

# Supplementary material for:

## *Heterosis counteracts hybrid breakdown to forestall speciation by parallel natural selection*

by:

Ken A. Thompson & Dolph Schluter

Contents: Supplementary methods, Table S1 & S2, Fig. S1–S14

### Supplementary methods

#### Determining experiment size

Before the experiment, we conducted a power analysis to determine a sample size that would be necessary to detect a 2% difference in the mean of a fitness component between any two cross types within a pond at the 5% significance level with 80% probability. To estimate the level of variability in fitness we might observe in a pond experiment, we used the standard length data of Arnegard et al. (2014) as our fitness metric. We determined that a comparison of approximately 100 fish in the two cross types was sufficient for this purpose. We anticipated that 50% of fish would perish during the experiment—so for each pond we introduced 200 fish per cross type (i.e., pure, F<sub>1</sub> hybrid, and F<sub>2</sub> hybrid) per ecotype.

#### Experimental animals

Wild fish were collected from Priest and Paxton Lakes (Texada Island, BC, Canada) and Little Quarry Lake (Nelson Island, BC) from 2017–2019. Two Paxton benthic males were collected from a pond population on UBC campus founded with wild Paxton Lake benthic fish in 2016. All fish in the experiment were therefore  $\geq 2$  generations removed from the wild. All benthics were captured using minnow traps as were most limnetic males. Gravid limnetic females were caught almost invariably by dip-netting. One Paxton limnetic family and one Little Quarry limnetic family was raised from a nest-collected clutch. In these cases we closely examined the resulting fish to ensure none were benthic  $\times$  limnetic hybrids.

Wild fish were crossed either at the lakeside or at the lab by gently stripping the eggs from a gravid female fish into a small Petri dish filled with lake or aquarium water. Male parents were euthanized with

an overdose of MS-222, and their testes were removed and placed in the Petri dish. A fine paintbrush was then used to release sperm from the testes and ensure it was well mixed among the eggs in the clutch. Crosses for the present experiment were all made in the lab in much the same way. Males were occasionally used to fertilize multiple clutches. Due to logistical constraints, we made crosses as females became gravid.

Crosses were made from 4 March to 9 April, 2020, and raised in 5 ppt saltwater (Instant Ocean) within 110 L aquaria with a small amount of methylene blue added as a fungicide. These two additions reduce the loss of clutches to fungus and also reduce labour required to raise healthy fish (because *Artemia* nauplii can live for 24 hr in 5 ppt saltwater). Larval fish were fed *Artemia* nauplii daily. Due to logistical constraints we could not monitor hatching success but note that we observed very little mortality. In general, there is very limited evidence that F<sub>1</sub> or F<sub>2</sub> hybrid stickleback exhibit ‘intrinsic’ deficiencies (Lackey and Boughman, 2017).

After the fish were large enough to handle without risk of mortality (approx. 3–4 weeks), we split large families into multiple tanks and culled excess fish. At this time we began feeding fish chopped frozen bloodworms and conducting weekly 50% water changes with fresh dechlorinated water. After splitting and culling we kept the density of fish to fewer than 30 individuals per aquarium. When fish were large enough, we added unchopped bloodworms, spirulina adult brine shrimp, and chopped mysis shrimp to their diets. Eventually, the mysis shrimp were fed whole. Fish were collected under Species at Risk Act (SARA) permits from Fisheries and Oceans Canada (SARA 16-PPAC-00004, 17-PPAC-00002, 18-PPAC-00006, 19-PPAC-00006) and the British Columbia Ministry of Forests, Lands and Natural Resource Operations (SU17-258923, MRSU18-288855, MRSU18-454239). All animal care protocols were approved by the University of British Columbia Animal Care Committee (A16-0044, A20-0050).

## **Coded wire tagging**

Here we fully document our methodology for using sequential coded wire tags (hereafter ‘tags’). We ordered sufficient quantities of tags from Northwest Marine Technology (<https://www.nmt.us/>; Anacortes, WA, USA). The tags come on sheets with two columns—one ‘fish’ column and one ‘reference’ column. Although tags appear in sequential order, the numbers do not increase in perfect 1:1 association with tag position. Because of this, the unique sequences on tags cannot be inferred without some possibility of error. It is therefore necessary to pre-read some tags. We never used adjacent (i.e., on the same row) tags for fish bound for the same pond. This allowed us to pre-read fewer tags without risking error.

Specifically, we pre-read every third tag in the ‘reference’ column using a Magniviewer (Northwest Marine Technology) and could reliably infer other tags because we know which pond each fish was retrieved from.

Before tagging, fish were anaesthetised with MS-222. We then used a single-shot CWT injector (Northwest Marine Technology) to inject tags beneath the skin on the fish’s dorsal musculature. Fish were injected while laying on their side atop a large sponge, and their head was covered with a paper towel soaked in water from their original tank. Light pressure was applied to the head and caudal peduncle to stabilize the fish during injection. Injection was easiest if the lateral plates were used as a leverage point to implant the head of the needle beneath the skin. (Limnetic fish, with their denser musculature and invariable presence of lateral plates, were easier to inject). We found maximal success when the push rod was not pushed to be maximally extended—in fact this can cause the CWT to emerge from the fish, rather an extension of about  $\frac{3}{4}$  worked best. When injecting, we took care that the pectoral fin was either oriented toward the head or down toward the ventral area, so as to not be pierced by the tagging needle. Fish were kept temporarily in aerated water from their original tank for recovery and then moved back to their original tank before introduction into the ponds three days later. Methylene blue was added to their original tank to prevent infection of the tagging wound. We estimate that each fish took approximately 30 s to anaesthetize, weigh, and tag, and it took about two min to pre-read a page of tags.

At the end of the experiment we retrieved and read tags from each fish. We located tags under a dissecting microscope. Tags were always handled delicately because metal forceps can easily damage them. Most often, a scalpel was used to cut away skin and visually identify the tag. After cutting away muscle, the magnetic tag is attracted to the scalpel, where it can be gently placed onto the end of the magnetic ‘pencil’ that is used with the MagniViewer. Removed tags were read with the MagniViewer and then stored for later reference if necessary. A magnetic T-wand (NMT) was occasionally used to confirm the presence of a tag in a fish (or a piece of its flesh).

Ambiguities and possible transcription errors were identified by ensuring a match between ecotype (benthic or limnetic; assigned visually from re-captured fish) and pond, checking for duplicates and implausible values, and ensuring each recovered tag was assigned to a fish.

### **Estimating fitness via fecundity and overwinter survival**

In the main text we use a conservative fitness estimate that incorporates only survival and growth. A more complicated and possibly more accurate estimate of fitness would account for overwinter survival

and fecundity. This estimate of composite fitness considers three components. The first, survival during the experiment (i.e., ‘summer survival), was directly measured. The next two fitness components consider the fitness effects of body size (directly measured) on overwinter survival and fecundity (both estimated). Absolute fitness was calculated as summer survival  $\times$  (estimated) winter survival  $\times$  (estimated) fecundity. Composite fitness is reported as the absolute fitness of each cross type divided by the cross type with the highest absolute fitness (within ecotype), and thus is relative to the pure crosses.

We estimated relative overwinter survival and fecundity using previously published data. These previous studies used standard length to predict fitness components, and we measured approximately 50 individuals of both ecotypes to estimate the mass–length relationship for both. Quadratic models had a high explanatory pattern for crosses of both ecotypes ( $r^2_{\text{limnetic}} = 0.88$ ;  $r^2_{\text{benthic}} = 0.79$ ) (Fig. S13).

To estimate relative overwinter survival, we conservatively interpret the analysis of Carlson et al. (2010), who found that standard length was positively associated with overwinter survival in Alaskan stickleback. The specific relationship between length and survival was either quadratic (concave) or linear, depending on the year (Carlson et al., 2010). We conservatively assume that relative overwinter survival is a linear function of standard length. I estimated fecundity using two previously published datasets from stickleback experiments in the UBC Experimental Ponds (Fig. S14). Schluter et al. (2021) estimated fecundity in  $F_2$  marine  $\times$  freshwater stickleback hybrids. Specifically, the fitness of over 200  $F_2$  hybrid females was quantified as the tally of her surviving  $F_3$  offspring (from a sample of 500  $F_3$ s). Male fitness was not estimated by Schluter et al. (2021), but the authors speculate that selection acted similarly on males and females because the evolutionary response observed was highly similar to what was expected from estimates of female fitness. Bay et al. (2017) estimated the number of mating events for Paxton Lake  $F_2$  benthic  $\times$  limnetic hybrid females, and pure (i.e., pure) benthic and limnetic males. Analysis of the data suggested that length–mating relationships were the same across groups (i.e., no significant interaction) so all data were grouped for analysis. Data were analyzed using simple linear models from which intercepts and slopes were extracted. Relationships were significant (in a Poisson generalized linear model) and positive in both datasets, but we chose to base fecundity estimates on the estimates from Schluter et al. (2021) because differences in composite fitness among groups were smaller (i.e., it is more conservative).



## Supplementary Tables & Figures

Table S1: Summary of fish numbers and survival (re-capture numbers) for the 2020 pond experiment.

pond	origin	ecotype	type	$n$ released	$n$ recaptured	survival proportion
4	Little Quarry	b	Pure	102	46	0.45
4	Paxton	b	Pure	107	70	0.65
4	hybrid	b	F <sub>1</sub>	206	123	0.60
4	hybrid	b	F <sub>2</sub>	210	130	0.62
4	Little Quarry	l	Pure	106	31	0.29
4	Paxton	l	Pure	107	39	0.36
4	hybrid	l	F <sub>1</sub>	207	101	0.49
4	hybrid	l	F <sub>2</sub>	208	62	0.30
9	Little Quarry	b	Pure	104	84	0.81
9	Priest	b	Pure	106	93	0.88
9	hybrid	b	F <sub>1</sub>	203	185	0.91
9	hybrid	b	F <sub>2</sub>	205	169	0.82
9	Little Quarry	l	Pure	105	36	0.34
9	Priest	l	Pure	105	57	0.54
9	hybrid	l	F <sub>1</sub>	204	86	0.42
9	hybrid	l	F <sub>2</sub>	205	100	0.49
19	Priest	b	Pure	103	95	0.92
19	Paxton	b	Pure	103	87	0.84
19	hybrid	b	F <sub>1</sub>	204	174	0.85
19	hybrid	b	F <sub>2</sub>	205	161	0.79
19	Priest	l	Pure	105	32	0.30
19	Paxton	l	Pure	105	42	0.40
19	hybrid	l	F <sub>1</sub>	206	88	0.43
19	hybrid	l	F <sub>2</sub>	205	72	0.35

Table S2: Fitness components and composite fitness estimates for cross types within ecotype. We used estimates of the relationship between mass and standard length (Fig. S13), standard length and fecundity (Fig. S14) and standard length and overwinter survival (Carlson et al., 2010) to generate estimates of composite fitness (survival  $\times$  final mass) for each cross type for crosses of both ecotypes ('alt. fit' column). Because the patterns were less conservative than a simple multiplicative estimate of composite fitness ('comp. fit..' column), we opted to use the more conservative method.

ecotype	cross	sum. surv.	mass (g)	std. lgth. (cm) <sup>*</sup>	win. surv. <sup>*</sup>	fecundity <sup>*</sup>	alt. fit. <sup>†</sup>	comp. fit.
benthic	pure	0.799	2.06	5.74	0.98	3.53	1	1
	F <sub>1</sub>	0.818	2.14	5.84	1.00	3.69	1.08	1.07
	F <sub>2</sub>	0.797	2.01	5.67	0.97	3.43	0.98	0.98
limnetic	pure	0.387	1.01	4.42	0.97	1.49	1	1
	F <sub>1</sub>	0.454	1.09	4.56	1.00	1.69	1.34	1.27
	F <sub>2</sub>	0.386	0.98	4.37	0.96	1.41	0.94	0.97

<sup>\*</sup> estimate; <sup>†</sup>  $w = \text{sum. surv} \times \text{wint. surv.} \times \text{fecundity}$

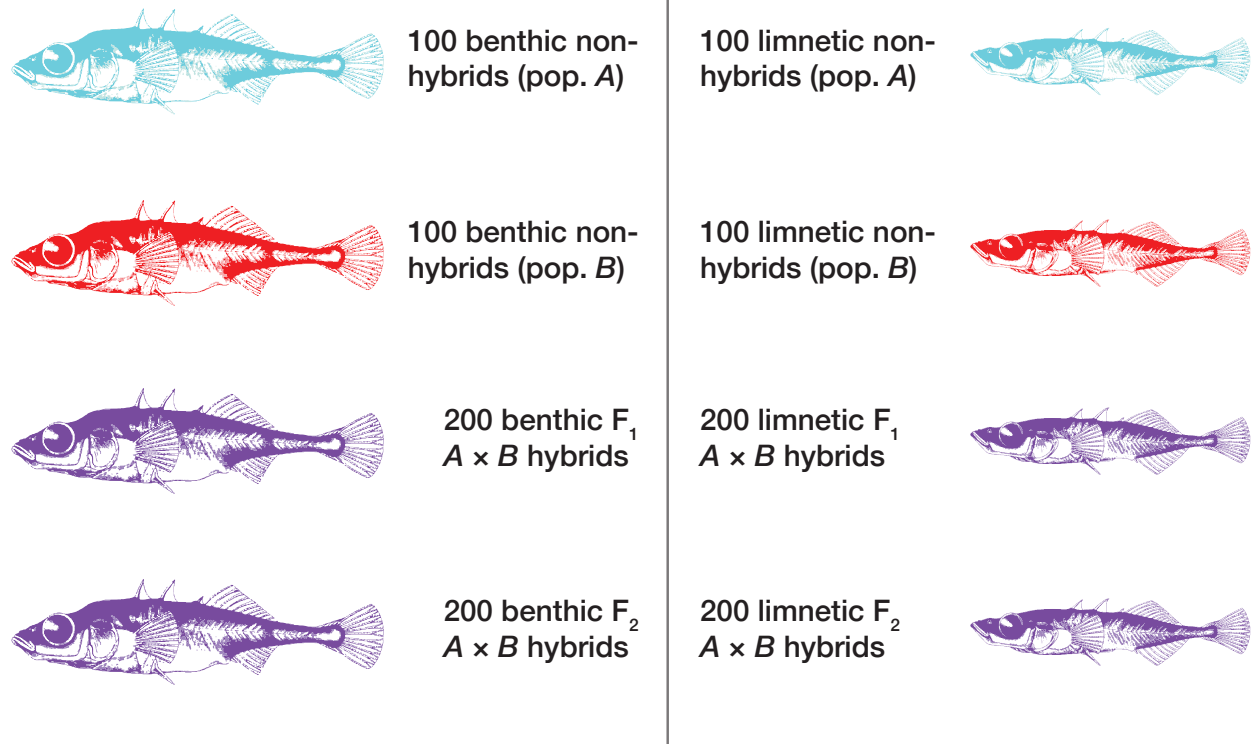


Figure S1: **Overview of the experimental design.** Each pond contained individuals from four cross types from each of the two ecotypes. In this figure (and not in the main text), red and cyan are the two different colours that represent two different pure populations ('non-hybrids'). Purple represents (F<sub>1</sub> and F<sub>2</sub>) hybrids.

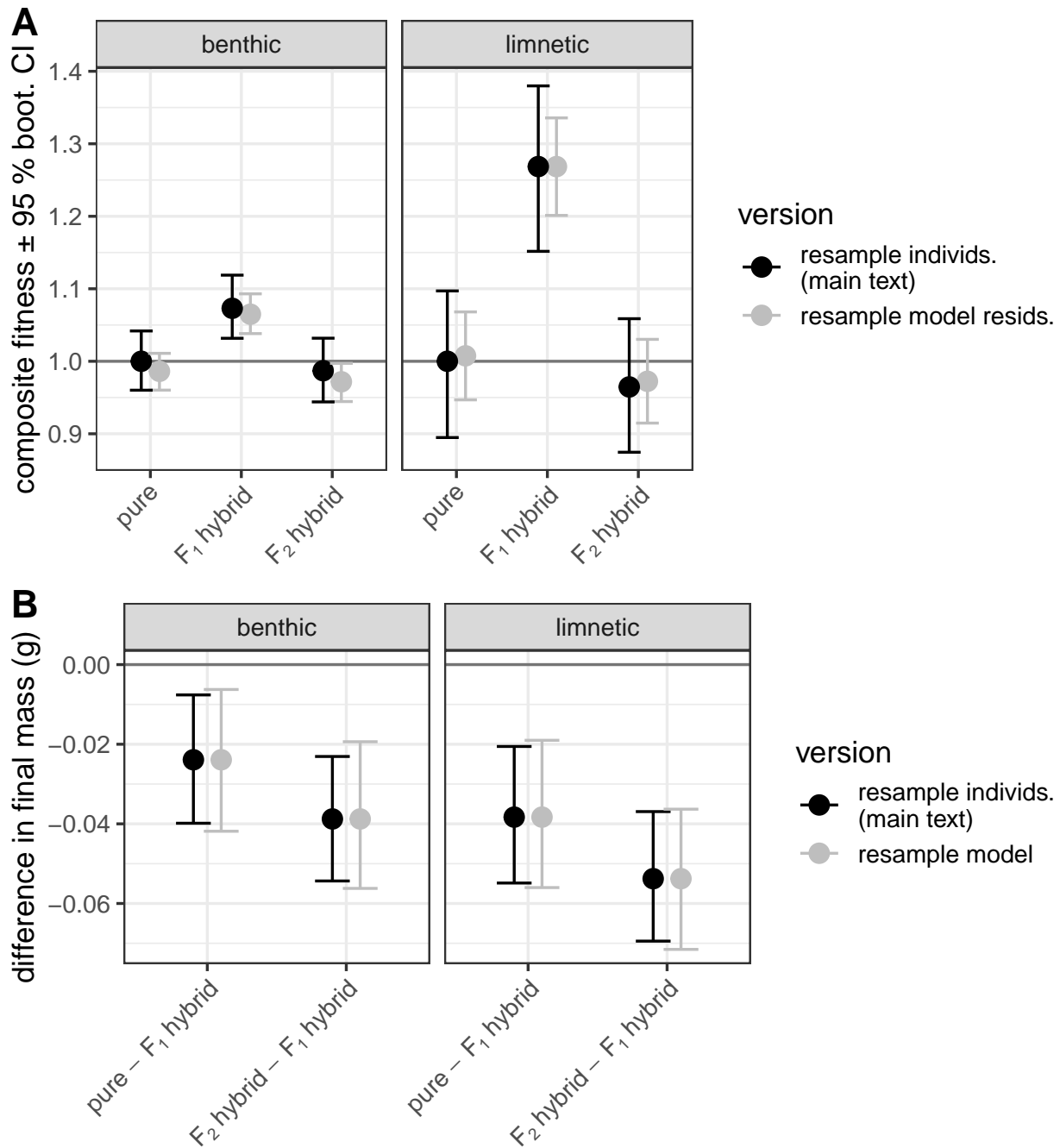


Figure S2: **Conclusions of alternative bootstrap procedures.** Panel (A) shows a case where model residuals are resampled (grey) in comparison to the resampling of individuals. Means are from the resampling but absolute fitness values are divided by the observed fitness of pure crosses. Panel (B) shows estimates of group differences in final mass ( $F_2 - F_1$  and pure -  $F_1$ ) from a parametric bootstrap resampling procedure implemented via the `lmeresample` R package (Loy et al., 2021). This procedure has slightly wider confidence intervals than our approach because it simulates the error term for the ‘lake pair’ random effect, which with only three levels is estimated with poor accuracy.

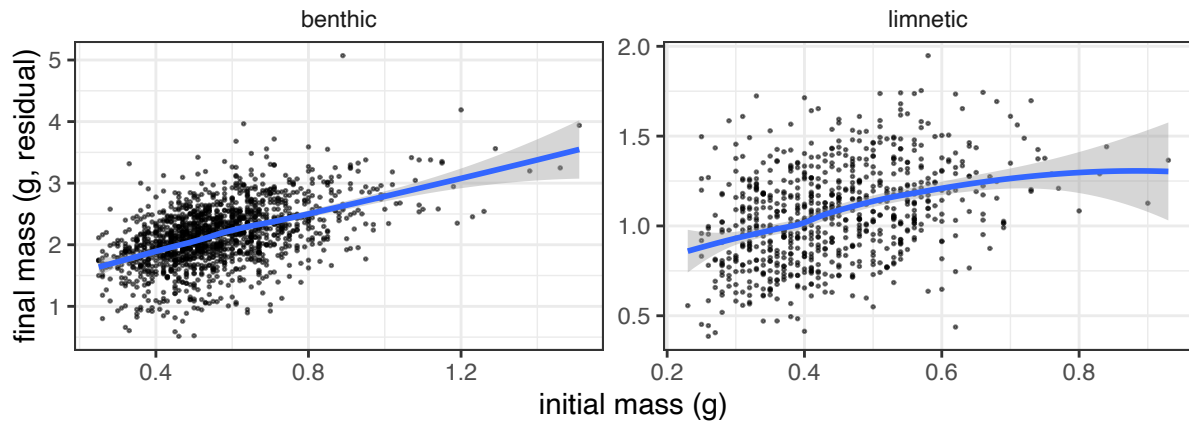


Figure S3: **Relationship between initial and final mass in the pond experiment.** The left plot shows benthic crosses and the right shows limnetic crosses. Points are partial residuals from a mixed model with lake pair as a random effect (pair-as-random models), extracted using `visreg` (Breheny and Burchett, 2017). We fit a smooth curve to the data here to illustrate patterns, but we analyze the mass data on a  $\ln(\text{mass} + 1)$  scale. Note the difference in axis scales.

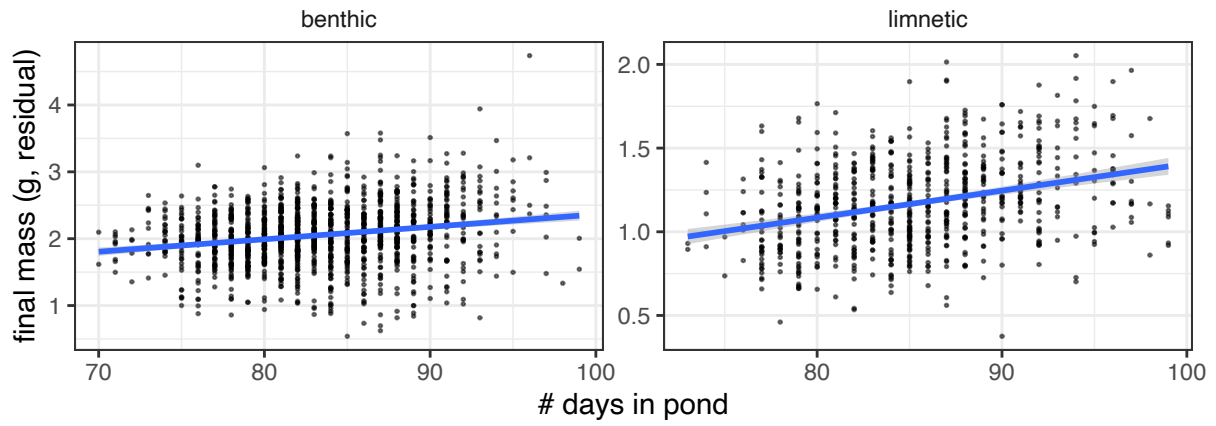
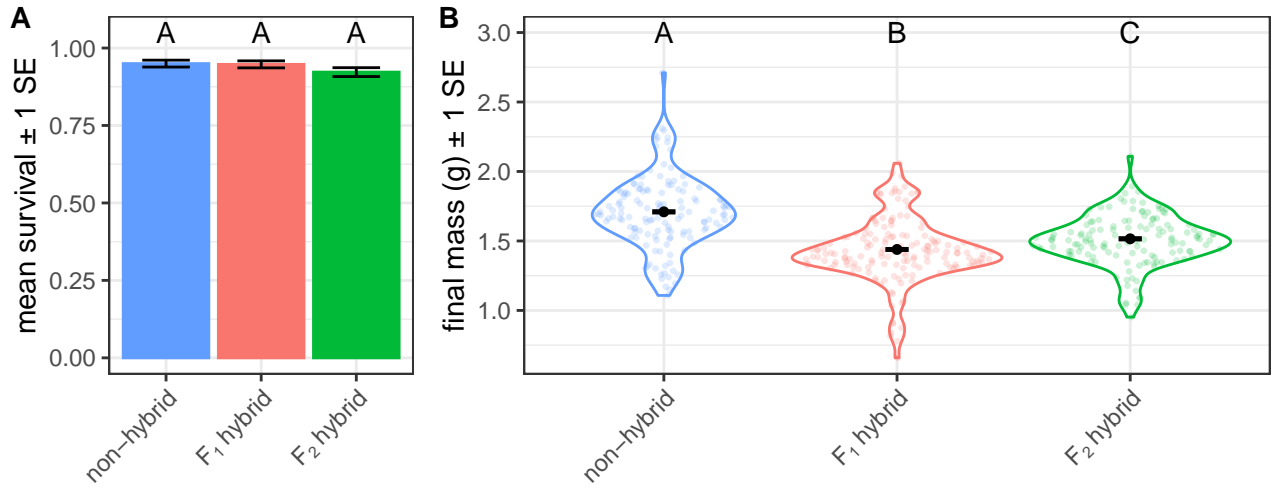


Figure S4: **Relationship between the number of days in pond and final mass in the 2020 pond experiment.** Both relationships are significant and positive. Note the difference in axis scales.

### fitness components in aquaria (benthic)



### fitness components in aquaria (limnetic)

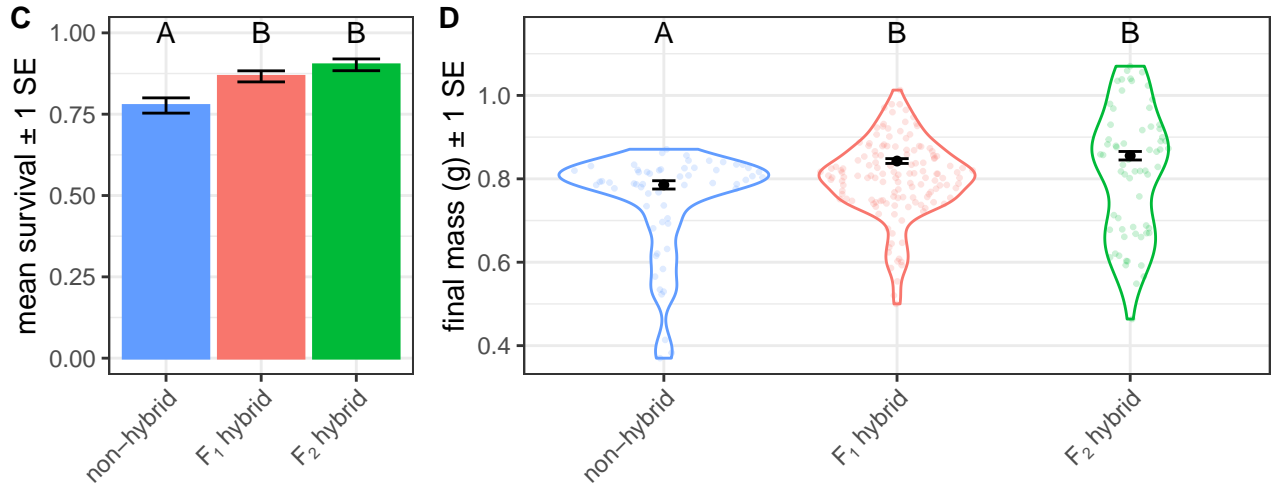


Figure S5: **Components of fitness in aquaria.** These data were used to generate the estimates of composite fitness in Fig. 3.

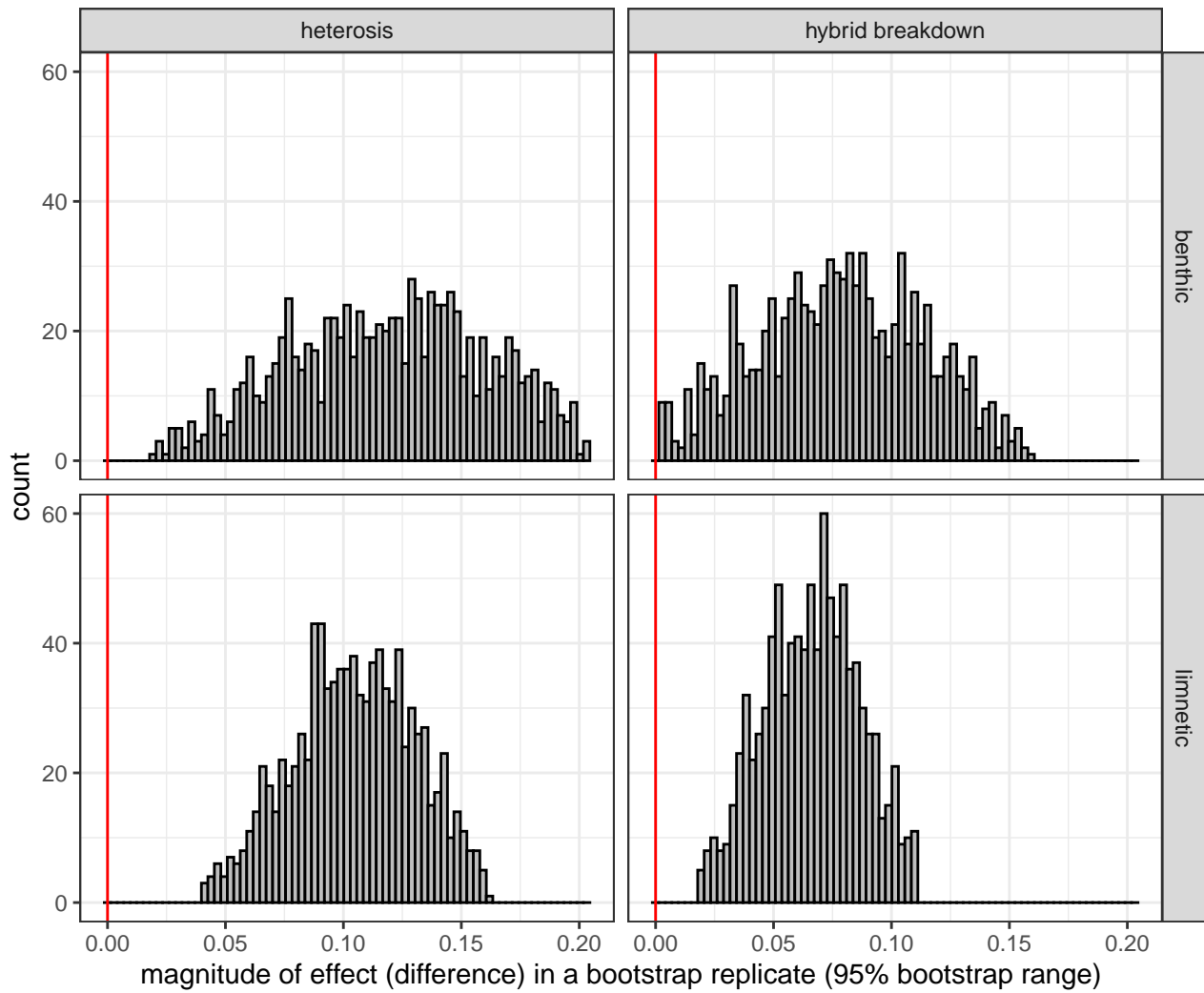


Figure S6: **Distribution of bootstrap estimates of heterosis (left) and hybrid breakdown (right) via composite fitness for both benthic crosses (top) and limnetic crosses (bottom) (95% range) in ponds.** All 95% ranges do not include 0.



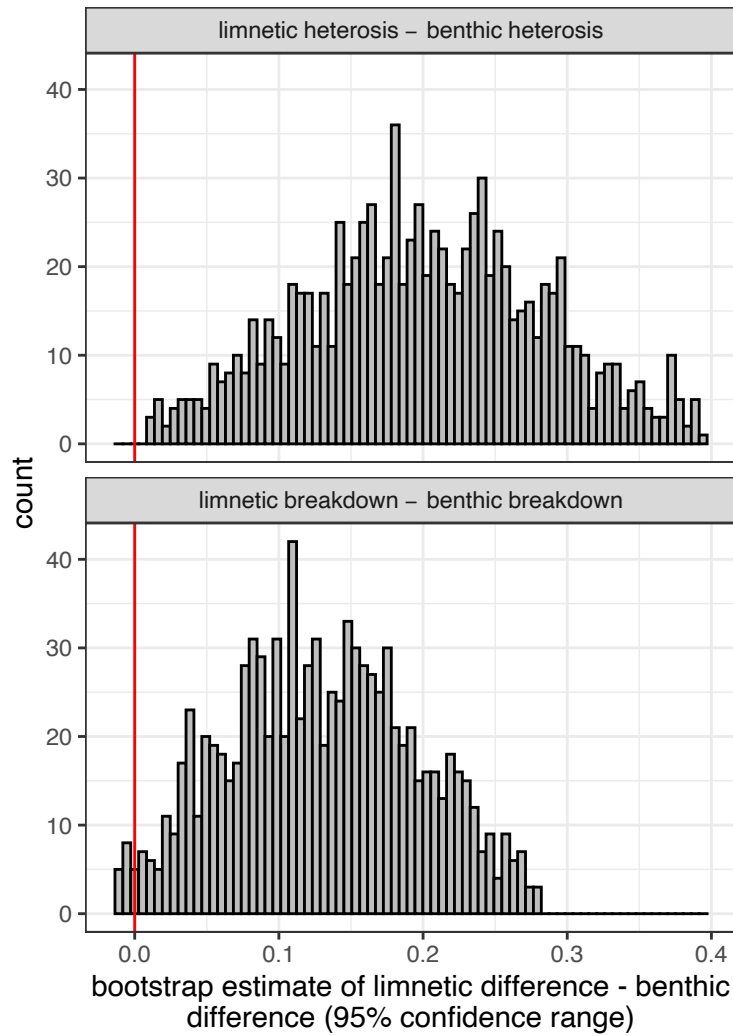


Figure S7: **Comparison of the magnitude of heterosis and breakdown between benthic crosses and limnetic crosses.** Heterosis and breakdown are both estimated as differences, and this analysis is therefore testing for a difference in these differences. Values show the 95% bootstrap confidence range, and ranges that do not include zero (red vertical line) are considered significant.

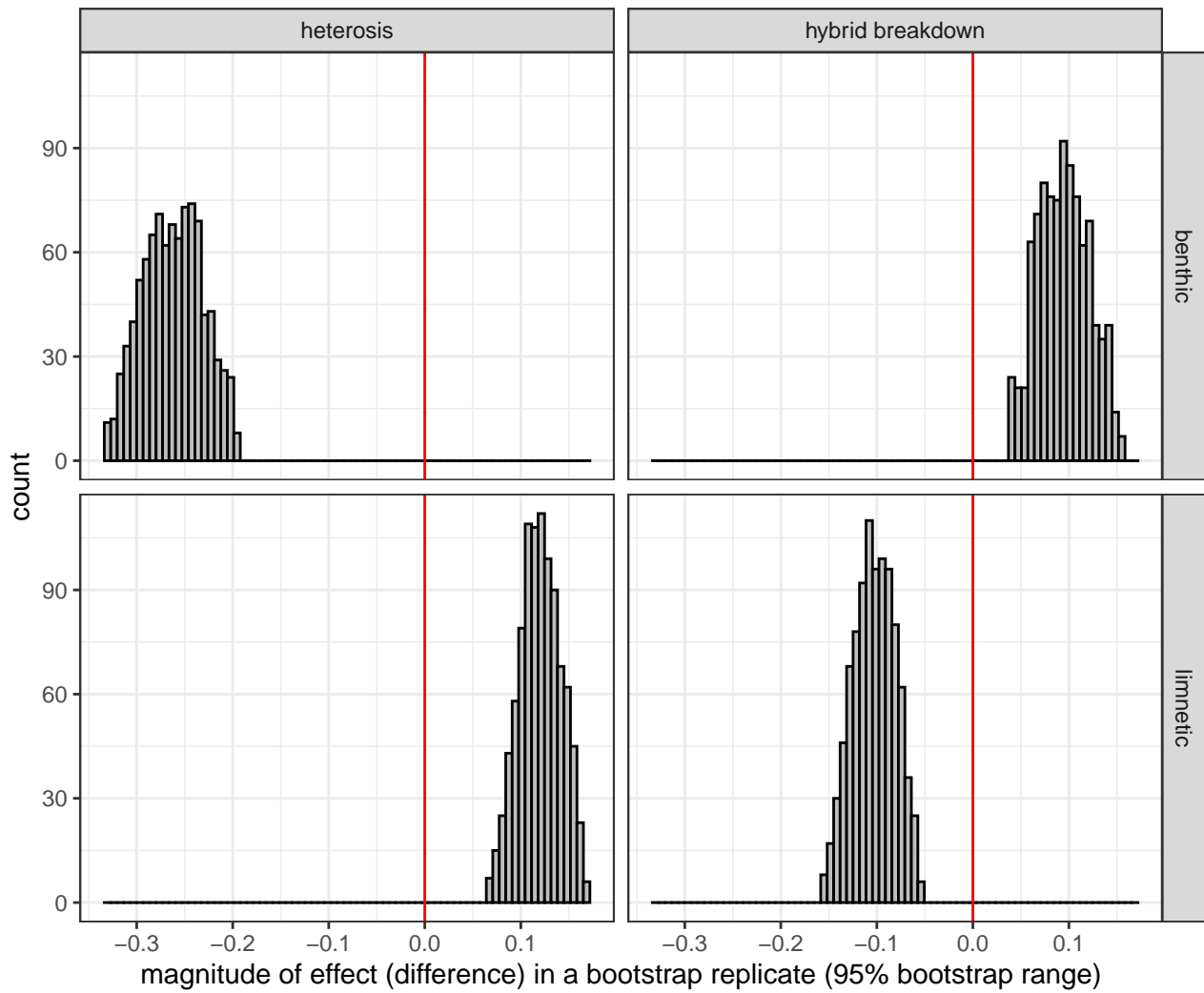


Figure S8: **Distribution of bootstrap estimates of heterosis (left) and hybrid breakdown (right) via composite fitness for both benthic crosses (top) and limnetic crosses (bottom) (95% range) in *aquaria*.** All 95% ranges do not include 0.

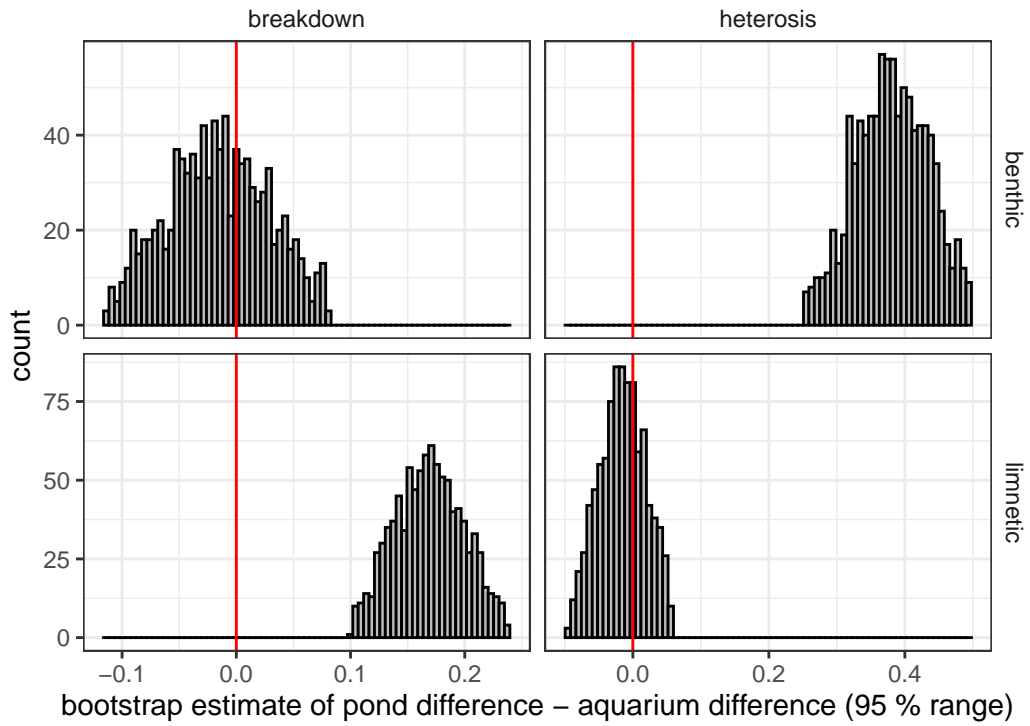
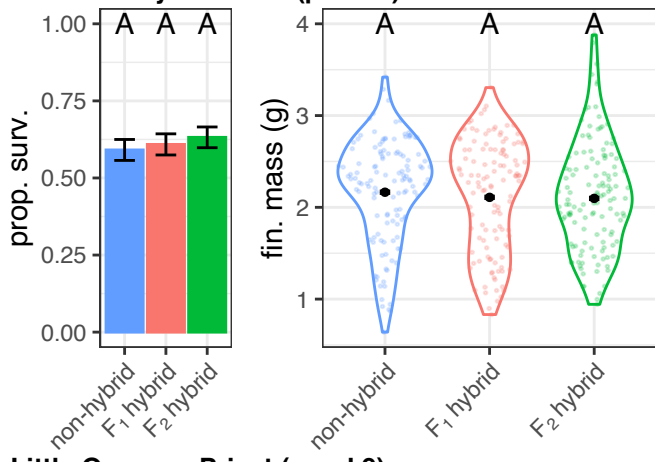


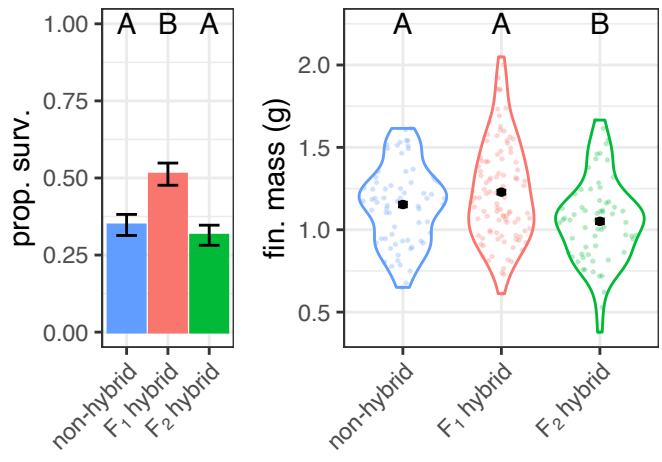
Figure S9: **Comparison of the magnitude of heterosis and breakdown between pond and aquarium data (within ecotype).** Heterosis and breakdown are both estimated as differences, and this analysis is therefore testing for a difference in these differences between the same metric in pond data vs. aquaria. Ranges that do not include 0 (red vertical line) are considered statistically significant.

### benthic

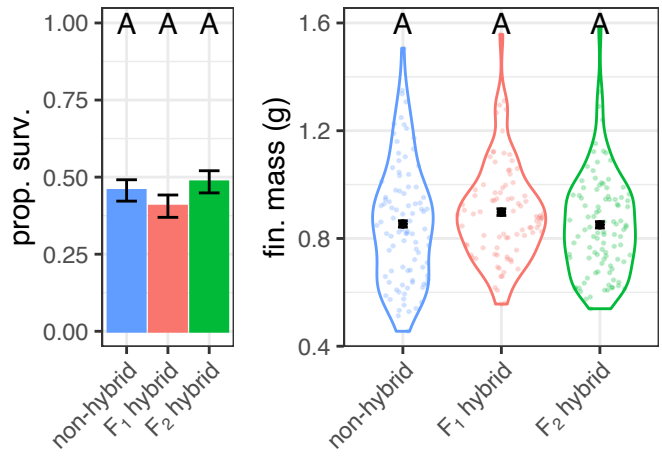
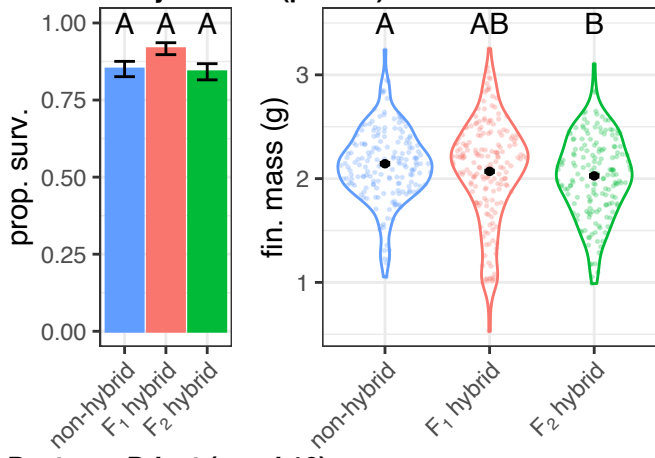
Little Quarry x Paxton (pond 4)



### limnetic



Little Quarry x Priest (pond 9)



Paxton x Priest (pond 19)

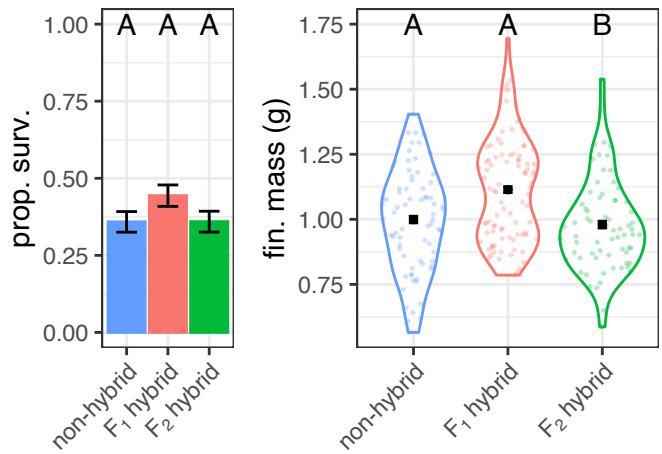
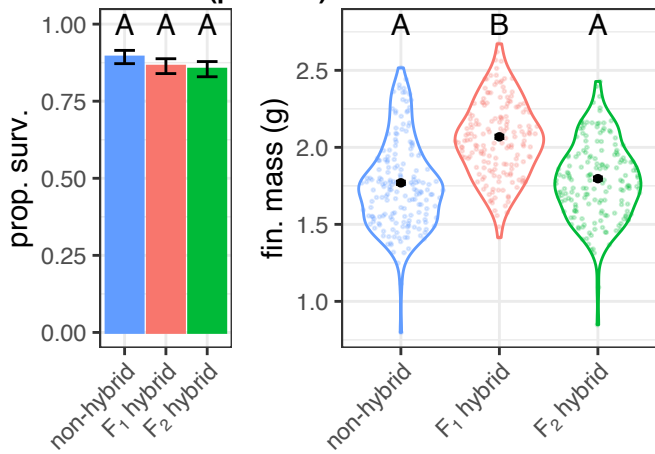


Figure S10: Components of fitness in each experimental pond (rows; see labels) for benthic crosses (left) and limnetic crosses (right). Error bars are  $\pm 1SE$ .

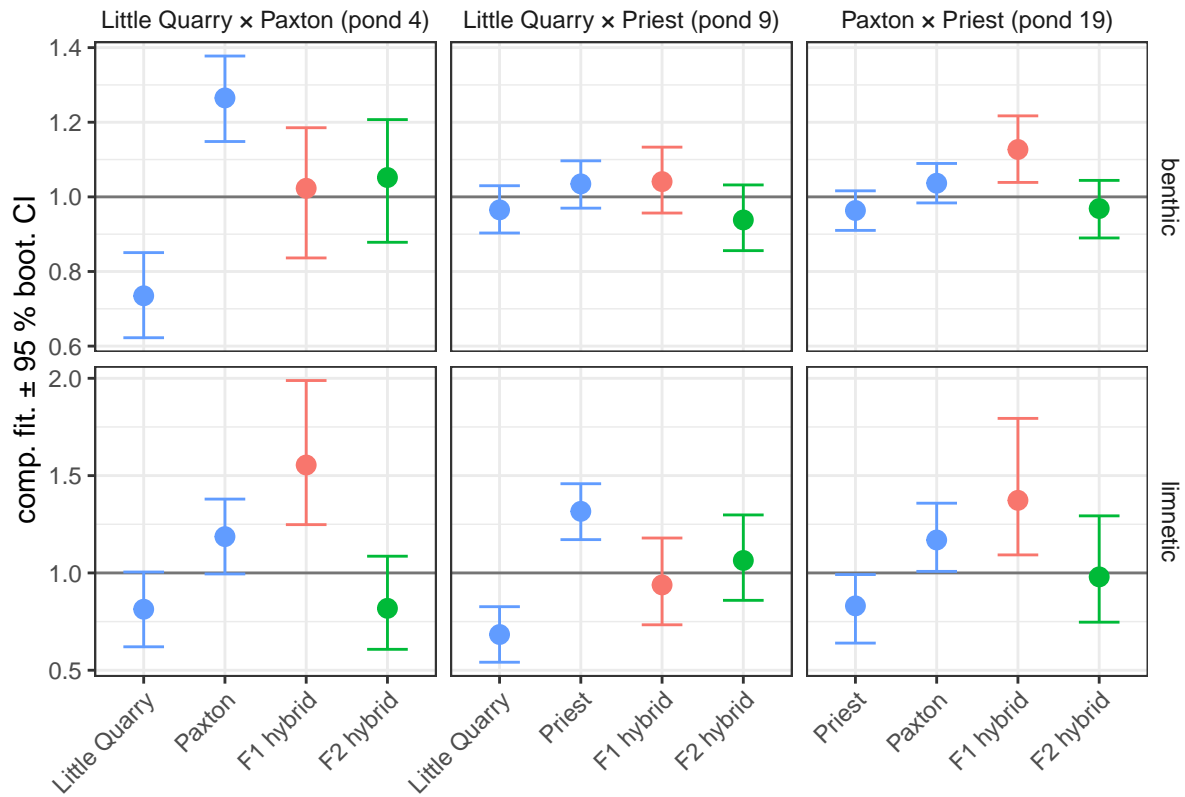


Figure S11: **Fitness components with the two parent populations plotted separately.** Points are estimated marginal means (emmeans package; Lenth et al. 2020), and 95% confidence intervals from 1,000 bootstrap replicates. In all cases where heterosis is observed, the F<sub>1</sub> is either superior to both parents or significantly superior to the less fit parent with an overlapping interval (though higher mean) than the fitter parent.

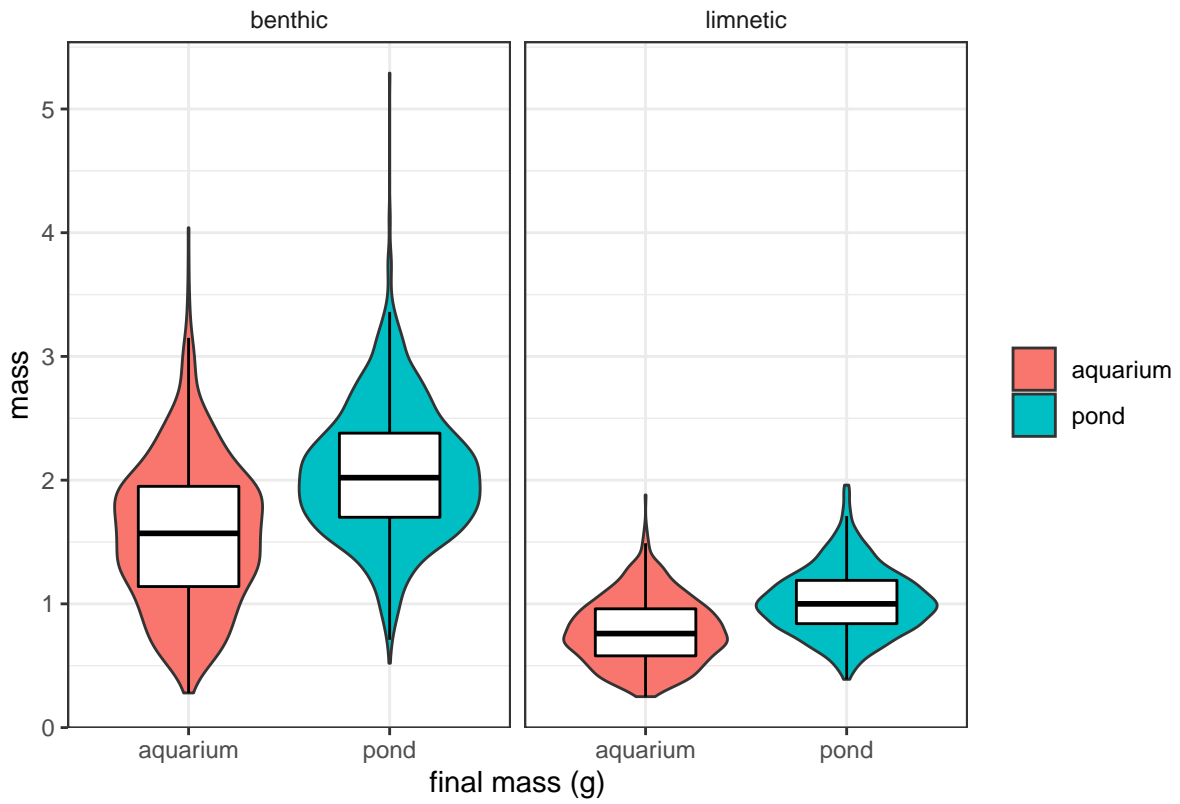


Figure S12: **Final mass (g) of benthic (left) and limnetic (right) fish from aquarium (red) and pond (blue) experiments.** Values are raw data, not from models. Fish from ponds were significantly larger than fish from aquaria for both ecotypes (both  $P < 0.0001$ ).

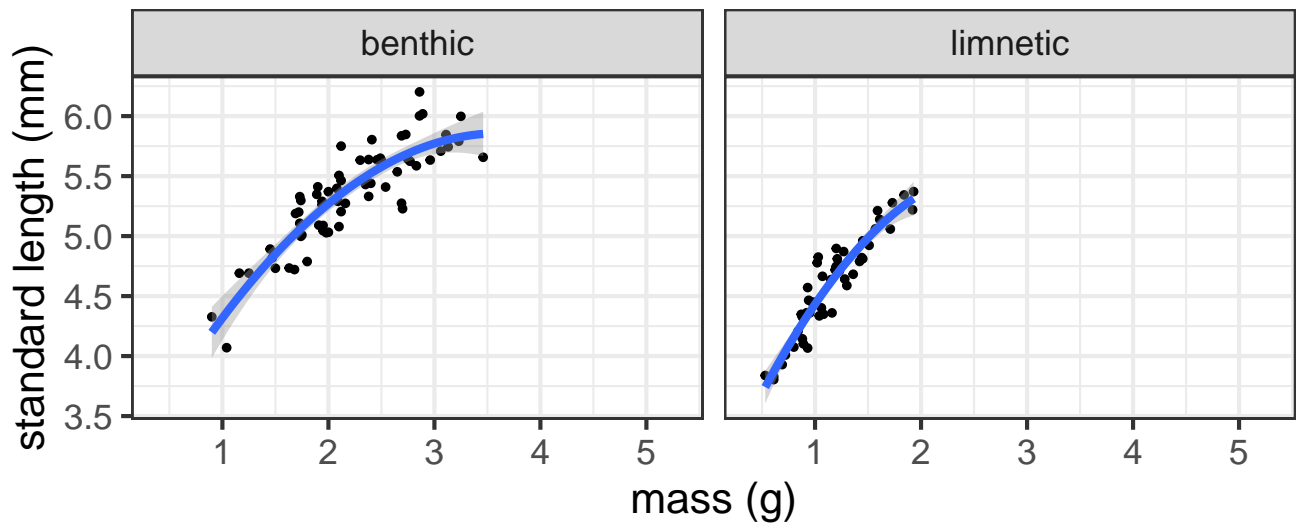


Figure S13: **Relationship between mass and standard length for fish collected from ponds.** Relationships are shown separately for benthic crosses and limnetic crosses. Both are significant and positive. These data were used to estimate standard length across the dataset.

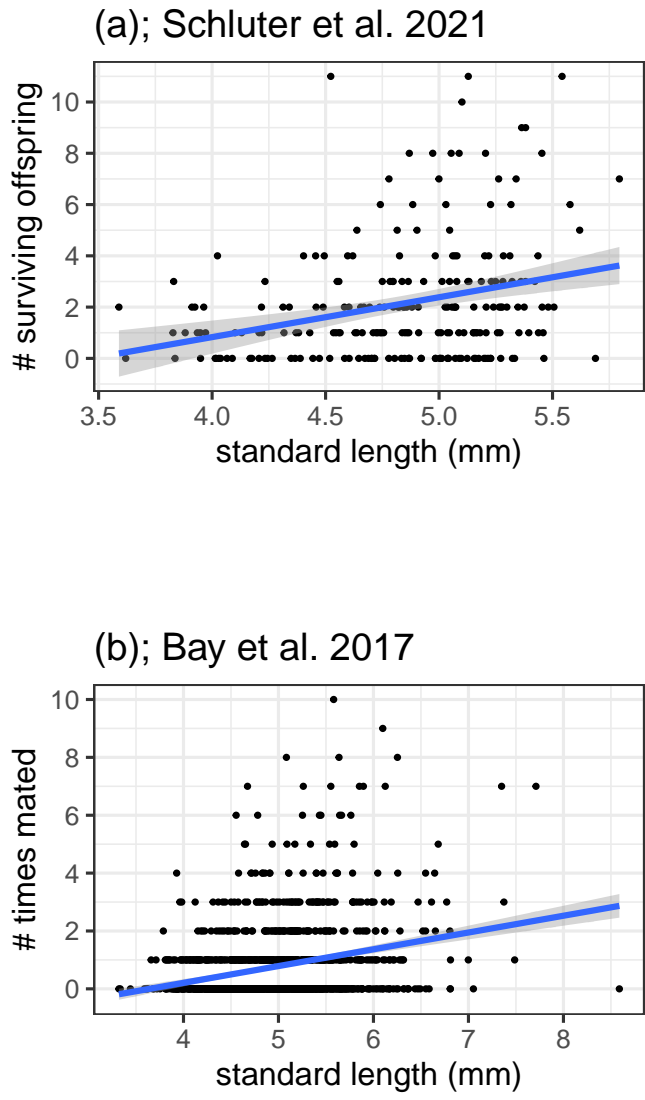


Figure S14: **Relationship between standard length and fecundity in previously published stickleback pond experiments.** Schluter et al. (2021) considered  $F_2$  marine-freshwater hybrid females and measured the number of  $F_3$  hybrids to which they could be confidently assigned parentage. Bay et al. (2017) considered  $F_2$  benthic-limnetic (Paxton) hybrid females and pure males of both ecotypes and recorded the number of successful mating events (inferred from genotyped eggs and offspring). A generalized linear model did not reject the hypothesis that the slope differed by group ( $F_2$  female, pure benthic male, or pure limnetic male), so we plot them together. Both relationships are significant and positive as evaluated with a generalized linear model.