope to count the number of dorsal fin rays, anal fin rays, lateral armour plates, and gill rakers. We also measured the length of the longest gill raker using an ocular micrometer. We photographed the left and ventral sides of each fish with a Nikon D300 camera and used ImageJ (Abramoff et al. 2004) to make linear measurements of body dimensions and bones (see Fig.1 for more details on measurements).

Data analysis

We evaluated whether adaptive divergence between parent populations was associated with the phenotype observed in hybrids. We specifically investigated patterns of trait dominance, variation, and mismatch.

All data processing and model-fitting was done using R (R Core Team 2019). Data processing and plotting was greatly aided by the packages contained within the `tidyverse` (Wickham 2017).

Mixed models were fit using `lme4` (Bates et al. 2014) and analysed using `lmerTest` (Kuznetsova et al. 2014 with the Kenward-Roger approximation for the denominator degrees of freedom (Kenward and Roger 1997). The the 'map' function in `purrr` (Henry and Wickham 2019), and associated functions in `broom` (Robinson et al. 2020), were used to streamline code for iterating models over variables. The `ggpubr` (Kassambara 2020) and `egg` (Auguie 2019) packages were used to create, customize, and annotate graphs. Partial residuals were plotted using `visreg` (Breheny and Burchett 2017). `for` loop code was streamlined with the functions in `magicfor` Makiyama 2016). We used the `emmeans` package (Lenth et al. 2020) and the 'cld' function in `multcomp` (Hothorn et al. 2008) to assist with post-hoc comparisons. The 'r2beta' function in `r2glmm` (Jaeger 2017) to calculate the partial r^2 coefficient for for the fixed effects following the method of Nakagawa and Schielzeth (2013). The functions in the 'correlation' package (Makowski et al. 2019) improved code for tidy correlation matrices.

We size-corrected all linear measurements by calculating residuals from a log-log linear regression model on standard length. Some measurements are affected if fish are fixed with an expanded buccal cavity, so we also corrected for fixation position by assigning all fish a number (0, 1, or 2) depending on the extent to which the mouth was open and then performing a further correction as above using residuals for gape width, snout length, and head length. Linear measurement traits that had no measurement values—which is possible when features such as the first dorsal spine or pelvic girdle are absent—were given a value of 0 mm.

Dominance

When measuring dominance of traits in hybrids we restrict our analysis to only traits and populations where the parents have different trait values. We did this because calculating dominance for traits that do not differ can result in extreme values of dominance due to sampling error. Specifically, we retained traits where the freshwater parent populations were statistically different (t-test $P < 0.05$) and/or ≥ 1 SD apart (in units of anadromous SDs; both statistics and SDs used to be as inclusive as possible). We primarily generate inferences about dominance from F_1

hybrids, but also calculate metrics in F_2s as a point of comparison.

We measure dominance on a scale where values are expected (though not restricted) to fall between 0 and 1. Values of 0 indicate that hybrid trait values are the same as the anadromous parent (P_{anad}; i.e., ancestral trait is dominant), values of 1 indicate that trait values are the same as the freshwater parent (P_{fresh}; i.e., derived trait is dominant),and values of 0.5 indicate the trait is additive. On this scale, transgressive trait values have values < 0 or > 1. We calculate our dominance metric (D) for a given trait in a given individual (F_1) hybrid as:

$$\mathrm{D} = \frac{\mathrm{F}_n - \mu_{\mathrm{P}_{\mathrm{anad}}}}{\mu_{\mathrm{P}_{\mathrm{fresh}}} - \mu_{\mathrm{P}_{\mathrm{anad}}}}, \tag{1}$$

where F_n is individual hybrid's (either F_1 or F_2) trait value, and $\mu_{P_{fresh}}$ and $\mu_{P_{anad}}$ represent the trait means of the freshwater and common anadromous parental populations, respectively. The pooled means reflect the mean of families within a given cross type.

To test whether dominance differed among traits and/or populations, we fit linear mixed models with 'D' values (eqn. 1) of individual fish as the response variable and family was a random effect. In models testing whether dominance coefficients varied among traits, 'trait' was the fixed effect. In models testing whether dominance differed among populations for a given trait, 'population' was the fixed effect.

Because dominance varied among populations for some traits (see **Results**), we wished to test if dominance evolves predictably with the magnitude of phenotypic divergence between each derived population and their common ancestor. To do this, we calculated the phenotypic distance between cross parents for each trait in units of anadromous phenotypic standard deviations as:

$$d(\mathrm{P}_{\mathrm{anad}}, \mathrm{P}_{\mathrm{fresh}}) = \frac{|\mu_{\mathrm{P}_{\mathrm{anad}}} - \mu_{\mathrm{P}_{\mathrm{fresh}}}|}{\sigma_{\mathrm{P}_{\mathrm{anad}}}}, \tag{2}$$

where $\mu_{P_{anad}}$ and $\mu_{P_{fresh}}$ are the pooled means for a given trait the anadromous within a given freshwater population, respectively, and $\sigma_{P_{anad}}$ is the anadromous phenotypic SD for the trait. This was then regressed against individual dominance values (eqn. 1) for a given trait to test

the hypothesis that dominance evolves in association with the magnitude of divergence between parents.

Phenotypic variation

We next examined patterns of phenotypic variation observed in hybrids. We first determined how many populations exhibited appreciable segregation variance for each trait. To do this, we calculated segregation variance (σ_s^2) for each trait in each population as:

$$\sigma_s^2 = \frac{4\sigma_{F_2}^2}{2\sigma_{F_1}^2 + \sigma_{P_{anad}}^2 + \sigma_{P_{fresh}}^2}, \tag{3}$$

where σ^2 represents the phenotypic variance and all other notation is as above (Wright 1968). We generated 1000 bootstrap estimates of segregation variance for each trait in each population to produce confidence intervals and determine if it was significantly greater than the null expectation of 1. Re-sampling was done by drawing individuals from families with the same sample size as in the original families, such that the number and size of families was the same in the original dataset as in the bootstrap replicates. For these analyses, we accounted for possible differences in trait means among families by taking each individual's trait value as the difference between it's observed value and the family mean. Bootstrap replicates were constrained to have equal numbers of individuals from each family. We used Spearman's rank-order correlations to evaluate if there was a significant relationship between segregation variance (eqn. 3) and parent trait divergence (eqn. 2).

Going beyond a single trait approach, we next investigated whether hybrid phenotypic variation in multivariate trait space is predicted by the phenotypic distance between parents. To do this, we consider all traits at once with all traits scaled to a mean of 0 and an SD of 1. We use this approach rather than dimensionality-reduction (e.g., principal components analysis) because genetic correlations in our data are low (median r = 0.20; 12.7 % of pairwise correlations significant at P = 0.05; Fig. S1). In multivariate space, parent trait divergence was calculated as

the Euclidean distance between the pooled mean phenotypes of the freshwater and anadromous parents.

We calculated each hybrid individual's mean Euclidean phenotypic distance to all other hybrids within its generation (F_1 or F_2) and population. We analyzed the data with mixed models that had each individual hybrid's mean distance as the response. Fixed effects were were (i) hybrid trait category (F_1 or F_2); (ii) the Euclidean distance between parent mean trait values; the interaction between (i) and (ii); and (iii) multivariate parental variation (distance among freshwater parent individuals). Family was a random effect.

Trait 'mismatch'

Our final line of investigation concerned the degree to which hybrids had 'mismatched' combinations of traits and whether this is associated with the degree of phenotypic divergence between parents. We used two approaches for this, one considering all traits at once in multivariate space and the other considering pairs of traits at a time. The former has the advantage of considering high-dimensional trait space and the latter has the advantage of having a more intuitive biological interpretation.

Mismatch is the Euclidean distance between a hybrid's phenotype and the line that connects the two parental mean phenotypes (see Fig. 4A & 4B for visual overview). This is calculated as:

$$d_{\text{mismatch}} = (\vec{\mathrm{F}}_n - \vec{\mathrm{P}}_{\text{anad}}) - (\vec{\mathrm{P}}_{\text{fresh}} - \vec{\mathrm{P}}_{\text{anad}}) \times \frac{(\vec{\mathrm{F}}_n - \vec{\mathrm{P}}_{\text{anad}}) \cdot (\vec{\mathrm{P}}_{\text{fresh}} - \vec{\mathrm{P}}_{\text{anad}})}{\vec{\mathrm{P}}_{\text{fresh}} - \vec{\mathrm{P}}_{\text{anad}}^{\,2}}, \tag{4}$$

where $\vec{\mathrm{F}}_n$, $\vec{\mathrm{P}}_{\text{anad}}$, and $\vec{\mathrm{P}}_{\text{fresh}}$ are the vectors of individual hybrid, pooled mean anadromous, and pooled mean freshwater (scaled) trait values, respectively.

The goal of the mismatch analysis was to test the hypothesis that mismatch is more substantial in hybrids formed between more divergent parents. Because freshwater populations are phenotypically variable, hybrids might appear mismatched even if they have similar variation

to the freshwater parent. We therefore accounted for this by calculating the 'mismatch' for each freshwater parental population (eqn. 4). We fit mixed models with mismatch (either multivariate or pairwise) as the response, and $d_{(\mathrm{P_{anad}}, \mathrm{P_{fresh}})}$ (i.e., multivariate Euclidean distance between the parental populations), hybrid category (F_1 or F_2), and their two-way interaction as fixed effects. Parent 'mismatch' was also a fixed effect, and family was a random effect.

For two-trait mismatch metrics, regressions were looped across all trait pairs. We evaluated the statistical significance of the regression models, as well as the distribution of regression coefficients across models to generate inferences about the relationship between adaptive divergence between parents and mismatch in hybrids.

Results

Evolution of dominance in hybrids

Dominance, as observed in hybrids, varied among traits and populations. In both F_1 and F_2 hybrids, we found that dominance coefficients differed widely among traits, with most traits being dominant toward the anadromous ancestor (Fig. 2A). The only trait that was consistently transgressive was body width, where hybrids were often more slender than the streamlined anadromous parent. Of 16 traits, only four exhibited variation in dominance among freshwater populations (Fig. S2). Dominance in two traits showed consistent associations with the amount of phenotypic divergence between parents for that trait. For lateral armour plates, the freshwater phenotype is increasingly dominant in hybrid offspring of more derived (i.e., fewer plates) parents (Fig. 2B). By contrast, dominance of pelvic spine length is less dominant in hybrid offspring of more derived (i.e., shorter spines) parents (Fig. 2C).

Evolution of hybrid phenotypic variation

Phenotypic variation was positively associated with the magnitude of divergence in hybrids. Considering traits individually, only lateral armour plates exhibited significant segregation vari-

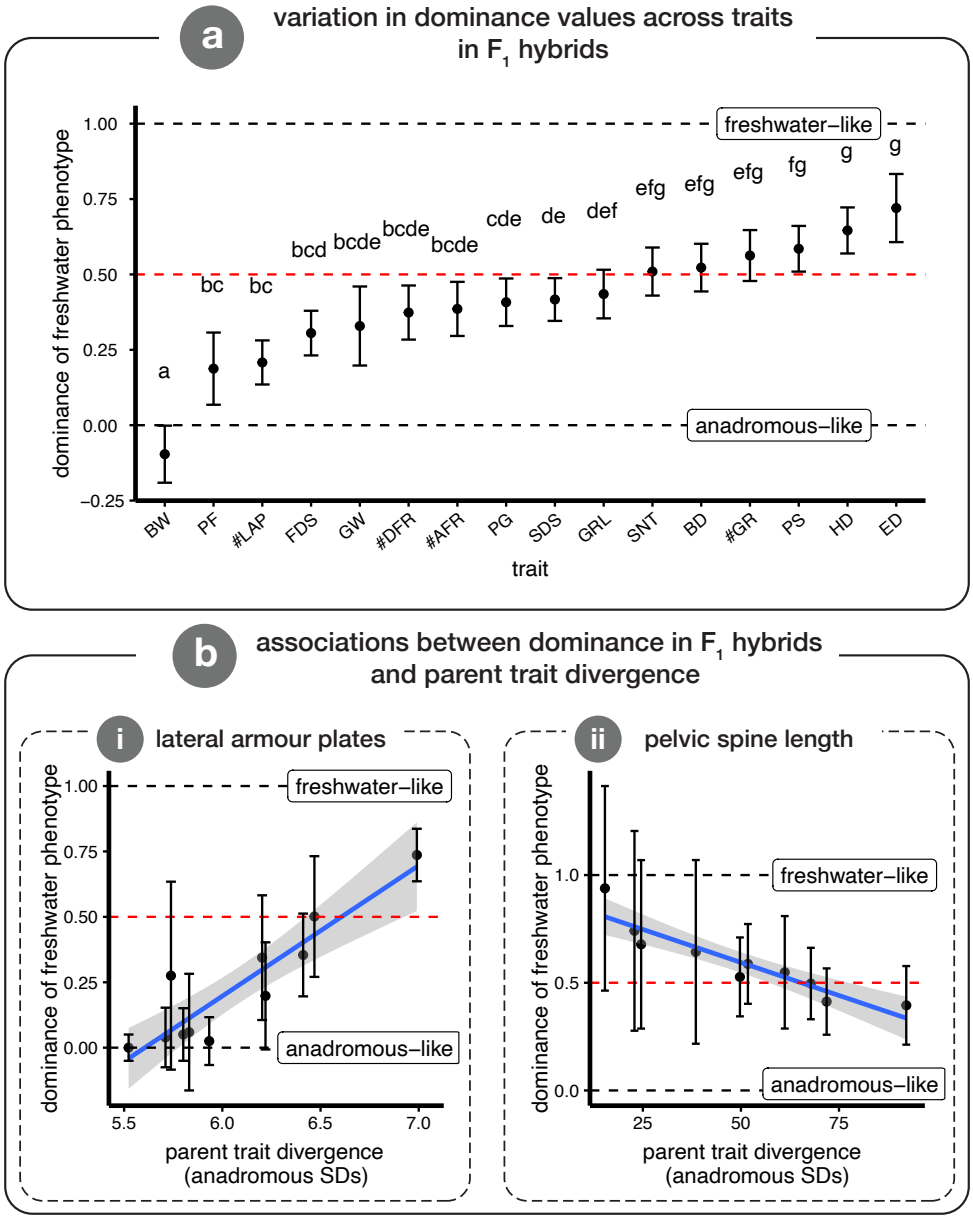


Figure 2: **Dominance, as observed in F_1 hybrids, differs among traits and populations.** For all panels, the dashed lines at 0 and 1 represent the ancestral anadromous parent and derived freshwater parent trait values, respectively, and the red dashed line at 0.5 represents the expected value with no dominance. Panel (a) depicts the estimated marginal mean ($\pm$ 95% CI) dominance coefficient—calculated across all populations—for all measured traits (N = 862 fish including F_1 hybrids and parents). Letters indicate traits that differ at $P < 0.05$ after correcting for multiple comparisons. Panel (b) depicts the relationship between dominance and parental trait divergence (in units of anadromous SDs; eqn. 2) for (i) lateral armour plates ($\beta = 0.50$; $F_{1,9} = 47.19$; $r^2 = 0.84$; $P = 7.31 \times 10^{-5}$) and (ii) pelvic spine length ($\beta = -6.13 \times 10^{-3}$; $F_{1,8} = 47.25$; $r^2 = 0.85$; $P = 1.28 \times 10^{-4}$).

ance (i.e., bootstrap CI excluding 1) for all populations (Table 1). All other traits had significant segregation variance in fewer than half of studied populations. Moving beyond a univariate approach, we analysed hybrid phenotypic variation in multivariate trait space and tested whether it increased with the phenotypic distance between parents. In support of our prediction, we found that hybrids show increasing multivariate phenotypic variation (i.e., increasing distance among individuals) in more phenotypically divergent crosses (Fig. 3. The category (i.e., F_1 or F_2) $\times$ parent phenotypic divergence interaction term was marginally significant (P = 0.081), so we cannot reject the null hypothesis that both F_1 and F_2 hybrids have increasing variation with parent divergence.

Trait mismatch

Mismatch in hybrids was generally positively associated with phenotypic divergence between parents. We first investigated the relationship between hybrid mismatch and parent phenotypic divergence in multivariate trait space. We found that there is a statistically significant positive relationship between multivariate trait mismatch and parent phenotypic divergence (Fig. 4C).

We next examined the relationship between adaptive divergence and hybrid mismatch for pairs of traits at a time. Because trait pairs are not all independent, we view this analysis as an intuitive heuristic meant primarily to complement the analysis considering all traits at a time. We found that pairwise trait mismatch was significantly associated with parent trait divergence for 22 of 120 trait pairs in F_1s (18.3 %) and 10 of 120 trait pairs in F_2s (8.3 %; Fig. 4D). Slopes of significant relationships were positive in all cases except for one trait pair in F_1s. To see an example of pairwise mismatch for individual hybrids in 2D trait space with real data, see Fig. S3).

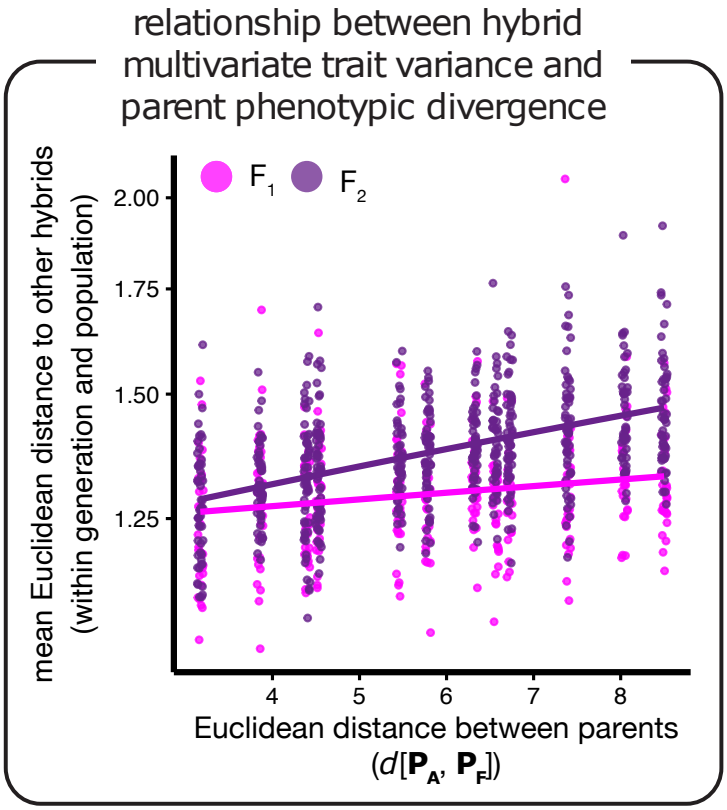


Figure 3: **Hybrid phenotypic variation increases with the magnitude of phenotypic divergence between parents.** Points represent individual hybrids' mean distance to all other hybrids within the same category (i.e., F_1 or F_2) and population. Partial residuals are plotted controlling for parent phenotypic variation. The main effect of parent phenotypic distance is positive and statistically significant (β = 0.013; $F_{1,76.38}$ = 6.16; P = 1.28 $\times$ 10^{-4}; F_1 hybrid r^2 = 0.034; F_2 hybrid r^2 = 0.045). Note the $\log^{10}$ scale on the y-axis. Points are slightly jittered to ease visualization.

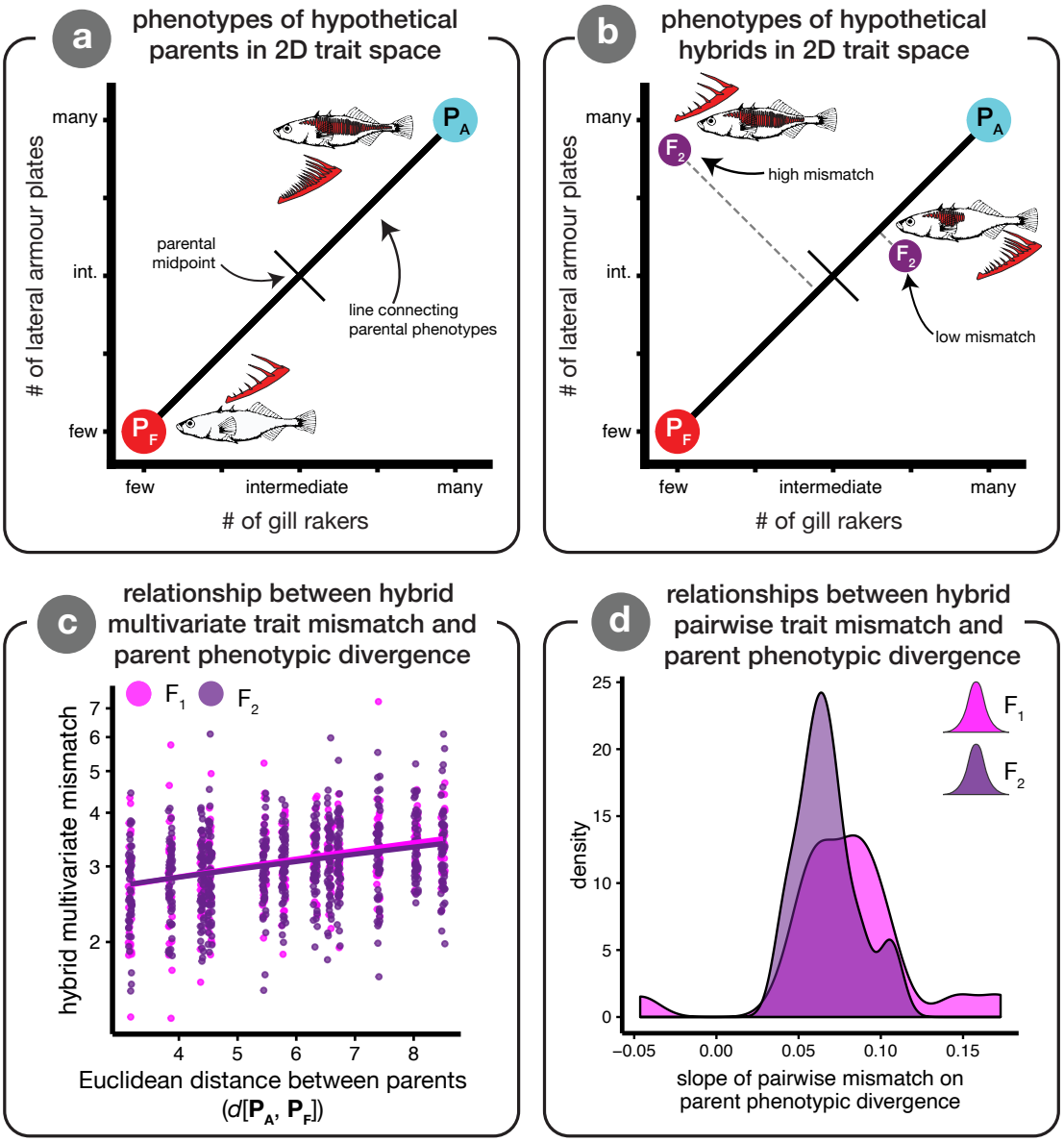


Figure 4: **Visual overview of mismatch calculation and observed mismatch results.** Panel (a) shows hypothestical parent phenotypes for lateral armour plates and gill raker number in 2D trait space. Panel (b) shows two hypothetical hybrids in the same trait space with low or high values of mismatch (calculated as the length of the dashed lines). Panel (c) shows the statistically significant relationship between hybrid multivariate trait mismatch and parent phenotypic divergence at in multivariate trait space for both F_1 and F_2 hybrids (β = 0.13, $F_{1,70.9}$ = 10.4, P = 0.0019; F_1 partial r^2 = 0.098, F_2 partial r_2 = 0.011). Points are partial residuals from the model, plotted using and slightly jittered horizontally to aide visualization. Panel (d) is a density plot depicting the distribution of slopes for the pairwise mismatch analyses (restricted to statistically significant relationships only; N trait pairs: 22 [F_1] & 10 [F_2]).

Discussion

In this article, we used experimental hybridization to investigate how hybrid phenotypes evolve over the course of adaptive divergence. We were motivated by the fact that, although studies have documented a seemingly general relationship between ecological divergence and reproductive isolation (Funk et al. 2006; Shafer and Wolf 2013), no general mechanistic foundations for this relationship have been uncovered. Our results suggest that, at least in stickleback, more divergent parental populations tend to have hybrids with increasingly novel and mismatched multivariate phenotypes. Such a pattern might be partially responsible for the observed reduction in gene flow associated with adaptive ecological divergence. Surprisingly, this relationship was clearer in F_1 hybrids than in F_2s, which has implications for the importance of this relationship for speciation.

Variation in dominance

We observed variation in dominance among traits, and among populations for a given trait. Much of the observed dominance might have to do with the history of adaptation in the stickleback system, where alleles that allow populations to adapt to freshwater are maintained in the sea for years and are 'transported' back to freshwater (Schluter and Conte 2009). The observed dominance of a trait in the present is therefore the outcome of a tug-of-war between selection for the recessivity of freshwater-adapted alleles in the sea (i.e., selection to *not* express the freshwater phenotype in heterozygotes) and selection for dominance of freshwater-adapted alleles inland (i.e., selection *to* express the freshwater phenotype in heterozygotes). The distribution of dominance coefficients among traits might indicate that selection on dominance in the sea tends to be the stronger of the two forces, although it is uneven among traits. This pattern might be expected from earlier quantitative trait locus mapping studies in marine-freshwater stickleback crosses that have found that QTL are slightly more dominant in the direction of the marine ancestor (Miller et al. 2014). Thus, as Haldane's sieve favours dominant alleles in freshwater, the

'Transporter sieve' favouring recessive ones is slightly more influential in this system.

The predictable association we observed between dominance and the freshwater parent's trait value hints at a possible adaptive explanation. For lateral armour plates—where anadromous individuals typically have > 30 plates and the most derived freshwater individuals have none—we observed that the freshwater phenotype was increasingly dominant in more phenotypically derived populations. Such a pattern could either arise due to selection in the sea or selection in freshwater. If the alleles underlying plate reduction were different across freshwater populations, those reducing it the most would be most deleterious in the sea and would experience countervailing selection for modifiers. However, since plate reduction is known to be largely caused by a single large-effect variant at the *Eda* locus (Archambeault et al. 2020; Colosimo et al. 2005), and we observe large shifts in plate number due to dominance, this is consistent with additional modification of *Eda*'s expression. Indeed, other studies looking specifically at the fitness differences among *Eda* genotypes have found that the dominance patterns differ among populations (Schluter et al. 2020). In freshwater, dominance modifiers favouring the expression the freshwater phenotype might be most strongly selected for in the most derived populations because it is probably in those populations where expression of the ancestral phenotype is most deleterious. By contrast with armour, pelvic spine length was more *recessive* in derived populations than in more ancestral ones. Because variation in this trait is more continuous than lateral armour plates, it is possible that selection in the sea modifies pelvis-reduction alleles with large effects more than those with small effects.

Explanations for the evolution of phenotypic variation

The only trait for which we found universal segregation variance was lateral armour plates. This is not surprising given that the phenotype is largely governed by a single large-effect locus (*Eda*) and that appreciable segregation variance for this trait has been shown many times before in marine-freshwater crosses. For example, Schluter et al. (2004) crossed marine and stream fish and found significant segregation variance for armour plates. The same cross, however, did

not show segregation variance for body shape, and the authors suggest this is due to the high polygenicity of this trait (Schluter et al. 2004). Other studies measuring the same traits as we did in crosses between low-plated stickleback populations found that plate number is a large outlier in variance compared to traits such as gill raker number or length (Hatfield and Schluter 1999).

We found that multivariate phenotypic variation in hybrids increased with the magnitude of divergence between cross parents. There was marginal evidence that this relationship was stronger in F_2s than in F_1s. Previous studies in stickleback have found that the average effect size of alleles used for adaptation to freshwater is larger in more phenotypically derived populations (Rogers et al. 2012), which might partially underlie this relationship. Other studies in stickleback have found that many alleles used for freshwater adaptation are also pleiotropic (Albert et al. 2008). It is possible that this pleiotropy obscures the patterns when traits are considered one at a time, and why patterns are most evident in multivariate trait space.

Mismatch findings and their implication for speciation

The key result of this study is that hybrids between phenotypically divergent parents tend to be more mismatched than hybrids from phenotypically similar parents. The association between parental phenotypic divergence and mismatch is not perfect (i.e., a rank-order correlation is not 1), and parent divergence explains a fairly low amount of the variation (all partial $r^2 < 0.12$). Previous studies comparing gene flow between pairs of natural populations, or fitness among crosses spanning a continuum of 'ecological' (broadly defined) divergence, have typically found imperfect relationships as well (Edmands 1999; Funk et al. 2006; Shafer and Wolf 2013). Therefore, we emphasize that the relationships we documented have low explanatory power, which is perhaps unsurprising in light of these previous studies.

The sign of dominance coefficients, as measured in hybrids, varied markedly among traits and was more often than not consistent across populations. This ultimately results in hybrids having some trait values that would be well-suited to life in the sea, and others that are better suited to life in lakes. In British Columbia, adaptation to marine life in anadromous stickleback

populations results in a streamlined (i.e., narrow) body with large pectoral fins and a small head compared to freshwater populations (Dalziel et al. 2012; Jones et al. 2006). There might be no possible environment in which the trait combinations that we find to typically characterize F_1 hybrids—for example a streamlined body and a large head—are optimal. Such trait mismatches could readily contribute to the extrinsic selection against hybrids known to occur in anadromous-freshwater stickleback hybrid zones (Jones et al. 2006; Vines et al. 2016), including the Little Campbell River (Hagen 1967) studied here.

We found that pairwise mismatch increased with parent divergence in both F_1 and F_2 hybrids. If all QTL are additive, one expects the pattern of increasing mismatch with phenotypic divergence to hold only for the F_2 generation. This is because F_1s will be intermediate always and more (and larger) QTL will segregate in F_2s from 'wider' crosses (Barton 2001; Chevin et al. 2014; Slatkin and Lande 1994). While we did observe the predicted pattern in F_2s, the same pattern not only held but was in fact clearer (i.e., higher r^2) in F_1s. In light of our finding that dominance is inconsistent among traits and (largely) consistent among populations, this result is easily understood and predictable. Specifically, because mismatch in F_1s is caused by unbalanced dominance among traits, a given amount of dominance imbalance generates a magnitude of mismatch that is positively associated with the phenotypic divergence between parents. If a hybrid is 25 % more similar to one parent for trait x, and 25 % more similar to the *other* parent for trait y, that hybrid will have a greater magnitude of mismatch if the parents have a greater magnitude of divergence. Because F_1s are less phenotypically variable than F_2s, this causes all F_1s to be similarly mismatched within a cross thus causing the relatively high explanatory power. If mismatch reduces gene flow, it can act immediately after hybridization instead of only acting in recombinant hybrids (Matsubayashi et al. 2010). To further explore the consequences of this result, it would be useful for theoretical work to evaluate the relative strength of reproductive barriers that act in the F_1 and recombinant hybrids to those that only act in recombinant hybrids, and also to examine how variation in barrier strength among individual hybrids affects reproductive isolation.

A limitation of our study is that we can only indirectly link our findings to fitness and reproductive isolation. In stickleback, Arnegard et al. (2014) identified a fitness landscape that implicated mismatch between two traits—jaw protrusion and jaw-opening inlever—as being responsible for a significant reduction in growth rate in the field. Thompson et al. (2021) illustrated that individual-level variation in pairwise mismatch in sunflowers was associated with fitness where more mismatched individuals had lower fitness than less mismatched individuals. It would be useful for future studies to conduct comparative experimental studies where mismatch can be linked to fitness in several crosses spanning a continuum of divergence.

Acknowledgements

We are grateful to many Schluter lab members & volunteers and kind BC residents, for assistance in the lab and field. The manuscript was aided by feedback from members of Dolph Schluter's group at The University of British Columbia and Molly Schumer's group at Stanford University. Anadromous stickleback were collected on the lands of the Semiahmoo First Nation, and we are grateful to J. Charles and J. Cook for allowing us to sample on their land. Freshwater populations were collected on the traditional territory of the Huu-ay-aht and Sechelt people. Funding was provided by an NSERC Discovery Grant to DS, an NSERC CGS-D and UBC Four-Year Fellowship to KAT, and an NSERC USRA to AKC.

Data availability

All data and analysis code will be made available to reviewers following submission to a journal and archived permanently at time of acceptance.

Author contributions

KAT and DS conceived of the study. KAT conducted fieldwork, made crosses, raised animals, and collected samples. KAT and AKC collected data and co-wrote the first draft of the manuscript. AKC and KAT analyzed the data with input from DS. All authors revised the manuscript.

References

Abramoff, M. D., P. J. Magalhaes, and S. J. Ram. 2004. Image processing with ImageJ. Biophotonics International 11:36–42.

Albert, A. Y., S. Sawaya, T. H. Vines, A. K. Knecht, C. T. Miller, B. R. Summers, S. Balabhadra, D. M. Kingsley, and D. Schluter. 2008. The genetics of adaptive shape shift in stickleback: Pleiotropy and effect size. Evolution 62:76–85.

Archambeault, S. L., L. R. Bärtschi, A. D. Merminod, and C. L. Peichel. 2020. Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback ¡i¿Eda¡/i¿ locus. Evolution Letters pages 1–20.

Arnegard, M. E., M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte, S. Kabir, N. Bedford, S. Bergek, Y. F. Chan, F. C. Jones, D. M. Kingsley, C. L. Peichel, and D. Schluter. 2014. Genetics of ecological divergence during speciation. Nature 511:307–311.

Auguie, B. 2019. egg: Extensions for 'ggplot2': Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size. R package version 0.4.5.

Barton, N. H. 2001. The role of hybridization in evolution. Molecular Ecology 10:551–568.

Bates, D., M. Maechler, B. Bolker, and W. S. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Bell, M. A., and S. A. Foster. 1994. The evolutionary biology of the threespine stickleback. Oxford University Press.

Breheny, P., and W. Burchett. 2017. Visualization of regression models using visreg. R Journal 9:56–71.

Chevin, L.-M., G. Decorzent, and T. Lenormand. 2014. Niche dimensionality and the genetics of ecological speciation. Evolution 68:1244–1256.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, H. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928–1933.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer.

Dalziel, A. C., T. H. Vines, and P. M. Schulte. 2012. Reductions in prolonged swimming capacity following freshwater colonization in multiple threespine stickleback populations. Evolution 66:1226–1239.

Edmands, S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. Evolution 53:1757–1768.

Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proceedings of the National Academy of Sciences of the United States of America 103:3209–3213.

Hagen, D. W. 1967. Isolating Mechanisrns in Threespine (Gasterosteus). Journal Fisheries Research Board of Canada 24:1637–1692.

Hatfield, T., and D. Schluter. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution 53:866–873.

Henry, L., and H. Wickham. 2019. purrr: Functional programming tools.

Hothorn, T., F. Bretz, P. Westfall, and R. M. Heiberger. 2008. multcomp: simultaneous inference for general linear hypotheses, 2008.

Jaeger, B. 2017. r2glmm: Computes R Squared for Mixed (Multilevel) Models.

Jones, F. C., C. Brown, J. M. Pemberton, and V. A. Braithwaite. 2006. Reproductive isolation in a threespine stickleback hybrid zone. Journal of Evolutionary Biology 19:1531–1544.

Kassambara, A. 2020. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr.

Kenward, M. G., and J. H. Roger. 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997.

Kuznetsova, A., P. B. Brockhoff, and H. B. Christensen. 2014. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). v 2.0-11. http://CRAN.R-project.org/package=lmerTest.

Lackey, A. C. R., and J. W. Boughman. 2016. Evolution of reproductive isolation in stickleback fish. Evolution .

Lenth, R., H. Singmann, J. Love, P. Buerkner, and M. Herve. 2020. Package ' emmeans ' R topics documented :. R package version 1.15-15 34:216–221.

Makiyama, K. 2016. magicfor: Magic Functions to Obtain Results from for Loops.

Makowski, D., M. S. Ben-Shachar, I. Patil, and D. Lüdecke. 2019. Methods and Algorithms for Correlation Analysis in R. Journal of Open Source Software 5:2306.

Matsubayashi, K. W., I. Ohshima, and P. Nosil. 2010. Ecological speciation in phytophagous insects. Entomologia Experimentalis et Applicata 134:1–27.

McPhail, J. D. 2008. Ecology and evolution of sympatric sticklebacks (¡i¿Gasterosteus¡/i¿): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. Canadian Journal of Zoology 70:361–369.

Miller, C. T., A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman, M. D. Shapiro, B. L. Cole, C. L. Peichel, D. Schluter, and D. M. Kingsley. 2014. Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. Genetics 197:405–420.

Miller, S. E., M. Roesti, and D. Schluter. 2019. A Single Interacting Species Leads to Widespread Parallel Evolution of the Stickleback Genome. Current Biology 29:530–537.

Morris, M. R., E. Bowles, B. E. Allen, H. A. Jamniczky, and S. M. Rogers. 2018. Contemporary ancestor? Adaptive divergence from standing genetic variation in Pacific marine threespine stickleback. BMC Evolutionary Biology 18:113.

Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4:133–142.

Nosil, P. 2012. Ecological speciation. Oxford University Press, Oxford, UK.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robinson, D., A. Hayes, and S. Couch. 2020. broom: Convert Statistical Objects into Tidy Tibbles.

Rogers, S. M., P. Tamkee, B. Summers, S. Balabahadra, M. Marks, D. M. Kingsley, and D. Schluter. 2012. Genetic signature of adaptive peak shift in threespine stickleback. Evolution 66:2439–2450.

Schluter, D. 1996. Ecological speciation in postglacial fishes. Philosophical Transactions of the Royal Society B: Biological Sciences 351:807–814.

———. 2000. The Ecology of Adaptive Radiation. Oxford University Press, New York.

Schluter, D., E. A. Clifford, M. Nemethy, and J. S. McKinnon. 2004. Parallel evolution and inheritance of quantitative traits. The American Naturalist 163:809–822.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proceedings of the National Academy of Sciences of the USA 106.Sup1:9955–62.

Schluter, D., K. B. Marchinko, M. E. Arnegard, H. Zhang, S. D. Brady, F. C. Jones, M. Bell, and D. Kingsley. 2020. Fitness maps to a large-effect locus in introduced stickleback populations. Proceedings of the National Academy of Sciences of the United States of America page In press.

Shafer, A. B. A., and J. B. W. Wolf. 2013. Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. Ecology Letters 16:940–950.

Slatkin, M., and R. Lande. 1994. Segregation variance after hybridization of isolated populations. Genetical Research 64:51–56.

Thompson, K. A. 2020. Experimental hybridization studies suggest that pleiotropic alleles commonly underlie adaptive divergence between natural populations. The American Naturalist In press.

Thompson, K. A., M. M. Osmond, and D. Schluter. 2019. Parallel genetic evolution and speciation from standing variation. Evolution Letters 3:129–141.

Thompson, K. A., M. Urquhart-Cronish, K. D. Whitney, L. H. Rieseberg, and D. Schluter. 2021. Patterns, predictors, and consequences of dominance in hybrids. The American Naturalist page In press.

Vines, T., A. Dalziel, A. Albert, T. Veen, P. Schulte, and D. Schluter. 2016. Cline coupling and uncoupling in a stickleback hybrid zone. Evolution 70:1023–1038.

Vinšálková, T., and L. Gvoždík. 2007. Mismatch between temperature preferences and morphology in F1 hybrid newts (¡i¿Triturus carnifex¡/i¿ × ¡i¿T. dobrogicus¡/i¿). Journal of Thermal Biology 32:433–439.

Wickham, H. 2017. Tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1.

Wright, S. 1968. Evolution and the Genetics of Populations Vol. 1. Genetic and Biometric Foundations. University of Chicago Press, Chicago, IL.

Supplementary Material

Supplementary methods

Details of crossing protocol and fish husbandry

Female stickleback were selected for spawning when their abdomens were sharply angled at the cloaca and the first egg was visible. We gently squeezed the sides of female fish's body to release the eggs into a petri dish containing water from her source habitat (tank, lake, or river water). Mature males were identified by their bright blue colouration and red throat. Males were euthanized with an overdose of MS-222, and then testes were extracted from the body cavity using fine forceps after making a small incision beginning at the cloaca. We used a small paintbrush to release sperm from testes and ensure sperm contacted all eggs. The live fish and eggs that were crossed in the field were transported to the InSEAS aquatic facility at the University of British Columbia, Vancouver, British Columbia, Canada.

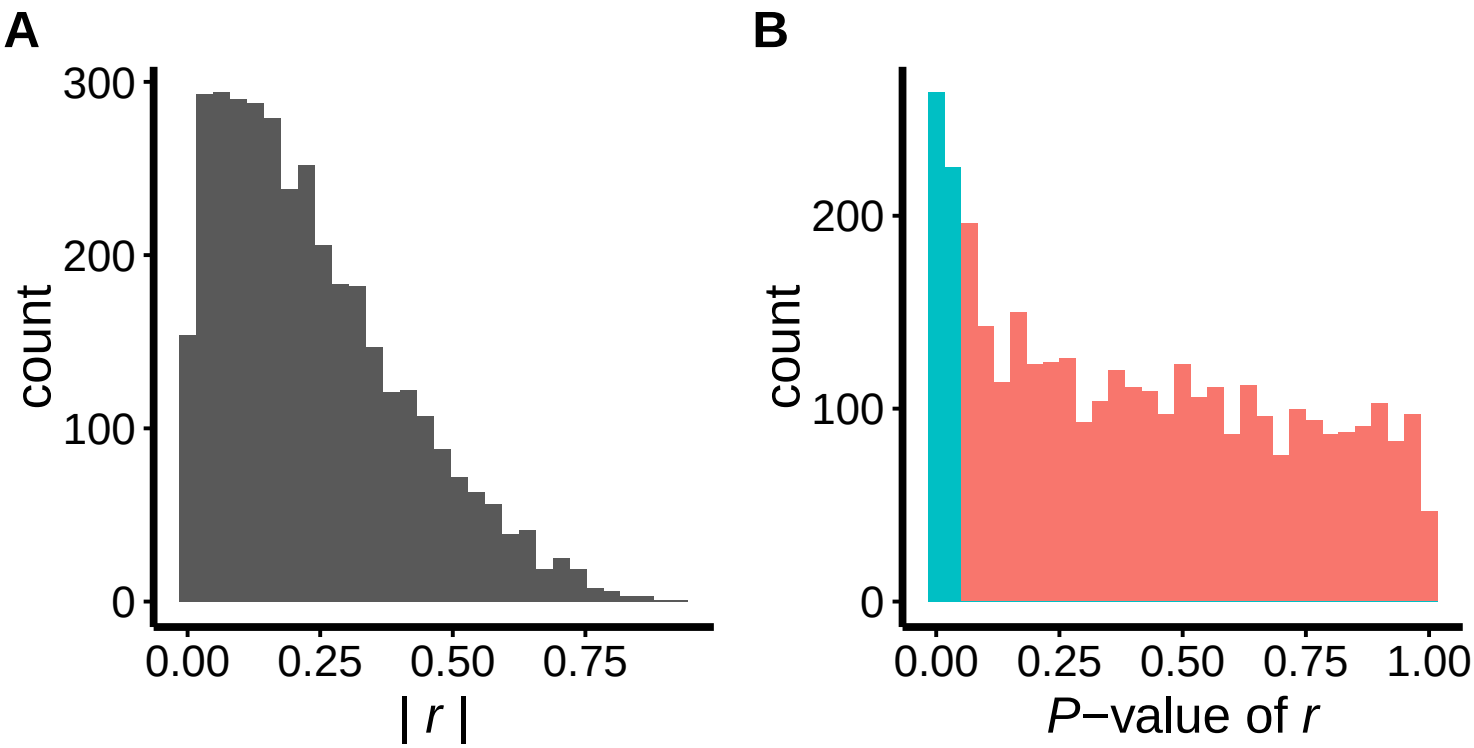
All fish were hatched in 100 L aquaria with room temperature between 17 and 19 °C and a photoperiod that followed local dawn and dusk times. Instant Ocean® Sea Salt was added to maintain a salinity of 5 ppt in all tanks. Fry were fed live brine shrimp nauplii. Chopped frozen bloodworms were added to the diet when fish were large enough, and then finally adult-sized fish were fed full size frozen bloodworms and frozen mysis shrimp *ad libitum* (Hikari Bio-Pure®).

We sampled fish for phenotype measurements typically when the mean standard length of a family was approximately 40 mm. Sticklebacks have adult morphology at this stage and are not sexually reproducing. Due to occasional logistic constraints, some tanks were sampled at earlier or later mean standard length sizes. Also, due to logistic constraint, all populations except for Paxton benthic and Paxton limnetic were collected from lakes and river in 2017 (see Fig. 1)

Data diagnostics

We checked for outliers in the raw data and evaluated outlier individuals to ensure they were not caused by measurement or transcription error. If fish were inadvertently measured twice, we averaged trait values across measurements. Two fish had broken second dorsal spines, so we divided the family average of size-corrected second dorsal spine length by the family average of size-corrected first dorsal spine length to replace the missing values. One fish (KT_Gac_1186) was removed because it had an unusual body shape—qualitatively appearing as if it had failed to inflate its swim bladder—and it was an extreme outlier in Normal Q-Q plot and in standardized residuals vs. leverage plot. ln-transformations were applied if they improved the diagnostic plots.

537 *Supplementary figures*



.

Figure S1: **Histogram summaries of pairwise trait correlations in F_2 hybrid families.** Panel (A) shows the distribution of Pearson's r values for all trait pairs in all F_2 hybrid families where $\geq$ 10 fish were measured. The median r was 0.20. Panel (B) shows the distribution of uncorrected P-values for the correlation matrices. 13.6 % of correlations are significant (blue bars) at $\alpha = 0.05$.

Table 1: Fraction of traits exhibiting statistically significant segregation variance (Eqn. 3), determined using the bootstrap.

	trait	fraction of populations with sig. segregation variance
1	pectoral fin length	1 / 12
2	second dorsal spine length	1 / 12
3	pelvic spine length	1 / 12
4	anal fin rays	2 / 12
5	dorsal fin rays	2 / 12
6	body depth	3 / 12
7	gape width	3 / 12
8	body width	3 / 12
9	gill raker number	4 / 12
10	gill raker length	4 / 12
11	pelvic girdle length	5 / 12
12	lateral armour plates	12 / 12

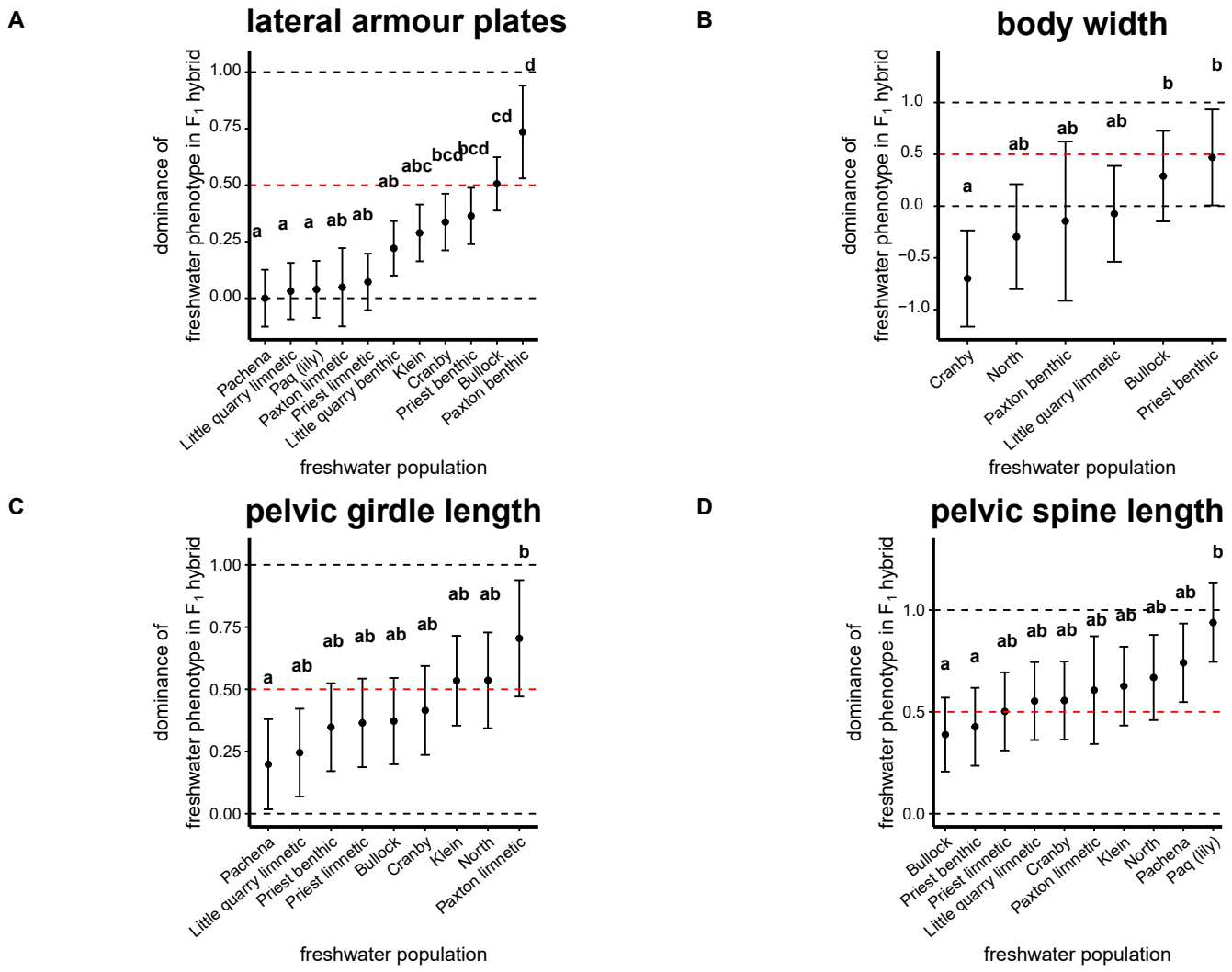
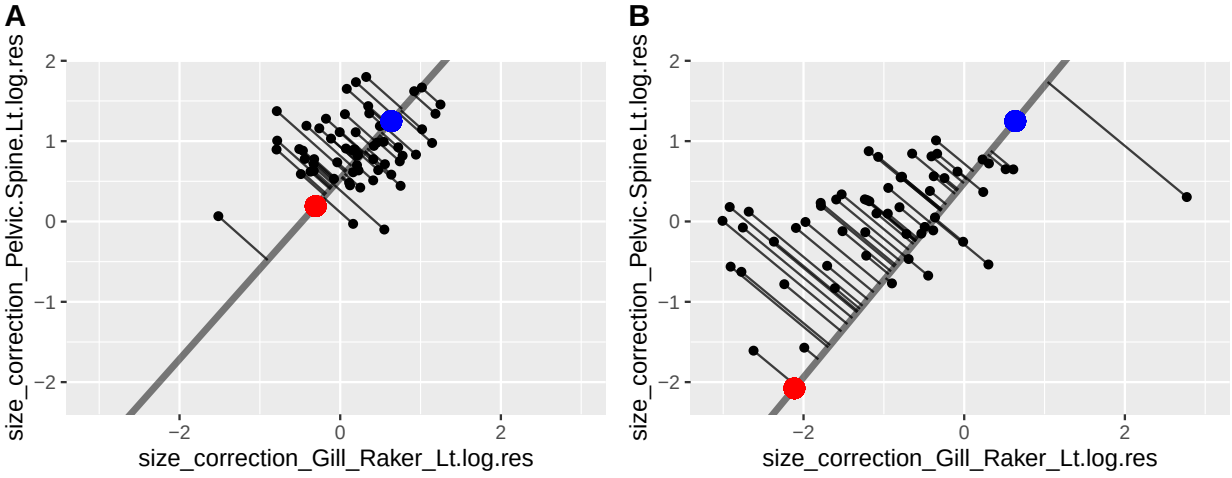


Figure S2: **Plot-of-means showing traits for which dominance differs between at least two freshwater populations** Points and error bars represent mean dominance and 95% confidence intervals for each trait and each freshwater population. Letters indicate freshwater populations that are different at $P < 0.05$ (N = 870 fish including F_1 hybrids and parents). The dashed black lines at 0 and 1 represent ancestral anadromous parent and derived freshwater parent trait values, and the red line at 0.5 indicates no dominance.



.

Figure S3: **Visualization of individual-level mismatch in F_2 hybrids between gill raker number and length in crosses with low (A; Pachena Lake) and high (B; Priest lake benthic) divergence for these traits.** In each plot, the red point is the freshwater parent mean and the blue point is the anadromous parent mean. Small black points are the individual hybrid phenotypes. The thick line connects the parent means and is the reference point for mismatch vectors. Thin black lines connect each hybrid to the line that connects parents; these lines are the mismatch vectors and their length is what we call 'mismatch'. The mean length of the vectors in (A) is 1.02 and in (B) is 1.34—this indicates that mismatch for these two traits in F_2 Priest Lake Benthic hybrids is approximately 31 % greater than in F_2 Pachena hybrids.