

# Cline coupling and uncoupling in a stickleback hybrid zone

Tim Vines<sup>1,5</sup>, Anne Dalziel<sup>2</sup>, Arianne Albert<sup>3</sup>, Thor Veen<sup>4</sup>, Patricia Schulte<sup>1</sup> and Dolph Schluter<sup>1</sup>

<sup>1</sup>Biodiversity Research Centre, University of British Columbia, 6270 University Blvd  
Vancouver BC, Canada, V6T 1Z4.

<sup>2</sup>Department of Biology, Université Laval, 1030 Avenue de la Médecine, Laval, QC,  
Canada, G1V 0A6

<sup>3</sup>Women's Health Research Institute, 4500 Oak Street, Vancouver, BC, Canada V6H 3N1

<sup>4</sup>Integrative Biology, University of Texas at Austin, One University Station C0990,  
Austin, TX 78712, USA

<sup>5</sup>Molecular Ecology Editorial Office, 6270 University Blvd Vancouver BC, Canada, V6T  
1Z4.

Running Head: Cline coupling in a hybrid zone

Keywords: *Gasterosteus aculeatus*, reproductive isolation, hybridization, Ectodysplasin,  
linkage disequilibrium

Data archiving location: all data will be archived on Dryad

## Abstract

Strong ecological selection on a genetic locus can maintain allele frequency differences between populations in different environments, even in the face of hybridization. When alleles at divergent loci come into tight linkage disequilibria, selection acts on them as a unit and can significantly reduce gene flow. For populations interbreeding across a hybrid zone, linkage disequilibria between loci can force clines to share the same slopes and centers. However, strong ecological selection can push clines away from the others, reducing linkage disequilibria and weakening the barrier to gene flow. We looked for this ‘cline uncoupling’ effect in a hybrid zone between stream resident and anadromous sticklebacks at two genes known to be under divergent natural selection (*Eda* and *ATP1a1*) and five morphological traits that repeatedly evolve in freshwater stickleback. We used 10 anonymous SNPs to characterize the shape of the zone. We found that the clines at *Eda*, *ATP1a1*, and four morphological traits were concordant and coincident, suggesting that direct selection on each is outweighed by the indirect selection generated by linkage disequilibria. Interestingly, the cline for pectoral fin length was much steeper and displaced 200m downstream, and two anonymous SNPs also had steep clines.

## Introduction

Progress towards speciation can be hindered by hybridization because gene flow erodes allele frequency differences between the populations, and because recombination in hybrids breaks down linkage disequilibria. In general, only alleles under strong selection themselves or close to strongly selected loci are expected to show large allele frequency differences between hybridizing populations (Yeaman & Otto 2011, Yeaman & Whitlock 2011). When strongly selected alleles at different loci come into linkage disequilibrium (i.e. are found together more often than predicted by their respective frequencies), selection tends to act on them as unit, reducing hybrid or immigrant fitness by their combined effects (Bazykin 1969, Barton 1983, Kruuk et al 1999, Barton & de Cara 2009), which helps to maintain divergence between the hybridizing populations. The key processes in achieving further divergence between hybridizing populations are thus a) having increasing numbers of loci with large allele frequency differences between the populations, and b) building stronger linkage disequilibria between the alleles at the diverged loci (Felsenstein 1981, Barton & de Cara 2009, Smadja & Butlin 2011, Flaxman et al. 2013, 2014).

One situation that is particularly informative about the interplay between divergent selection and linkage disequilibria (LD) during the process of speciation is a hybrid zone, where parapatric populations interbreed across a shared boundary (Barton & Hewitt 1985, Harrison 1990). Multiple generations of hybridization between adjacent populations typically produce allele frequency clines at the differentiated loci. The slope

and position of each cline is a product of the balance between dispersal and total selection at that locus. Total selection is given by direct selection on the locus plus the sum of indirect selection on diverged loci elsewhere in the genome, weighted by the strength of LD between the focal locus and the other diverged loci (Barton 1983, Barton & Gale 1993). For loci where direct selection is outweighed by indirect selection, all of them will experience approximately the same selection regime, and their clines will cluster together. The clustering of clines in turn increases LD, as ‘pure’ individuals will have alleles characteristic of one or other population at the selected loci, which in turn increases indirect selection (Slatkin 1975, Barton 1983). Clines brought together by strong indirect selection thus tend to be steeper than expected in the zone center (where LD is strongest), and the clines appear to be ‘stepped’ (Szymura & Barton 1991).

By contrast, when direct selection on a locus is stronger than indirect selection, cline shape will reflect the regime imposed by selection on that locus alone. If the direct selection has an ecological component (e.g. the alleles are each favoured on either side of an environmental transition), the cline may be pushed away from the clines at other loci. This scenario is especially likely when the hybrid zone is centered on an ecotone, but the transition points for a subset of environmental variables are displaced away from the boundary between the two environments.

Strong direct selection thus plays two potentially opposing roles: it ensures that allele frequency differences persist despite gene flow (Yeaman & Otto 2011), but it may also push clines apart, reducing linkage disequilibria and thereby weakening the total indirect

selection felt by the remaining loci (Nürnberg et al. 1995). Evidence for this ‘cline uncoupling’ effect of strong direct selection is obtained when the clines for genetic loci or phenotypic traits known to be under selection are nearby but do not share cline centers. Conversely, finding that clines for loci under varying types of selection are centered together suggests that direct selection on each locus has been overwhelmed by indirect selection. Further evidence comes from the cline shapes: concordant, stepped clines suggest that indirect selection predominates, while finding a steeper or shallower unstepped cline for a particular locus would indicate that its shape is determined by direct selection alone.

Genetic divergence under gene flow has been well studied, both theoretically (Wright 1931, Bulmer 1972, Yeaman & Otto 2011, Yeaman & Whitlock 2011) and empirically (reviewed in Smadja & Butlin 2011, Feder et al. 2013). By contrast, the circumstances under which clines come together or move apart are less well studied, either theoretically (with the exception of Slatkin 1975, Nürnberg et al. 1995, Bierne et al 2011) or empirically (see Abbott et al. 2013 for a review).

Here, we look for evidence of cline uncoupling in a natural hybrid zone by examining clines for two loci known to be under selection, five ecologically-relevant morphological traits, and 12 anonymous SNPs in a zone between two differentially adapted threespine stickleback populations. In this hybrid zone, one population (‘stream’) inhabits the upper reaches of a small river, Bonsall Creek, throughout the year. The other (‘anadromous’) spends most of the year in the sea and migrates into the lower half of the river to breed

over the summer (Hagen 1967). The hybrid zone between the two types is located close to the ecotone between the estuary and the freshwater environment. The estuary itself represents a gradient for many environmental variables, including salinity, tidal range, vegetation, and the community of predators, parasites and competitors. These gradients end abruptly with the transition to freshwater. Sticklebacks can easily swim from one side of the zone to the other, so an active preference for the estuary or freshwater part of the river likely plays a role in structuring the hybrid zone. Moreover, lab crosses between anadromous and stream fish from Bonsall Creek show no evidence of hybrid inviability or infertility (Dalziel et al. 2012, Dalziel & Schulte 2012), and it seems likely that the majority of reproductive isolation in the hybrid zone is extrinsic.

The two selected genes we examine are Ectodysplasin (hereafter *Eda*), and  $\text{Na}^+$ ,  $\text{K}^+$  ATPase's catalytic  $\alpha 1$  subunit (hereafter *ATP1a1*). *Eda* is located on chromosome IV and is responsible for one of the major phenotypic differences between these populations: stream fish have 4-8 lateral armour plates per side, whereas the anadromous fish have 30-35 (Hagen 1967). The low plated allele at *Eda* has risen to fixation in many freshwater environments (Colosimo et al. 2005), and shows evidence of repeated selection (Mäkinen et al. 2008, Shimada et al. 2011, DeFaveri et al. 2011, Jones et al. 2012, Raeymaekers et al. 2014). *ATP1a1* is located on chromosome I, and is a multi-subunit, membrane bound enzyme that maintains electrochemical gradients by moving  $\text{K}^+$  ions into and  $\text{Na}^+$  ions out of the cell (reviewed by Kaplan 2002). In fish, this enzyme plays a critical role in osmoregulation in both fresh and salt water (reviewed by Evans et al. 2005). Like *Eda*, the *ATP1a1* isoform has an allele that has risen to high frequency in many freshwater

stickleback populations (Jones et al. 2006, Shimada et al. 2011) and selection is related to salinity (DeFaveri et al. 2011, 2013; Shimada et al. 2011).

All of the morphological traits we examined have repeatedly evolved after freshwater colonization, and are predicted to be under strong selection. Anadromous populations have long pelvic and dorsal spines that likely defend against gape-limited predators, while freshwater populations typically have shorter spines (Bell et al. 1993, Reimchen 1994, Marchinko 2009). Anadromous fish also have large, long fins that are capable of powering prolonged swimming during migration. Their smaller caudal peduncles are predicted to streamline the fish and reduce drag (Dalziel et al. 2012). Stream fish have repeatedly evolved smaller pectoral fins, which may help fish maneuver in smaller spaces, and deeper caudal peduncles that are predicted to increase burst swimming capacity (Taylor & McPhail 1986). The median fins (which includes the dorsal fin) are involved in maneuvering and force generation during steady swimming (Lauder et al. 2002), and also repeatedly shrink in lakes without predatory fish (Walker 1997; Walker and Bell 2000).

The selection regimes for alleles and traits advantageous in either anadromous or stream fish will not necessarily transition in the same part of the river. For example, selection on loci involved in osmoregulation, such as *ATP1a1*, may favor the stream allele where the stream bed salinity is less than isosmotic (~13ppt) and the marine allele at higher salinities (Shimada et al. 2011; DeFaveri et al. 2011). By contrast, stream alleles at loci underlying reductions in lateral plates (*Eda*) and other body armor traits (pelvic and

dorsal spine length) may only be favored in full freshwater (< 1 ppt), where insects become a significant part of the predator regime. Strong direct selection on these loci and traits would force their clines to be centered in different parts of the river. Our goal in this paper was to thus test whether the clines at body armor loci and traits (*Eda*, pelvic and dorsal spine length), osmoregulation loci (*ATP1a1*) and morphological traits related to swimming capacity (pectoral and dorsal fin length, caudal peduncle depth) have different slopes and centers, which would in turn suggest that strong direct selection has uncoupled the clines at these loci. By contrast, finding that these loci and traits have concordant and coincident clines indicates that indirect selection generated by LD has overwhelmed the presumably strong direct selection and forced the clines into having the same shapes. Either way, our data shed light on the relative roles of direct and indirect selection in promoting (or inhibiting) progress towards speciation in these species.



## Methods

### Data collection

We collected sticklebacks from Bonsall Creek between 15<sup>th</sup> May and 8<sup>th</sup> June 2006. The creek is a relatively short (15km) and narrow (~3m wide) tidal river on the South Eastern end of Vancouver Island, British Columbia (see Hagen 1967). Anadromous stickleback migrate into Bonsall Creek in late spring (Hagen 1967). The majority remains in the tidal part of the creek (<2350m in Figure 1), but a few move further up into the freshwater stretches. Fish that resemble phenotypically pure anadromous sticklebacks (see Figure 3) are rare beyond 2700m upstream. The stream resident population reaches its highest densities upstream of 3300m, but individuals with stream resident phenotypes are common to around 2300m and can occasionally be found as far downstream as 1650m from the sea. Hybrid fish with intermediate phenotypes can be found where the stream resident and anadromous populations overlap, but are most common between 2200m and 2600m.

We captured fish at 100 to 150m intervals through the hybrid zone, using 5 minnow traps per site (Figure 1). Fish were killed with an overdose of MS222 anesthetic and frozen on dry ice or preserved in ethanol. We used a YSI 85 Handheld Conductivity and Oxygen meter (YSI Inc., Ohio, USA) to measure the surface and creek bed salinity at each site at low tide and at the peak of the highest tide during sampling period (11pm on 14<sup>th</sup> June 2006), and found the limit of the salt finger to occur 2350m from the sea (at 48°52'48.46"N, 123°40'26.93"W).

## **Morphological data**

We selected ten morphological traits that have been found to repeatedly diverge between anadromous and freshwater sticklebacks (Hagen & Gilbertson 1972). Ten marine and ten stream individuals from each side of the hybrid zone were measured for standard length, body depth, head depth, eye diameter, dorsal fin length, caudal peduncle depth, pectoral fin length, caudal peduncle width, left pelvic spine length, and second dorsal spine length. Eye diameter, head depth and body depth showed no consistent differences between marine and stream fish, even when each trait was regressed onto standard length to remove the effects of size, and these traits were discarded. The remaining seven traits were measured on the 428 ethanol preserved fish. We measured standard length and dorsal and pelvic fin length with calipers, the length of the second dorsal and pelvic spine and the caudal peduncle depth and width with an ocular micrometer fitted to a binocular microscope. As standard length showed a strong relationship with all traits, we took the residuals from a regression of each trait on standard length. These residuals were then log transformed to improve their normality.

## **Genetic data**

We extracted DNA from tail fin tissue from all 428 fish using the protocol in Peichel *et al.* (2001); the DNA was resuspended in 50µl of double distilled water and stored at -20°C.

218

219 *Ectodysplasin (Eda)*

220 The *Eda* locus is located on linkage group IV. We chose a T/C SNP at position 421 of an  
221 amplicon spanning the 7<sup>th</sup> and 8<sup>th</sup> exon of the *Eda* gene (Colosimo et al. 2005) for  
222 genotyping. The primers are given in Table S1.

223

224 *Na<sup>+</sup>, K<sup>+</sup> ATPase subunit  $\alpha$ 1a (ATP1a1)*

225 There are two Na<sup>+</sup>, K<sup>+</sup> ATPase  $\alpha$ 1 paralogs in the stickleback genome, which are found in  
226 tandem within approximately 9,000 bp of each other on contigs 7064/7065 and 7066  
227 within Linkage group/Chromosome 1. We studied the isoform on contig 7066, which is  
228 the same isoform studied by Jones et al. (2006) and Barrett et al. (2008). Our initial  
229 genotyping of the anadromous and stream alleles at the 16<sup>th</sup> to 18<sup>th</sup> intron of this gene  
230 followed the procedure in Jones et al. (2006), but as these sites are not variable in our  
231 populations we could not achieve allele-specific amplification. We therefore sequenced  
232 this genomic region in 10 pure stream (above 4000m) and 10 pure anadromous fish (from  
233 1650m). We found a number of other diagnostic SNPs within this region, and selected an  
234 A/G SNP at position 446 of the amplicon (see Table S1 for sequences and primers).

235

236 *Anonymous SNPs*

237 We began with the list of 25 loci used to construct the phylogenetic tree in Colosimo *et*  
238 *al.* (Figure 3C in Colosimo et al. 2005). We excluded loci on the same linkage group as  
239 *Eda* (LG 4) or Na<sup>+</sup>, K<sup>+</sup> ATPase (LG 1), although loci with an unknown location were  
240 retained. We then tested the primers pairs from Colosimo et al. (2005) on ten stream

(4000m) and ten anadromous (1650m) fish from Bonsall Creek. Of the 16 loci that amplified successfully in both populations, we retained 13 loci that had a SNP minor allele frequency greater than 0.1 in at least one of the samples.

These sequences of these 13 loci (Supplementary Table 1) were used by the McGill University and Génome Québec Innovation center to design a custom genotyping assay using Sequenom® iPLEX®Gold Genotyping Technology. Genomic DNA from all 428 fish was quantified and diluted to 20 ng/μl and at least 30μl of each sample were sent on dry ice to the McGill University and Génome Québec Innovation center for genotyping.

## Statistical methods

### *Cline fitting: genetic loci*

The clines at *Eda* and *ATP1a1* and the anonymous SNP loci were characterized with the program CFit version 7 (Gay et al. 2008). This program uses a Metropolis algorithm to find the cline shape that best fits the available data. The program fits two basic parameters that describe a sigmoid cline, cline slope  $s$  and cline center  $c$ , and also allows for deviations from Hardy-Weinberg equilibrium (fitted as  $F_{IS}$ ) within each site. Initial tests found no support for including  $F_{IS}$  at any of the genetic loci and we do not consider it further.

We also fitted a more complex stepped cline model, where the center part of the cline remains sigmoid, but exponential curves are permitted for each of the tails. This model

has six parameters: slope  $s$ , center  $c$ , two locations where the exponential tails begin ( $xpos1$ ,  $xpos2$ ) and the two slopes of the exponential curves ( $tslope1$  and  $tslope2$ ); the latter two range between 0 and 1 and are the proportion by which the slope of the sigmoid curve is reduced. Note that  $xpos2$  and  $tslope2$  are the parameters for the anadromous side of the hybrid zone.

#### *Cline fitting: morphological data*

CFit requires morphological traits to be at their maximum at the left hand side (i.e. anadromous) side of the hybrid zone. This is true for the logged residuals of dorsal fin length, left pectoral fin length, pelvic spine length, and second dorsal spine length. However, the logged residual of caudal peduncle depth is greatest in stream fish, and so we used the negative of the logged residuals in the CFit analysis for this trait.

We used a simple clinal model to characterize the morphological data across the hybrid zone (Gay et al. 2008). Observed measurements were assumed to be drawn from a normal distribution  $N(\mu_x, \sigma_x)$ , where  $\mu_x$  and  $\sigma_x$  are functions of distance  $x$  across the hybrid zone. More specifically,  $\mu_x$  was modeled as  $\mu_{\min} + (\mu_{\max} - \mu_{\min}) p_x$ , where  $\mu_{\min}$  and  $\mu_{\max}$  measure the mean in the stream and anadromous population, respectively, and  $p_x$  is the sigmoid function of distance (with maximum slope  $s$  and center  $c$ ). The variance  $\sigma_x$  was modeled as  $p_x^2 \sigma_1 + 2 p_x(1 - p_x) \sigma_2 + (1 - p_x)^2 \sigma_3$  to account for differences among individuals from either the anadromous or the stream populations ( $\sigma_1$  and  $\sigma_3$ , respectively) and between either pure population and hybrids from the zone center ( $\sigma_2$ ). This model assumes that in each location, the distribution of each trait is unimodal

(drawn from a simple normal distribution). As with the genetic data, parameters were estimated by maximum likelihood using CFit 7.

For each trait, we compared the fit when the trait variance is constant through the zone ( $\sigma_1 = \sigma_2 = \sigma_3$ ) to a model where  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$  were allowed to vary independently. The best fit from this comparison (as determined by comparing the Akaike Information Criteria, Akaike 1973) was then retained for the cline comparison step described below. As with the genetic data, we fitted the stepped cline to the morphological data as well, leading to a model with a maximum of eleven independent parameters.

#### *Comparing cline shapes*

We tested for coincidence and concordance between clines by forcing them to have the same center and/or slope, respectively, and comparing the AIC values of these models to models where each cline was fitted independently. We considered a difference of 2 AIC units between models an indication of a ‘significant’ difference, with the caveat that a difference of 7 to 10 AIC units is needed before concluding that the worse model has essentially no support (Burnham and Anderson 2002).

#### *Cline Analysis with the Hybrid Index*

When genetic loci are more or less fixed on either side of the zone they can be treated as diagnostic, and can thus be used to calculate a Hybrid Index (HI) for each individual. The Hybrid Index is calculated by counting the number of anadromous alleles across the four loci with large differences in allele frequency between the two populations (*Eda*,

310 *ATP1a1*, P3D05, and P6A10, see results). We can use the variance in HI in the center of  
 311 the zone to estimate linkage disequilibria ( $\bar{D}$ ) and hence the dispersal rate  $\sigma$  (p23 of  
 312 Barton & Gale 1993). Since cline width is a product of the relationship between dispersal  
 313 and selection, we can use this information to estimate the strength of selection acting on  
 314 the genetic loci and the morphological traits.

315

316

## Results

### *Sample sizes*

We genotyped 428 adult (>40mm) fish from 16 sites through the hybrid zone at *Eda*, *ATP1a1*, and 13 anonymous SNP loci. These data are available on Dryad (doi:XXXX). Locus P7A09 failed to amplify in most anadromous individuals and was excluded from the subsequent analyses. Amplification success was otherwise very good, with an average of 9.7 (2.2%) missing genotypes per locus. Two of the anonymous loci (P3A06 and P6B12) showed no allele frequency differences across the hybrid zone, and were also excluded at this stage.

We also obtained morphological data for these 428 fish (Table 1 and Dryad doi:XXXX). We discovered that variation in caudal peduncle width was found to result from two confounding sources: anadromous fish had a bony ‘keel’ which adds 1-2mm to the total width of the peduncle, but this feature is not present in stream individuals. However, when the keel is ignored, stream fish had a wider caudal peduncle than anadromous fish. As the total width of the peduncle could not be expected to fit a simple cline model it was discarded at this stage.

### *Comparing cline centers and slopes*

Our central question is whether the clines for *Eda*, *ATP1a1*, and the five morphological traits have the same center and slope. We therefore fit a model where these seven clines



were constrained to share the same slope and center, and compared it to one where slope and center were allowed to vary independently. A joint center (2298m) and slope (-2.83) was a better fit ( $AIC = 4206.9$  for joint clines vs  $4212.8$  for independent clines). The parameters for the individually fitted clines at all genetic loci are given in Table 2a.

The cline for pectoral fin length is quite different from the other morphological traits (see above, Figure 3c and S1), and correspondingly the model fit is somewhat improved when it is allowed to have an independent slope and center ( $AIC = 4201.4$  when pectoral fin is independent vs  $AIC = 4206.9$  when it is not). The joint cline for these six traits and loci was at 2296m and had a slope of -2.87.

The clines for six of the seven selected traits and loci are thus centered together in the same part of the river, about 50m downstream from the upper limit of the salt finger at 2350m (Figure 4). We next tested whether these six clines share the same slope. We fit a model where these six clines are constrained to share the same center, but slope is allowed to vary independently. For comparison with the models above, the cline for pectoral fin length was included but had its own slope and center. This model was not an improvement over the model where all six clines shared the same center and slope ( $AIC = 4201.4$  for joint slopes vs  $4207.6$  for independent slopes). As mentioned in the introduction, it is possible indirect selection is responsible for the shape of most clines, but individual traits or loci may be under strong direct selection and therefore much steeper. We therefore fit a series of model where five of the clines had the same slope and one cline had an independent slope, with all clines sharing the same center. None of these

models was a significant improvement over the simpler model where all six clines are constrained to share the same slope and center.

The clines for two of the anonymous SNPs (P3D05 and P6A10) are also steep and centered near the joint cline for the selected traits and loci (Table 2a, Figure 4). As a *post hoc* test we fit models where each of these was forced to share the same slope and center as the joint cline. We found some evidence that P3D05 could be included with the joint cline (AIC = 4090.1 when P3D05 is included vs AIC = 4091.5 when P3D05 has its own slope and center). However, the model fit was worse when P6A10 was included with the joint cline (AIC = 4074.26 when P6A10 is included vs AIC = 4062.3 when P6A10 has independent slope and center). This is largely because the cline at P6A10 is located 100m upstream of the joint cline (at 2401m): forcing P6A10 to have the same center as the joint cline is a significantly worse model fit than forcing it to have the same slope (AIC = 4071.3 when P6A10 has joint center but independent slope vs AIC = 4062.8 when it has joint slope but independent center).

### *Fitting stepped clines*

There was a worsening in model fit when we fitted a stepped cline for both *Eda* and *ATP1a1* relative to the fit of the basic sigmoid cline (*Eda*: AIC=681.6 for sigmoid vs 685.5 for stepped; *ATP1a1*: 713.6 for sigmoid vs 721.36 for stepped). Similarly, a stepped cline model was a significantly worse fit for dorsal fin, peduncle depth, left

pelvic spine or 2<sup>nd</sup> dorsal spine. Pectoral fin length had a much steeper slope (-14.5) and a cline center at 2170m, about 130m downstream of the other traits (Table 2b and Figure 3c). The distribution of pectoral fin phenotypes in the hybrid zone was also quite different from the other four traits. First, fish from the furthest point upstream had proportionally much smaller fins than any of the other sites in the hybrid zone. Second, fish with proportionally very long pectoral fins (as is typical of anadromous fish) were very rare upstream of 2100m (see Figure 3c). We therefore used CFit7 to assess the fit of more complex cline models (see Gay et al. 2008 and the CFit7 manual for details). The best-fit model required 18 parameters and is shown in Figure S1.

Fitting a stepped cline led to an improvement in model fit for three of five anonymous SNP loci but the fit was hardly improved for the other two. A stepped cline with  $tslope1 = tslope2$  was a better fit for P6A10 (AIC = 674.4 for stepped vs 683.2 for sigmoid). However the improvement in fit of a stepped cline over a simple sigmoid cline is slight for P3D05, the other SNP having a steep cline (AIC = 711.4 for stepped vs 712.3 for sigmoid). Of the SNPs with shallower clines, a stepped cline was strongly supported for P7G05 and P9D09 (809.4 vs 815.4, and 816.7 vs 833.3, respectively) but not for P7C08 (AIC=761.6 for stepped vs 762.0 for sigmoid).

#### *Estimating dispersal and selection*

The four genetic loci with steep clines (*Eda*, *ATP1a1*, P3D05, and P6A10) are more or less fixed on either side of the zone and can be treated as diagnostic. We used these four

loci to calculate a Hybrid Index (HI), as described in Barton & Gale (1993; hereafter BG93). The distribution of the HI through the hybrid zone is shown in Figure 5, while the mean and variance for each sampling location are given in Table 1.

The variance in HI in the center of the zone was used to estimate mean linkage disequilibrium ( $\bar{D}$ ); this parameter can then be used to estimate the dispersal rate. A portion of the variance in HI in the zone center arises from the variance in allele frequency at the individual loci ( $\text{var}(p)$  in BG93), and the remainder is due to migration bringing in purer genotypes. Taking the 64 individuals sampled at 2300m and 2400m (between which the clines at these four loci are roughly centered), we estimate the variance in allele frequency to be  $\text{var}(p) = 0.0025$ . From the first term of equation 2b of BG93, the proportion of  $\bar{D}$  due to  $\text{var}(p)$  is 0.0309. The mean HI in this sample is  $\bar{z} = 0.515$ , and the variance in HI is  $\text{var}(z) = 0.087$ . The remaining variance ( $0.0875 - 0.0309 = 0.0566$ ) leads to the calculation  $\bar{D} = 0.151$ , with 95% confidence intervals (using  $F_{67, \text{Infinity}}$ ) of 0.112 to 0.228.

Next, we convert the jointly fitted slope of these four clines (-2.86) to cline width via  $4/\text{abs}(-2.86) = 1.39\text{km}$  (see Gay et al. 2008). We can use the estimates of  $\bar{D}$  and cline width to estimate effective dispersal within the hybrid zone, as measured by the standard deviation  $\sigma$  of the distance between where a fish is hatched and where it reproduces, using the equation  $\sigma = \sqrt{r\bar{D}w^2/(1+r)}$  (p24 of BG93). Our samples were restricted to fish over 40mm, which were very likely sampled after they had dispersed. None of these four loci are physically linked, so  $r = 0.5$ , and thus  $\sigma = 0.25 \text{ km gen}^{-1/2}$ .

432

433 Finally, we can use dispersal and cline width to estimate  $s^*$ , the difference in mean  
 434 fitness at that locus or trait between populations at the center of the zone and those at the  
 435 edge, using  $s^* = (1.732\sigma/w)^2$  (p16 of BG93). We estimated  $s^* = 0.097$  for the joint  
 436 cline at *Eda*, *ATP1a1*, dorsal fin, 2<sup>nd</sup> dorsal spine, left pelvic spine and peduncle depth.

437

438

439

## Discussion

In this paper we looked for evidence that strong direct selection could uncouple clines for loci and traits assumed to be under divergent ecological selection, thereby reducing the linkage disequilibria between them and impeding progress towards speciation. The clines for *Eda*, *ATP1a1*, dorsal fin length, caudal peduncle depth, pelvic spine length, and second dorsal spine length were found to be concordant and coincident. The cline for pectoral fin length was steeper and located 200m downstream from the other morphological traits. Interestingly, two anonymous SNPs (P3D05 and P6A10) also had steep clines located close to the joint cline for the selected traits and loci. The remaining eight SNPs had shallower clines with a wide range of slopes and centers (Table 2).

*Why do some clines have the same slope and centers?*

The concordance and coincidence of the clines at *Eda*, *ATP1a1*, dorsal fin length, second dorsal spine length, pelvic spine length, and caudal peduncle depth suggests that these traits and loci are experiencing roughly the same selection regime, with similar intensities of direct selection. It also suggests that the position along the cline at which selection switches from favoring the marine allele to favoring the freshwater allele is also very similar. However, it is difficult to explain why they would all experience similar direct selection, since their functions are diverse and include body armor, osmoregulation and swimming capacity. An alternative explanation for the equivalent clines among such diverse traits and genes is that indirect selection is strong and similar on all of them

because of strong linkage disequilibria (LD) between the loci and traits that pushes the clines to have the same slopes and centers (Barton 1983, Szymura & Barton 1991).

We can get an approximate estimate of the relative strengths of direct and total (i.e. direct + indirect) selection acting on a locus by fitting a stepped cline. In the center of the zone, LD between selected loci generates indirect selection on all of them together, steepening each of the clines (Szymura & Barton 1991). At the same time, an allele that has introgressed from the population on one side of the hybrid zone to the other side of the zone is at low frequency in the tail of the cline and is thus unlikely to be in LD there with introgressing alleles at other loci. The rate at which a focal allele moves further into the other population depends mostly on the relationship between dispersal and selection acting directly on that locus (and linked regions of the surrounding chromosome; Szymura & Barton 1991, Barton & Gale 1993, Baird 1995). Loci that are experiencing strong total selection but weak direct selection should therefore have a stepped cline, with a steep section in the center of the zone (where there is more LD) and shallower clines on either side.

The model fit for a stepped cline at *Eda* and *ATP1a1* was worse than the fit for a simple sigmoid cline. Similarly, a stepped cline was a worse fit for the four morphological traits in the joint cline (dorsal fin length, caudal peduncle depth, pelvic spine length, and second dorsal spine length). This is a puzzling finding. These clines are all steep compared to the estimated rate of dispersal, and highly concordant and coincident (Figures 2 and 3). There certainly is plenty of LD at the cline center: at 2300m, the

average  $D$  between *Eda*, *ATP1a1*, and P3D05 is 0.10, and  $D'$  is 0.46, indicating that LD is 46% of its maximum possible value in this site (estimated with the R package ‘genetics’: R Core Team, 2014, Gregory Warnes et al., unpublished). Indirect selection is therefore expected to be acting in the center of the hybrid zone and not at the edges, and these six clines should be stepped.

One alternative explanation is that our assumption above is wrong: despite the dissimilarities in function, these loci and traits are actually experiencing similar amounts of direct selection ( $s^* \sim 0.1$ , see last section of Results), and their selection regimes all transition between favoring anadromous and freshwater alleles in the same part of the river (close to the estuary-freshwater ecotone). Linkage disequilibria between the loci generate indirect selection, but perhaps this additional indirect selection is only strong enough to bring the clines into full concordance and coincidence, and not strong enough to generate stepped clines. Modeling work on the behavior of strongly selected loci in hybrid zones would be a useful way to examine this scenario.

*Why do some clines have different centers and slopes?*

We found that two anonymous SNPs had steep clines. The cline at P3D05 was very similar to the joint cline described above (centered at 2330m with a slope -2.59. The cline for P6A10 was slightly steeper (-3.3) and displaced about 100m upstream (at 2402m) from the joint cline, which is comparable to the dispersal distance (125 m). P3D05 is located at 10349kb along LGIII and is embedded within a gene called “amyloid beta (A4)



precursor protein-binding, family B, member 1 interacting protein”. P6A10 is at 12601kb on LGVII in a gene called “mitochondrial calcium uptake 2”. Neither gene has been explicitly connected to freshwater adaptation in sticklebacks. Jones et al. (2012) did not find significant anadromous-freshwater divergence in the vicinity of either SNP in their genome scan. Hohenlohe et al. (2010) saw no significant divergence in that part of LGIII, although they did highlight five divergent SNPs (at 12212041, 12505622, 12525947, 12801733, and 12987907) in the vicinity of P6A10 on LGVII (Figure 8a in Hohenlohe et al. 2010, and P. Hohenlohe, pers. comm).

Interestingly, there is support for a stepped cline at P6A10 (Table 2a and Figure 2d). The most plausible explanation is that it is experiencing both some direct selection ( $s_e = \text{tslope} \times s^* = 0.11$ , see Szymura & Barton 1991) on either side of the hybrid zone and a combination of direct and indirect selection in the center ( $s^* = 0.34$ ). The source of this indirect selection must be linkage disequilibria with other selected loci, either on the same chromosome or elsewhere in the genome.

#### *The non-concordant cline for pectoral fin length*

The cline for pectoral fin length is unlike any of the other morphological traits or genetic loci. The cline is centered between 2100 and 2200 m, a few hundred meters downstream of the other clines, and it is much steeper (-14.5 compared with c. -3 for the other morphological traits, Table 2b), with the caveat that the complicated distribution of phenotypes makes estimating cline shape quite difficult (Figure 3c and Figure S1). The

cline position seems to arise because very few fish with relatively longer fins were sampled further upstream than 2100m. There is a deep (3m) pool at 2100m and the streambed salinity here is 20ppt at all points of the tidal cycle. So, even though it is clearly possible for fish with long pectoral fins to swim into the upper parts of the estuary, and many fish with otherwise anadromous morphology are found upstream of 2100m (Figure 3a, c-e), it appears that long finned fish restrict themselves to the lower, more brackish, parts of the creek.

Why might pectoral fin length show a different cline center than other traits? It is possible that shorter pectoral fins are highly deleterious in migrating anadromous fish, such that returning anadromous sticklebacks are all long-finned. If anadromous fish only migrate into the lower part of the estuary then the cline for pectoral fin length will coincide with that limit. This scenario implies that there is a link between the phenotype of a fish and its preferred location in the river (see below). Fish with otherwise anadromous morphology found further upstream might remain in the estuary year round, or overwinter in the sea close to the mouth of the creek. While we do not currently know the migratory limit in the estuary for the anadromous population, it would be possible to use stable isotope ratios to distinguish migratory from non-migratory fish (e.g. Arai et al. 2003).

An additional explanation for the downstream position of the pectoral fin cline is that the stream alleles for the underlying loci are dominant, such that a genetically intermediate hybrid would have the short pectoral fin phenotype. However, lab crossed F1 hybrids

between stream and anadromous fish from Bonsall Creek showed no net dominance for pectoral fin area (Figure 3B in Dalziel et al 2012).

### *The structure of the Bonsall Creek hybrid zone*

One puzzling feature of this hybrid zone is that the observed clines are generally very steep compared to the movement capability of the fish. A mark-recapture study conducted on 23<sup>rd</sup> and 24<sup>th</sup> May 2006 (T. Vines & A. Dalziel, unpublished data) found several fish that had moved several hundred meters overnight. If fish move tens of meters per day in a random direction, selection would need to be extremely strong to maintain these steep clines: almost all fish crossing the hybrid zone would need to be eliminated before they could be sampled. It is difficult to imagine any selective force that could accomplish this. Adult sticklebacks are able to cope with a wide range of salinities without suffering large effects (e.g. Schaarschmidt et al. 1999, Barrett et al. 2009), so the salinity differences between the estuary and freshwater cannot be a big enough source of mortality. Moreover, piscivorous birds and fish are present throughout the hybrid zone, it seems very unlikely that their selective impacts would be restricted to just one side of the zone. The density of predators would also need to be very high to take out the constant flux of immigrants.

A more reasonable explanation is that while fish could easily traverse the zone, they instead choose a point along the river according to their genotype and phenotype and stay

there over the breeding season. This must be the case for the anadromous fish, which spend the winter in the sea and migrate into the river in the spring (Hagen 1967). The Bonsall Creek hybrid zone is thus similar to hybrid zones between migratory and resident bird species (e.g. Rohwer et al. 2001, Brelsford & Irwin 2009) where one species migrates long distances over the winter and the zone is only reformed in the breeding season. The key parameter in these zones is not the total distance travelled (which is orders of magnitude greater than the zone width), and instead is the distance between where the organism is born and where it reproduces. The standard deviation in the latter distance is estimated by  $\sigma = 0.25 \text{ km gen}^{1/2}$  above: in this hybrid zone, an individual reproduces an average of 125m upstream or downstream of where it was born.

The steep concordant clines in Bonsall Creek perhaps arise because breeding location is determined by a continuous mating habitat preference, and the preference loci are in tight linkage disequilibria with the traits and loci we examined here. In this scenario, pure anadromous fish prefer to nest in the saltiest part of the river near the sea, and fish with an increasing proportion of stream alleles at the preference loci are found closer and closer to freshwater. The strong LD between the alleles underlying habitat preference and the stream or anadromous alleles for other morphological or genetic clines should be maintained by strong total selection. Alternatively, there may be no segregating variance for habitat preference, and the fish may follow a simple rule like ‘remain in the part of the river where the effort to maintain osmotic balance and swimming effort is minimized’.

What is the source of this selection? Unlike adults, eggs and newly hatched offspring cannot change their location within the river, and must therefore experience the changing salinities associated with the tidal cycle. The effects of salinity on egg hatching success and later fish growth in stickleback appear to be significant. For example, the hatching success of eggs from Belgian freshwater populations is better in low than high salinity, while the converse is true for marine populations (Heuts 1947). More recent studies have found low hatching success and growth rate of freshwater fish in higher salinities (Marchinko & Schluter, 2007; DeFaveri and Merila 2014).

Interestingly, some hybrid zones between anadromous and stream sticklebacks are found exclusively in fresh water, such as on tributaries of the nearby Fraser River (Virgil & McPhail 1994, Dalziel et al. 2012). Since all of these hybrid zones are similarly narrow, it seems unlikely that adaptation to different salinities in the early life stages is the major source of selection against hybrids or against the offspring of fish breeding in the ‘wrong’ part of the river. Instead, as hypothesized by Hagen (1967), it seems plausible that the majority of selection occurs over the winter, and is largely against hybrids that are unfit in both the stream or the sea. Because linkage disequilibria are generally quite strong in this hybrid zone, hybrids will tend to have many features that make them poorly suited to surviving in either environment. For example, a hybrid that remains in the stream over the winter will need to cope with the combination of low temperatures and fresh water (which in combination can be challenging for anadromous fish, see Heuts 1947, Schaarschmidt et al. 1999), and the need to efficiently maintain position in the stream. It will also need to compete effectively with pure stream sticklebacks for (presumably)

limited food resources. By contrast, hybrid fish in the marine environment will need to cope with the higher density of vertebrate predators, and compete for food with the pure anadromous fish. Studies with sympatric stickleback species pairs have shown that selection against hybrids via competition for food can be strong (Rundle 2002), particularly when coupled with predation (Vamosi & Schluter 2002, Rundle et al. 2003). It is also possible that some hybrids do not choose between the stream or marine environment at all, and overwinter in the more intermediate environment of the estuary.

## Conclusions

This study looked for evidence that direct selection could uncouple clines at loci known to be under ecological selection. We examined two well-studied genes in a hybrid zone between stream and anadromous sticklebacks, and found that the clines at *Eda* and *ATP1a1* shared the same slope and center, as did two other anonymous SNPs and four morphological traits. The concordance and coincidence of these clines suggests that indirect selection is the major force in determining their shape and location, and that direct selection on the individual genes and traits has not been sufficient to uncouple them. The cline for pectoral fin length was much steeper and centered 200m downstream, perhaps because long-finned anadromous fish do not migrate any further into the creek.

Given the importance of pectoral fins for steady locomotion in sticklebacks, pectoral fin length may be under strong selection in migrating anadromous fish.

# Acknowledgments

The authors are grateful to Taylor Gibbons for help in the lab, Jeremy Schmutz for information on the SNP loci, Jessica Hill and Chrissy Spencer for help with sample collection, and to James Thomas and other members of the Halalt first nation for access to their land and information on the history of the creek. We thank Nick Barton, Sean Rogers and Thomas Lenormand for helpful discussions, while Rowan Barrett and Dave Toews provided comments on an earlier version. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Discovery and Discovery Accelerator grants to PMS, Discovery and Special Research Opportunity grants to DS, and a Canada Graduate Scholarship to ACD and AA. THV was supported by a European Union FP6 Outgoing International Fellowship and a real job. TV was supported by the NSERC-CREATE Training Program in Biodiversity Research and NSF grant IOS-1145468.

## References

- Abbott, R., D. Albach, S. Ansell, et al. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246
- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *International symposium on information theory, Akad´emiai Kiado, Budapest, Hungary*. 267–281.
- Arai, T., A. Goto, and N. Miyazaki. 2003. Use of otolith microchemistry to estimate the migratory history of the threespine stickleback, *Gasterosteus aculeatus* *Journal Of The Marine Biological Association Of The United Kingdom* 83: 223–230
- Baird, S. J. E. 1995. A simulation study of multilocus clines. *Evolution* 49: 1038–1045.
- Barrett, R. D. H., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. *Science* 322: 255–257.
- Barrett, R. D. H., T. H. Vines, J. S. Bystriansky, and P. M. Schulte. 2009. Should I stay or should I go? The Ectodysplasin locus is associated with behavioural differences in threespine stickleback. *Biology Letters* 5: 788–791.
- Barton, N. H. 1983. Multilocus clines. *Evolution* 37: 454–471.
- Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.* 16: 113–148.
- Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pages 13–45 in R. G. Harrison, ed. *Hybrid zones and the evolutionary process*. Oxford University Press, Oxford.



685     Barton, N. H., and M. A. R. de Cara. 2009. The evolution of strong reproductive  
686     isolation. *Evolution* 63: 1171–1190.

687     Bazykin, A. D. 1969. Hypothetical mechanism of speciation. *Evolution*, 23 685-687.

688     Bell, M. A., G. Ortí, J. A. Walker, and J. P. Koenings. 1993. Evolution of pelvic  
689     reduction in threespine stickleback fish: A test of competing hypotheses. *Evolution* 47,  
690     906

691     Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why  
692     genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–  
693     2072.

694     Brelsford, A., and D. E. Irwin. 2009. Incipient speciation despite little assortative mating:  
695     the yellow-rumped warbler hybrid zone. *Evolution* 63: 3050–3060.

696     Burnham, K. P. and D. L. Anderson. 2002. Model Selection and Multimodel Inference: A  
697     Practical-Theoretic Approach. Springer-Verlag.

698     Colosimo, P. F., K. E. Hosemann, S. Balabhadra, et al. 2005. Widespread parallel  
699     evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307:  
700     1928–1933.

701     Dalziel, A. C. and P. M. Schulte. 2012. Correlates of prolonged swimming performance  
702     in F2 hybrids of migratory and non-migratory threespine stickleback. *Journal of*  
703     Experimental Biology 215: 3587-3596.

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

707 DeFaveri, J., T. Shikano, Y. Shimada, A. Goto, and J Merilä. 2011. Global analysis of  
708 genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus*  
709 *aculeatus*). *Evolution* 65: 1800–1807.

710 DeFaveri, J., T. Shikano, Y. Shimada, and J. Merila. 2013. High degree of genetic  
711 differentiation in marine three-spined sticklebacks (*Gasterosteus aculeatus*). *Molecular*  
712 *Ecology* 22: 4811–4828.

713 Evans, D. H., P. M. Piermarini, and K. P. Choe. 2005. The multifunctional fish gill:  
714 dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of  
715 nitrogenous waste. *Physiological Reviews* 85: 97-177.

716 Feder, J. L., S. M. Flaxman, S. P. Egan, A. A. Comeault, and P. Nosil. 2013. Geographic  
717 mode of speciation and genomic divergence. *Ann. Rev. Ecol. Evol. Syst* 44: 73–97.

718 Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of  
719 animals? *Evolution* 35: 124–138.

720 Flaxman, S. M., J. L. Feder, and P. Nosil. 2013. Genetic hitchhiking and the dynamic  
721 buildup of genomic divergence during speciation with gene flow. *Evolution* 67: 2577–  
722 2591.

723 Flaxman, S. M., A. C. Wacholder, J. L. Feder, and P Nosil. 2014. Theoretical models of  
724 the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*  
725 23: 4074–4088.

726 Gay, L., P.-A. Crochet, D. A. Bell, and T. Lenormand. 2008. Comparing clines on  
727 molecular and phenotypic traits in hybrid zones: a window on tension zone models.  
728 *Evolution* 62: 2789-2806.

729 Hagen, D. W. 1967. Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). J.  
730 Fish. Res. Board Can., 24: 1637–1692.

731 Hagen, D. W., and L. G. Gilbertson. 1971. Geographic variation and environmental  
732 selection in *Gasterosteus aculeatus* in the Pacific Northwest, America. *Evolution* 26: 32–  
733 51.

734 Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. In: Oxford  
735 Surveys in Evolutionary Biology (eds Futuyma D, Antonovics J), pp. 69–128. Oxford  
736 University Press, New York, NY.

737 Heuts, M. J. 1947. Experimental studies on adaptive evolution in *Gasterosteus aculeatus*  
738 L. *Evolution* 1:89–102.

739 Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko.  
740 2010. Population genomics of parallel adaptation in threespine stickleback using  
741 sequenced RAD tags. *PLoS Genet.* 6:e1000862.

742 Jones, F. C., C. Brown, J. Pemberton, and V. Braithwaite. 2006. Reproductive isolation in  
743 a threespine stickleback hybrid zone. *J. Evol. Biol.* 19, 1531–1544.

744 Jones, F. C., M. G. Grabherr, Y. F. Chan et al. 2012. The genomic basis of adaptive  
745 evolution in threespine sticklebacks. *Nature* 484: 55–61.

746 Kaplan, J. H. 2002. Biochemistry of Na,K-ATPase. *Ann. Rev. Biochem.* 2002: 511-535.

747 Kruuk, L. E., S. J. E. Baird, K. S. Gale, and N. H. Barton. 1999. A comparison of  
748 multilocus clines maintained by environmental adaptation or by selection against hybrids.  
749 *Genetics* 153: 1959–1971.

750 Lauder, G. V., J. C. Nauen, and E. G. Drucker. 2002. Experimental hydrodynamics and  
751 evolution: function of median fins in ray-finned fishes. *Integr. Comp. Biol.* 42: 1009–  
752 1017

753 Mäkinen, H. S., T. Shikano, and J. Merila. 2008. Identifying footprints of directional and  
754 balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus*  
755 *aculeatus*) populations. *Molecular Ecology* 17: 3565–3582.

756 Marchinko, K. B., and D. Schluter. 2007. Parallel evolution by correlated response:  
757 lateral plate reduction in threespine stickleback. *Evolution* 61: 1084–1090.

758 Nürnberger, B., N. H. Barton, C. J. MacCallum, J. Gilchrist, and M. Appleby. 1995.  
759 Natural selection on quantitative traits in a *Bombina* hybrid zone. *Evolution* 49:1224–  
760 1238.

761 Peichel, C. L., K. S. Nereng, K. A. Ohgi et al. 2001. The genetic architecture of  
762 divergence between threespine stickleback species. *Nature* 414: 901–905.

763 R Core Team (2014) R: A language and environment for statistical computing. R  
764 Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

765 Raeymaekers, J. A. M., N. Konijnendijk, M. H. D. Larmuseau, B. Hellemans, L. De  
766 Meester, and F. A. M. Volckaert. 2014. A gene with major phenotypic effects as a target  
767 for selection vs. homogenizing gene flow. *Molecular Ecology* 23: 162–181.

768 Reimchen, T. E. 1994. Predators and evolution in threespine stickleback. in *Evolution of*  
769 *the threespine stickleback* (M.A. Bell and S. A. Foster, eds.), pp 240-273. Oxford  
770 University Press.

771 Rohwer, S., E. Bermingham, and C. Wood. 2001. Plumage and mitochondrial DNA  
772 haplotype variation across a moving hybrid zone. *Evolution* 55: 405–422.

773 Rundle, H. D. 2002. A test of ecologically dependent postmating isolation between  
774 sympatric sticklebacks. *Evolution* 56: 322–329.

775 Rundle, H. D., S. M. Vamosi, and D. Schluter. 2003. Experimental test of predation's  
776 effect on divergent selection during character displacement in sticklebacks. *Proc. Natl.*  
777 *Acad. Sci. USA* 100:14943–14948.

778 Schaarschmidt, T., E. Meyer, and K. Jurss. 1999. A comparison of transport-related gill  
779 enzyme activities and tissue-specific free amino acid concentrations of Baltic Sea  
780 (brackish water) and freshwater threespine sticklebacks, *Gasterosteus aculeatus*, after  
781 salinity and temperature acclimation. *Mar. Biol.* 135:689–697.

782 Shimada, Y., T. Shikano, and J. Merilä. 2011. A high incidence of selection on  
783 physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*.  
784 *Mol. Biol. Evol.* 28:181–193.

785 Slatkin, M. 1975. Gene flow and selection in a two-locus system. *Genetics* 81:787–802.

786 Smadja, C. M., and R. K. Butlin. 2011. A framework for comparing processes of  
787 speciation in the presence of gene flow. *Mol. Ecol.* 20: 5123–5140.

788 Szymura, J. M., and N. H. Barton. 1991. The genetic structure of the hybrid zone  
789 between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between  
790 transects and between loci. *Evolution* 45: 237–261.

791 Taylor, E. B., and J. D. McPhail. 1986. Prolonged and burst swimming in anadromous  
792 and freshwater threespine stickleback, *Gasterosteus aculeatus*. *Can. J. Zool.* 64: 416–420.

793 Vamosi, S. M., and D. Schluter. 2002. Impacts of trout predation on fitness of sympatric  
794 sticklebacks and their hybrids. *Proc. R. Soc. Lond. B Biol. Sci.* 269: 923–930.

795 Virgil, J. A., and J. D. McPhail. 1994. Spatiotemporal distribution of anadromous  
796 (trachurus) and fresh-water (leiurus) threespine sticklebacks, *Gasterosteus-aculeatus*.  
797 *Canadian Field-Naturalist* 108: 355-360.

798 Walker, J. A. 1997. Ecological morphology of lacustrine threespine stickleback  
799 *Gasterosteus aculeatus* L (Gasterosteidae) body shape. *Biol. J. Linnean Soc.* 61:3–50.

800 Walker, J. A., and M. A. Bell. 2000. Net evolutionary trajectories of body shape  
801 evolution within a microgeographic radiation of threespine sticklebacks (*Gasterosteus*  
802 *aculeatus*). *J. Zoology* 252: 293–302

803 Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

804 Yeaman, S., and M. C. Whitlock. 2011. Genetic architecture of adaptation under  
805 migration-selection balance. *Evolution* 65: 1897–1911

806 Yeaman, S., and S. P. Otto. 2011. Establishment and maintenance of adaptive genetic  
807 divergence under migration, selection, and drift. *Evolution* 65: 2123–2129.

808

809

Table 1. Sample sizes, mean allele frequencies for *Eda* and *ATP1a1*, mean and variance for the Hybrid Index (HI) as calculated from *Eda*, *ATP1a1*, P3D05, and P6A10. Also shown are mean body length and the mean logged residuals from the morphological traits, by site. The limit of the salt finger for the highest tide during the study period was 2350m from the sea.

Distance from sea (m)	N	<i>Eda</i> anadromous allele freq	<i>ATP1a1</i> anadromous allele freq	Mean Hybrid Index	Var in Hybrid Index	Body length (mm)	Dorsal fin	Peduncle depth	Pectoral fin	L Pelvic spine	2 <sup>nd</sup> dorsal spine
1650	28	0.80	0.83	0.81	0.04	51.0	0.58	-0.10	0.31	0.79	0.60
1800	22	0.81	0.76	0.81	0.04	48.3	0.26	0.07	0.79	0.54	0.27
1950	20	0.75	0.68	0.74	0.03	47.7	0.59	-0.07	0.55	0.76	0.37
2100	43	0.71	0.69	0.71	0.06	46.9	0.33	-0.06	0.20	0.45	0.25
2200	41	0.56	0.48	0.60	0.07	47.1	0.18	-0.02	-0.17	0.25	0.22
2300	35	0.57	0.45	0.53	0.11	45.4	0.14	0.00	0.00	0.25	0.12
2400	24	0.37	0.48	0.50	0.06	46.4	0.11	0.00	-0.36	0.11	0.03
2500	23	0.42	0.37	0.40	0.07	44.1	-0.12	-0.01	0.08	-0.14	0.02
2600	31	0.23	0.28	0.28	0.06	46.9	-0.20	0.05	-0.02	-0.12	-0.07
2700	35	0.24	0.26	0.24	0.06	47.5	-0.17	0.04	0.00	-0.39	-0.27
2800	11	0.38	0.15	0.23	0.04	43.1	-0.29	0.04	0.11	-0.40	-0.34
2900	27	0.19	0.15	0.14	0.03	45.1	-0.34	0.06	-0.01	-0.36	-0.26
3000	25	0.06	0.08	0.08	0.01	49.1	-0.45	0.10	-0.10	-0.81	-0.46
3200	9	0.00	0.17	0.04	0.00	47.3	-0.60	0.02	-0.25	-0.98	-0.53
3300	17	0.00	0.09	0.08	0.02	46.0	-0.42	0.02	-0.19	-0.71	-0.29
4000	21	0.00	0.00	0.02	0.00	46.6	-0.49	0.00	-1.10	-0.62	-0.42

Table 2a. Parameters for the individually fitted two parameter and stepped clines for all genetic loci. Note that the *xpos1* and *xpos2* variables give the number of meters upstream and downstream (respectively) from the center where the exponential tails begin. The best fit cline is highlighted in bold

Locus	Sigmoid cline parameters		Stepped cline parameters			
	Centre (m)	Slope	<i>xpos1</i> (m)	<i>xpos2</i> (m)	<i>tslope1</i>	<i>tslope2</i>
<b><i>Eda</i></b>	2308	-2.98				
<i>Eda</i> (stepped)	2334	-3.34	1660	330	0.99	0.26
<b><i>ATP1a1</i></b>	2270	-2.62				
<i>ATP1a1</i> (stepped)	2270	-2.75	2340	1514	0.81	0.85
<b>P3D05</b>	2330	-2.59				
P3D05 (stepped)	2465	-8.98	104	1	0.24	0.24
<b>P6A10</b>	2402	-3.32				
P6A10 (stepped)	2479	-5.39	418	102	0.34	0.34
<b>P4G01</b>	3320	-0.77				
P4G10 (stepped)	2374	-8.73	1	15	0.012	0.142
<b>P6D05</b>	2574	-0.40				
P6D05 (stepped)	2920	-9.92	30	9	0.056	0.007
<b>P7A07</b>	3216	-0.65				
P7A07 (stepped)	2909	-8.03	1	29	0.042	0.092
<b>P7C08</b>	3220	-1.23				
P7C08 (stepped)	2910	-2.02	1	960	0.14	0
<b>P7E08</b>	2730	-0.95				
P7E08 (stepped)	2570	-1.91	50	470	0.31	0
<b>P7G05</b>	2376	-1.32				
P7G05 (stepped)	2352	-9.41	32	66	0.11	0
<b>P7H05</b>	980	-1.47				
P7H05 (stepped)	1570	-8.86	120	1300	0.17	0.44
<b>P9D09</b>	2483	-1.22				
P9D09 (stepped)	2357	-2.38	260	710	0.045	0.88



Table 2b. Parameters of the individually fitted clines for all morphological traits. As there was no support for stepped clines for any morphological trait, only the variables for sigmoid clines are shown.

	Sigmoid cline parameters		Trait means		Trait variances		
	Centre (m)	Slope	$\mu_{\min}$	$\mu_{\max}$	$\sigma_1$	$\sigma_2$	$\sigma_3$
Dorsal fin	2410	-3.82	-0.50	0.56	0.72	0.72	0.72
Caudal peduncle depth	2330	-6.26	-0.05	0.08	0.14	0.17	0.15
Left pectoral fin	2170	-14.51	-0.13	0.40	0.67	1.14	0.61
Left pelvic spine	2440	-4.00	-0.75	0.79	0.68	0.72	0.52
Second dorsal spine	2340	-3.24	-0.46	0.61	0.62	0.39	0.41

Figure 1. Map of the Bonsall Creek study site on Vancouver Island (white dot on inset) and of the sampling locations along the river itself (main map). The grey shaded area represents salt marsh; the salt finger at high tide reaches 2350m along the creek, although there is some tidal fluctuation in water level at 2400m.

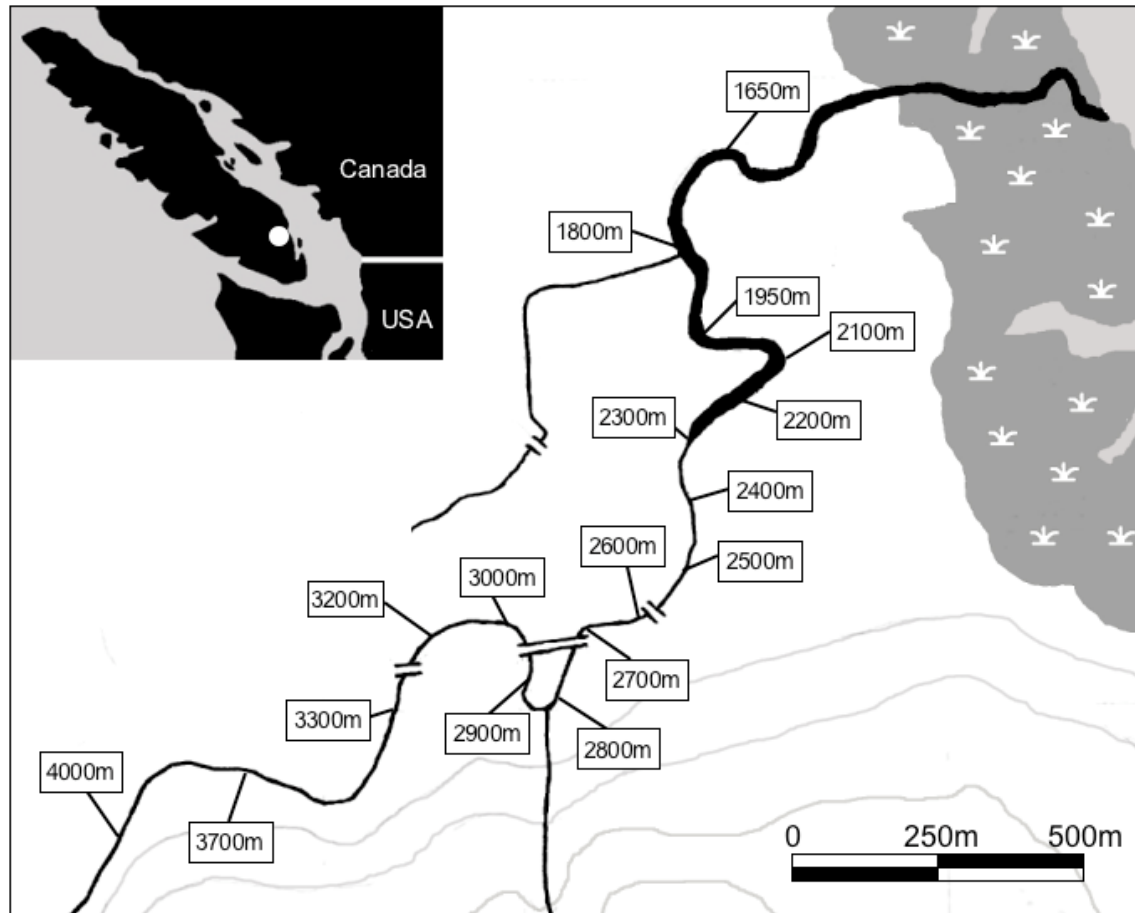


Figure 2. Individually fitted lines for the 12 genetic loci. The sigmoid cline is shown in black; the stepped cline is shown in red for the loci where this was a better fit.

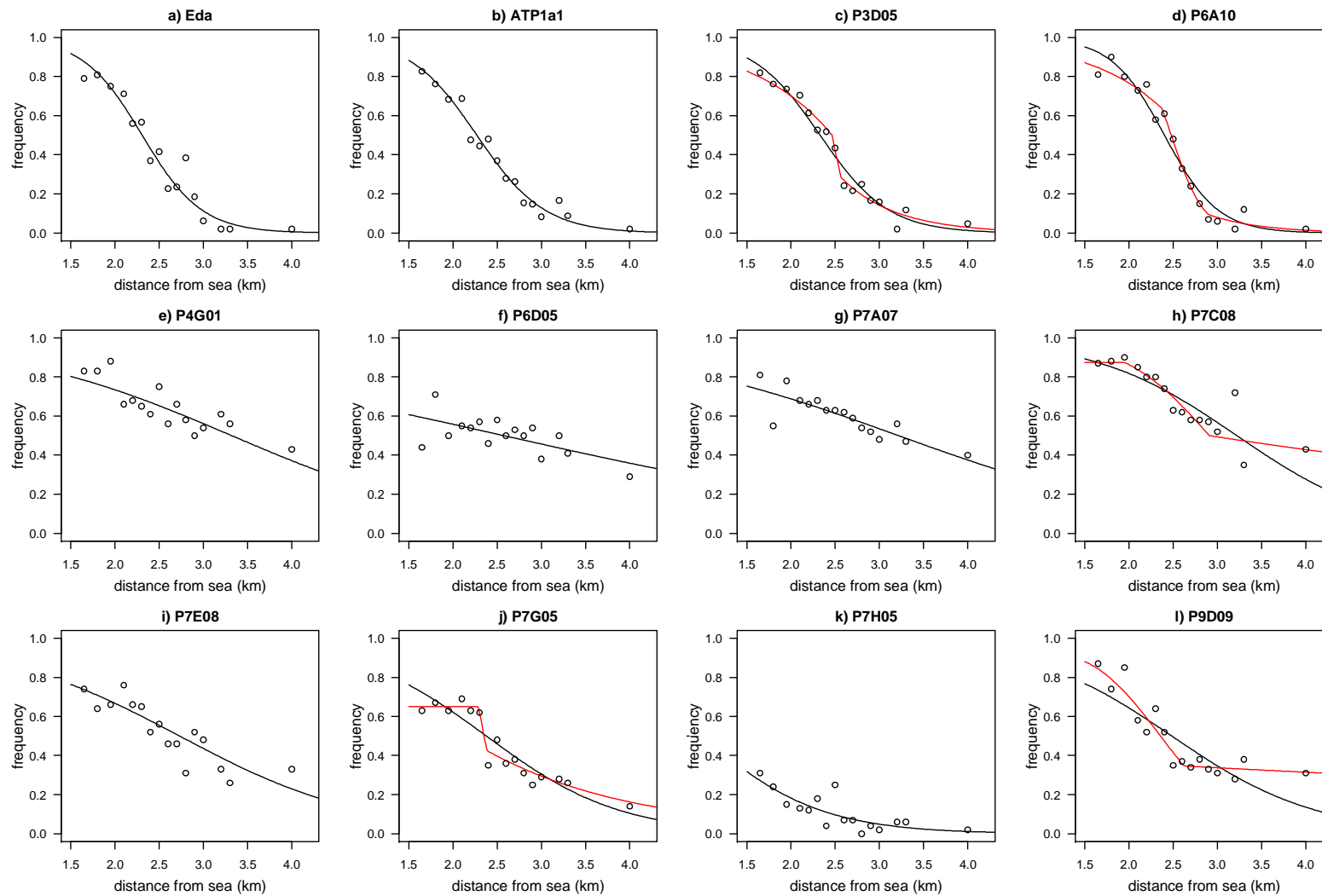


Figure 3a-e Clines in morphological traits through the hybrid zone. The width of the shaded band shows the trait variance through the zone, as calculated from  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$ .

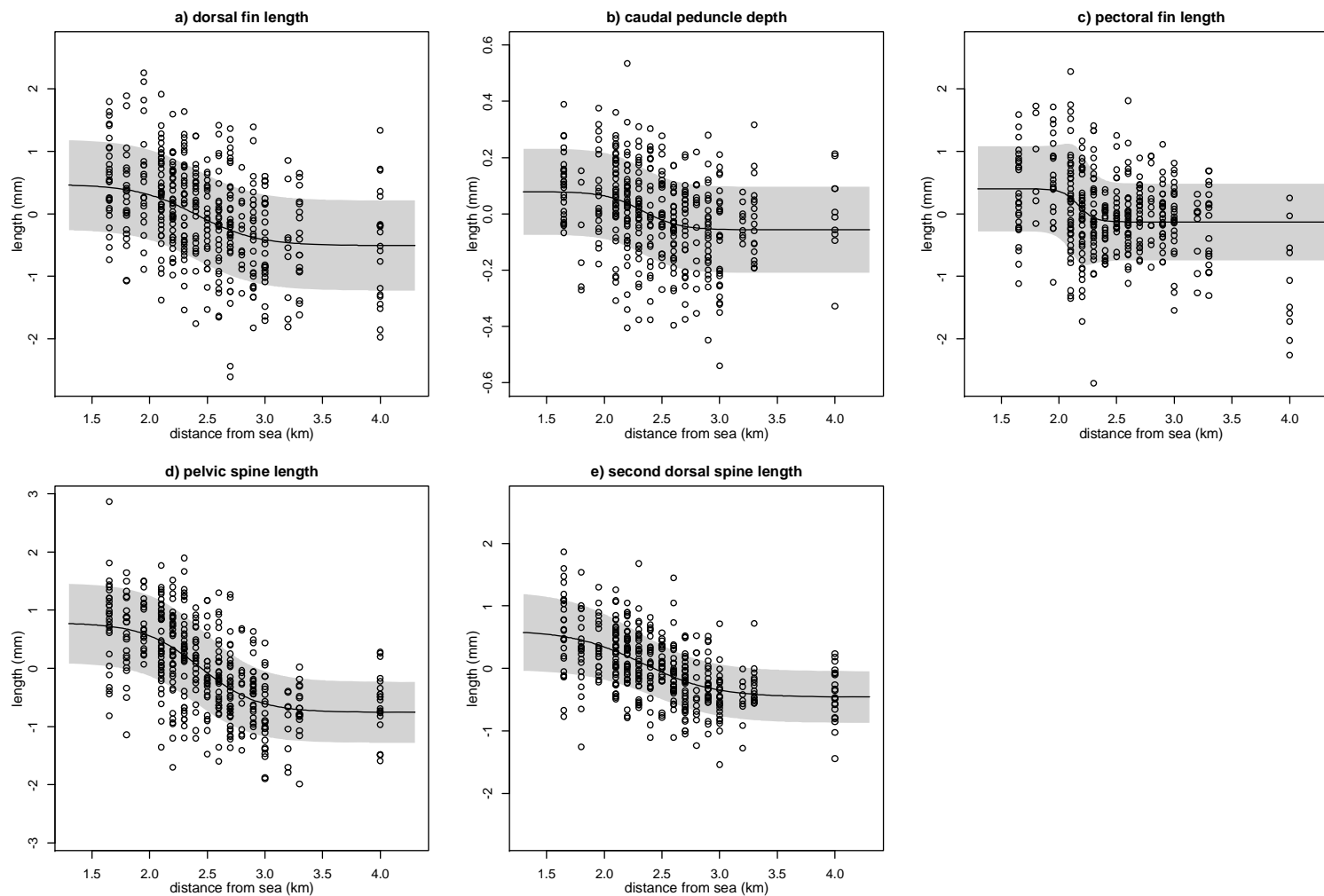


Figure 4. The best-fit clines in the hybrid zone. Salinity at high tide (green line) is plotted on an inverse scale on the right hand axis. The black line is the cline for pectoral fin length. The red line is the joint cline for *Eda*, *ATP1a1*, dorsal fin length, caudal peduncle depth, pelvic spine length, and 2<sup>nd</sup> dorsal spine length. The sigmoid cline for P3D05 is shown with a dot-dash blue line, and the sigmoid cline for P6A10 with a dashed blue line. The grey dotted lines are the best-fit sigmoid clines for the remaining eight anonymous SNPs.

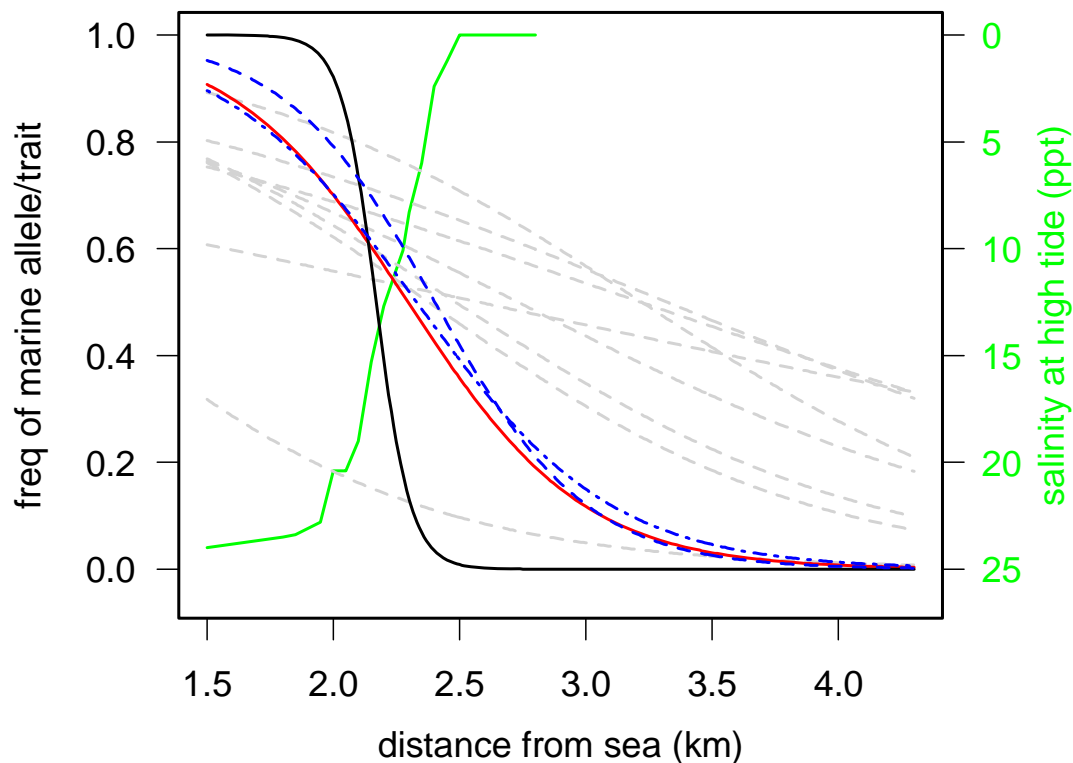


Figure 5. Bar plots of Hybrid Index through the hybrid zone. The Hybrid Index is given by the total number of anadromous alleles an individual carries at *Eda*, *ATP1a1*, P3D05, and P6A10.

