Sweeping genomic remodeling through repeated selection of alternatively adapted haplotypes occurs in the first decades after marine stickleback colonize new freshwater ponds

Susan Bassham^{1,%}, Julian Catchen^{2,%}, Emily Lescak^{3,4}, Frank A. von Hippel^{3,5}, and William A. Cresko¹

1. Institute of Ecology and Evolution, University of Oregon, Eugene, OR, 97403, USA; 2. Department of Animal Biology, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA; 3. Department of Biological Sciences, University of Alaska Anchorage, Anchorage AK 99508, USA; 4. College of Fisheries and Ocean Science, University of Alaska Fairbanks, Fairbanks AK 99775, USA; 5. Present address: Department of Biological Sciences & Center for Bioengineering Innovation, Northern Arizona University, Flagstaff, AZ, 86011, USA.

% These authors contributed equally to this work.

Direct correspondence to sbassham@uoregon.edu, wcresko@uoregon.edu, jcatchen@illinois.edu

Abstract

Heterogeneous genetic divergence can accumulate across the genome when populations adapt to different habitats while still exchanging alleles. How long does diversification take and how much of the genome is affected? When divergence occurs in parallel from standing genetic variation, how often are the same haplotypes used? We explore these questions using RAD-seq genotyping data, and show that broad-scale genomic re-patterning, fueled by standing variation, can emerge in just dozens of generations in replicate natural populations of threespine stickleback fish (*Gasterosteus aculeatus*). After the catastrophic 1964 Alaskan earthquake, marine stickleback colonized newly created ponds on seismically uplifted islands. We find that freshwater fish in these young ponds differ from their marine ancestors across the same genomic segments previously shown to have diverged in much older lake populations. Outside of these core divergent regions the genome shows no population structure across the ocean-freshwater divide, consistent with strong local selection acting in alternative environments on stickleback populations still connected by significant gene flow. Reinforcing this inference, a majority of divergent haplotypes that are at high frequency in ponds are shared across independent freshwater populations and are detectable, at low frequency, in the sea even across great geographic distances. Building upon previous work in this model species for population genomics, our data suggest that a long history of divergent selection and gene flow across stickleback in oceanic and freshwater habitats has created balanced polymorphism in large genomic blocks of alternatively adapted DNA sequences, ultimately stoking - and potentially channeling - rapid, parallel evolution.

contemporary evolution – ecological divergence – population genomics –Gasterosteus aculeatus – threespine stickleback

Introduction

When populations of a species adapt to different environments or habitats, parts of the genome may chart divergent evolutionary courses, leading to heterogeneous patterns of genomic differentiation (1-3). Recent research demonstrates that complete genetic isolation is not a prerequisite for such divergence (4-6). While these genomic patterns can occur neutrally via genetic drift in isolated populations, pronounced heterogeneous genomic divergence might often be shaped by the action of local adaptation to different habitats combined with gene flow among the populations (7, 8). Under strong natural selection, alleles and genotypes that are alternatively adapted to different habitats can be maintained at divergent frequencies even when the populations are still exchanging alleles. Local adaptation and gene flow can result in differentiation at selected loci and homogenization of variation in neutral regions of the genome (8, 9). When such divergence with gene flow plays out over geographic space, the metapopulation structure of a species can promote the maintenance and reuse of standing genetic variation for adaptation and even speciation (10-12).

Cases of rapid adaptation (13-16) - so-called contemporary evolution (17-21) - often invoke the use of standing variation because the extended time needed to accumulate new beneficial mutations is presumed to limit the rate of adaptation (22-25). Although standing genetic variation permits accelerated change, the patterns of heterogeneous genomic differentiation that materialize during contemporary evolution still remain largely unexplored in natural systems (26) particularly in the context of local adaptation with gene flow (27). It is still not known what proportion of the genome is affected during contemporary evolution by direct selection on standing adaptive variants, and how much of the genome is influenced by linked selection such as genetic hitchhiking and genetic draft (28-31). When a species exists across diverse habitat types with distinct selective pressures, to what extent do independent populations adapting to similar habitats initially make use of the same haplotypes, and how and at what frequencies are these adaptive haplotypes distributed and maintained species-wide? A first step in tackling these questions is to survey heterogeneous genomic re-patterning using haplotype-level data in organisms while they are first evolving in response to different natural habitats. Across the northern hemisphere marine threespine stickleback fish (*Gasterosteus aculeatus*) have invaded and adapted to countless freshwater habitats

over multiple time scales, making them a fruitful model for vertebrate evolution (32-34) and for population genomics of rapid adaptation (35). Marine and freshwater stickleback differ considerably in many traits (36-40), and remarkably parallel phenotypic and genomic evolution has followed independent invasions by marine fish into freshwater habitats, including those on landscapes de-glaciated only thousands of years ago [41-49]. At least some of this phenomenon might be due to the reuse of standing genetic variation (41-43), enabling the rapid and parallel adaptive evolution documented in this species (44-51). Schluter and Conte (26) framed one possible scenario to explain these dynamics: a freshwater-adapted stickleback genotype, scattered into the marine population by gene flow and recombination, might be brought together and reintegrated during subsequent colonizations of freshwater habitats. What is still needed in stickleback is investigation of the time scale on which reassembly can transpire, which depends strongly on the frequencies and genomic granularity of freshwater adapted haplotypes carried by marine populations. Both of these qualities have been largely unquantified.

Here we explore the nature of this ‘freshwater genome scattering and reassembly’ during the earliest stages of freshwater adaptation by marine stickleback, using naturally formed stickleback populations for which convincing genetic and geologic evidence indicates they colonized new freshwater habitats within the last 50 years (35). Submarine terraces that encircled islands in the Gulf of Alaska and Prince William Sound were suddenly thrust above sea level in 1964 by the Great Alaska Earthquake (52-54), and the resulting changes in sediment deposition and erosion over the following years created new freshwater ponds (35). Threespine stickleback fish that now inhabit many of these ponds have become as morphologically distinct in degree and form from their immediate marine ancestors as have much older mainland Alaska freshwater stickleback populations (35, 54, 55), similar to phenomena documented in older freshwater populations throughout the northern hemisphere (33, 36, 56-58). Previously, using a relatively small number of randomly chosen genetic markers in these new populations, we showed that marine fish likely invaded and adapted to fresh water several independent times on different islands and even among ponds on the same island (35).

The present study yields a much denser dataset of loci in over 1200 fish from independently derived stickleback populations on three seismically uplifted marine islands (Danger, Middleton and Montague Island), as well as from mainland Alaska and Oregon. We use this dataset to dissect genome-wide patterns of haplotype frequency divergence across populations that range in age from just decades to thousands of years. By exploring these replicate natural experiments in adaptation to alternative environments, we find that a much larger proportion of the genome is affected - through combined effects of direct and linked selection - than had been previously reported in stickleback. Loci with alternative haplotypes that dominate in either marine or freshwater populations are clustered in the genome, particularly across broad regions of divergence where marine fish are relatively depleted in haplotype diversity. We show how marine stickleback, even at great geographic distances, act as carriers of a common set of freshwater genotypes, leading to repeated selection of the same haplotypes in independently colonized lakes and ponds.

Results

Parallel patterns of genomic divergence involving a quarter of the genome assemble in decades on a background of no population structure

Despite their young age and their isolation by ocean expanses, we find that stickleback populations newly adapted to fresh water on seismically uplifted Middleton and Montague Islands (south of mainland Alaska) reiterate genomic patterns first described in fish from much older, post-glacial lakes on the Alaskan mainland (46) (Fig 1). We calculated genome-wide population genetic differentiation statistics (F_{ST} , F_{ST}' , and Φ_{ST} ; Table S1) using RAD-seq genotyping in stickleback sampled from multiple populations, and we performed genome scans using continuous divergence values (Fig 1; Fig 2).

A history of multiple founding events from the sea is reflected in the clear population structure that we documented previously (35) among groups of freshwater ponds on three different seismically uplifted islands (Fig S1). Using these pre-defined population genetic groupings for our analyses, we found that the genomic patterns of divergence from marine genomes in one Montague Island group (MoSW) and two groups on Middleton (MiFW1, MiFW2), despite their independent colonization, are broadly

Fig 1. Nascent and old freshwater stickleback populations share parallel patterns of genomic divergence from marine fish. Smoothed, SNP-based F_{ST} is plotted along the genome in comparisons of mainland coastal marine versus mainland post-glacial Cook Inlet freshwater, and island marine fish (from collection sites Da02, Mi17, Mi23) versus freshwater populations that formed after the 1964 earthquake on geographically separated islands: Middleton Island (FW1: Mi07, Mi17; FW2: Mi06, Mi11, Mi16, Mi28) and Montague Island (Mo30, Mo33). These populations were found to form cohesive groups via genetic structure analyses in Lescak et al. [35]

congruent, resembling patterns typical of much older populations (Fig 1; Fig 2). A striking pattern is that, outside of foci of high divergence, baseline genetic divergence approaches zero between marine fish and freshwater fish in the new ponds. In southeast Montague (MoSE) and Danger Is. populations, though, we detected only a subset of the F_{ST} signatures of selection common in freshwater-adapted fish (Fig 2).

In order to quantify what proportion of the genome has diverged over such a short time frame, and to what degree these regions of divergence are shared across independently evolved populations (Fig 1; Fig 2), we developed a Hidden Markov Model (HMM). The emission and transition parameters of the HMM were trained on patterns of F_{ST} values for RAD loci across the genome in freshwater-marine comparisons. The trained model delineated two states: diverged or not diverged (Fig 3; Fig S2). We found that a significant proportion of the genome falls within well-defined regions that diverge between marine and fresh water in all comparisons: 23% for Middleton FW1, 29% for Middleton FW2, 24% for SW Montague, and 18% for a mainland group (Table 1; Table S2). As expected from the conspicuously parallel genomic divergence across comparisons (Fig 1; Fig 2), the genomic blocks are largely nested when compared

Fig 2. Patterns of genomic divergence are similar among young, independently colonized freshwater locales. Smoothed, haplotype-based F_{ST} is plotted along the genome. Concentric rings are numbered from the center out, and labeled with 5 Mb intervals. Marine populations (Mi17 vs. Da02) from different islands about 100 km apart show little differentiation (Ring 1). While Middleton Island (MiFW2, Ring 2), sympatric freshwater vs. marine ecotypes both from Mi08 (Ring 3), and SW Montague Island (Ring 4) all show parallel patterns of marine-freshwater divergence, the likely younger populations on Montague (SE, Ring 5) and Danger Island (Da04, Ring 6) differ strongly from marine on only a subset of the typically divergent linkage groups.

among all four freshwater groups. The divergent regions of the two Middleton freshwater groups overlap by over 73%, Middleton and SW Montague overlap by over 70%, and these island populations overlap by between 51 and 59% with the much older mainland populations (Table 1; Table S2). Clearly, the majority of the patterns of genomic divergence are shared across populations, young and old, and this shared, persistent divergence amounts to nearly 15% of the genome.

Fig 3. Genomic blocks of freshwater-marine divergence largely overlap among freshwater populations. A) A Hidden Markov Model was trained on F_{ST} patterns to discriminate between diverged and not diverged regions. Shown here, for example, is the outcome of the model on three linkage groups that have large regions of divergence and one linkage group that has little divergence. From this, the proportion of the genome diverged in each freshwater group (dashed lines) and the divergence overlap (bold green line) among these populations were calculated. B) The total proportion of the genome diverged from marine in the compared freshwater populations ranges nearly to 30%, the majority of which (totaling $\approx 15\%$ of the genome) overlaps across freshwater populations.

	MiFW1	MiFW2	MoSW	Cook Inlet
MiFW1	23.26%	22.12%	20.07%	15.15%
MiFW2	73.27%	29.06%	22.01%	16.27%
MoSW	72.94%	70.15%	24.33%	15.95%
Cook Inlet	56.65%	51.79%	59.02%	18.64%

Table 1. Regions of the stickleback genome that diverge because of adaptation highly overlap between freshwater populations, both young and old. The total amount of the genome allocated to the FW vs. OC “diverged” state by a hidden Markov Model is listed, along the diagonal in bold, for each population. Above the diagonal are the proportions of the genome populations have in common. Below the diagonal is the degree of overlap of the divergent stretches of the genome for each population comparison. Despite the much older time since founding, the Cook Inlet freshwater group still shares more than 50% of divergent regions with the much younger uplift populations.

Marine and freshwater populations differ in patterns of haplotype diversity and composition

F_{ST} statistics quantify relative divergence affected by differences in genetic frequencies among, as well as diversities within, populations (59-61). Because it generates DNA sequence data, RAD-seq allowed us to

examine both the relative and absolute haplotype diversities within differentiated genomic regions in each population, as well as document the distribution and abundance of particular haplotypes across populations. We found that divergent genome segments correspond to regions where marine and freshwater populations differ in haplotype diversity. Marine genomes have lower absolute haplotype diversity across these regions, relative to non-divergent parts of the genome (Fig 4; Fig S3).

At individual RAD loci with high estimates of F_{ST} both the marine and freshwater populations each often have a single but different majority haplotype (Table 2; Table S3), where we define majority haplotypes as the highest frequency haplotypes that - singly or in combination - account for at least 60% of genotypes at a given locus (Fig 5). Independent freshwater populations most often have the same majority haplotype across all shared loci

Population	Total loci	Case 1	% Rare	% Strict Alt	Case 2	% Rare	Case 3	% Rare	Case 4
MiFW1	4847	2034	74.09	66.87	1800	82.83	363	98.94	610
MiFW2	3021	1294	84	75.73	1230	91.14	286	98.25	192
MoSW	2114	494	75.51	70.24	1408	90.63	87	93.1	108
MoSE	4370	1121	38.72	29.17	1703	73.69	261	89.27	1245
Da04	3800	1063	29.44	19.47	1249	67.17	188	85.11	1265

Table 2. Freshwater versus marine divergent loci categorized by haplotype characteristics. Loci with $F_{ST} \geq 0.2$ are tallied for each population group. Of these, the number that fall into each Case (as defined in Fig 5) is given. Also listed, where applicable, is the proportion ("% Rare") of these loci carrying a freshwater majority haplotype that occurs only rarely (i.e., less than 10% allele frequency) in the sea. For Case 1 is given the proportion ("% Strict Alt") of loci that satisfy these criteria: they carry a single majority haplotype in freshwater that is rare in the sea and they carry a different single majority haplotype in the sea. The island populations with a pattern of divergence that closely parallels older mainland populations have the largest proportion these of loci with radically alternative FW vs. OC haplotype frequencies.

at which there is a single, alternative majority freshwater haplotype (which we call "Case 1 loci"; see Fig 5 for definitions of four qualitatively different cases). For instance, 98.8% (1036 of 1052) of the Case 1 loci shared by the two freshwater population groups on Middleton Island have the same majority freshwater haplotype, and of those shared by both MiFW1 and Montague SW, 99.2% (396 of 399) of them have the same majority haplotype. In MiFW1, MiFW2 and MoSW, loci with majority haplotypes alternative to those in OC (referring to samples combined from marine populations Mi17, Mi23 and Da02) are clustered in the divergent regions delineated by the F_{ST} -based HMM (Fig 5). Outside of these divergent regions, loci with freshwater vs. OC alternative haplotypes occur, but the density of them differs by population group, with MoSW having the fewest, MiFW1 having more than MiFW2, and MoSE have the most (Fig 5; Fig S4).

Another important qualitative difference between MoSW and Middleton can be seen in the loci falling within the divergent regions; MoSW retains more of the common marine haplotypes at these loci and therefore has fewer loci with only freshwater haplotypes in the majority (i.e., more Case 2 loci, plotted in blue in Fig 5). Taken together, these two distinctions might be expected if MoSW has greater introgression between freshwater and marine fish, as was inferred by Lescak *et al.* (35), who described morphological intermediates in this population. MoSE (which lacks most of the peaks of high F_{ST}) has higher F_{ST} relative to OC than do the other pond populations across genomic regions that do not usually diverge in fresh water (Fig 2).

This pattern is likely a consequence of the low overall haplotype diversity in MoSE, including in regions that house relatively higher haplotype diversity in the sea, perhaps due to small founding population size (Table 3; Fig S3). An interesting correlate of this pattern in

Population	Gene Diversity	Haplotype diversity
MiFW1	0.2771	0.395
MiFW2	0.3106	0.4408
MoSW	0.3071	0.4367
MoSE	0.2136	0.2971
Da04	0.2036	0.2844

Table 3. Genome-wide averages of gene and haplotype diversity in the uplift island populations. Diversity is lower in the potentially younger pond populations on Danger Island and in southeast Montague Island.

MoSE is that F_{ST} relative to OC drops across regions normally of high F_{ST} in other freshwater-marine comparisons because these genomic segments have low diversity in *both* MoSE and OC, and the diffuse number of higher F_{ST} loci found in MoSE predominantly result from subsampling of the majority OC haplotypes (such as from 7 to 21 Mb on LG VII; Fig 2; Fig 6; Fig S4).

Danger Island presents another qualitatively distinct scenario from the other freshwater populations in this study; only a fraction of the parallel pattern of freshwater-marine divergence - a single F_{ST} peak on LG I – is prominent in Da04. This segment encompasses a known inversion whose orientation has been shown to differ consistently between freshwater and marine genomes (62). Relative to OC, Da04 has diffusely elevated F_{ST} due to genome-wide reduction in haplotype diversity (Fig 2; Fig S3; Table 3), but F_{ST} crescendos at the LG I inversion. Unlike in the

Fig 4. Divergent genomic regions differ also in haplotype diversity between freshwater and marine fish. A) Concentric rings are numbered from the center out. Blocks of elevated F_{ST} in a comparison of Middleton FW2 vs. marine (Ring 1) highly correspond with relative increases in haplotype diversity in fresh water (Ring 2). On Ring 2, blue represents higher diversity in fresh water than in marine, and red is higher diversity in marine than fresh water. Smoothed haplotype diversity is plotted separately for FW2 (Ring 3) and marine (Ring 4). While a fraction of haplotype diversity outliers does exceed the value range shown, > 99.9% of the smoothed values were 1.0 or less. See Methods for how diversities and difference are calculated. B-D) Middleton FW2 and marine are compared across LG IV, a linkage group to which many morphological traits map that differ between marine and freshwater ecotypes [38]: B) divergence statistics, C) gene diversity, D) haplotype diversity, and E) gene and haplotype diversity difference. Where values in panel D are greater than zero, fresh water has greater diversity than marine.

other freshwater groups that share this F_{ST} peak, however, elevated F_{ST} relative to OC extends to the inversion's flanking regions in Da04 (Fig 2; Fig 6). We find that two types of loci primarily occupy the inversion and closely linked regions. In the inversion itself, there is a concentration of loci that have increased in alternative FW haplotypes, and flanking these are loci depleted for common marine haplotypes (Fig 6; Fig S5). Together, these could signal a recent or ongoing selective sweep in Da04.

Alternative freshwater majority haplotypes are readily detectable at thousands of loci in the marine population

At loci with $F_{ST} \geq 0.2$, most haplotypes that are in the majority in fresh water can be found in the sea, though they are typically rare (frequency $\leq 10\%$) (Table 2; Table S3). For example, 81% of alternative majority haplotypes identified in MiFW1 and 68% of those identified in MiFW2 are found in the marine comparators (131 marine fish from Mi17, Mi23, and Da02) (Table S3). Marine populations that were sampled at great geographic distances from one another - at Middleton Island, at Danger Island (> 100 km mainly across open ocean from Middleton) and even as far as the southern Oregon coast (> 2500 km away, along a coastal route) – carry many of the freshwater majority haplotypes identified on Middleton Is. (Fig 7).

Fig 5. Regions of parallel divergence contain loci with alternative FW versus OC haplotypes. A) To characterize the nature of loci with high F_{ST} in OC-FW comparisons, we assign these loci to several categories: at Case 1 loci OC and FW populations have single but alternative majority haplotypes (i.e., accounting for at least 60% of population haplotypes); at Case 2 loci, the FW population has one or more haplotypes that are majority (accounting singly or in sum for at least 60% of haplotypes) but are not majority in the sea, and they have at least one majority haplotype that is also a majority haplotype in the sea; at Case 3 loci, the FW population has more than one majority haplotype that is not majority in the sea and no majority haplotypes in common with the sea; and at Case 4 loci, the FW population has majority haplotypes that are a subset of the majority haplotypes in the sea. B) Loci of $F_{ST} \geq 0.2$ are assigned to cases and plotted across the genome for two FW groups from Middleton Is. FW and one Montague Is. While these cases are not strictly discrete, we do think they broadly capture qualitatively different patterns of population genomic processes. Sampling effects could move some loci with borderline allele frequencies (i.e., those with frequencies near our parameter boundaries) between Cases 1-3, but are less likely to move such loci to Case 4 (for example, see Table S3).

Fig 6. Changes in haplotype frequency reveal ongoing processes in the youngest freshwater populations. The set of haplotypes at each locus (where $F_{ST} \geq 0.2$) is represented as a stacked bar chart of haplotype frequencies. Red haplotypes correspond to marine majority haplotypes and blue to fresh water. A) In the Danger Is. FW pond, a LG I region of elevated F_{ST} (relative to OC) encompassing a known inversion polymorphism shows a strong pattern of FW-OC alternative haplotypes (Case 1 and Case 3 loci) flanked by regions with increased density of loci that are depleted in haplotype diversity relative to OC (i.e., loci classed as Case 4). B) In the southeast Montague ponds, LG VII lacks the high F_{ST} typically seen across the middle of the chromosome in other FW-OC comparisons, and instead there is elevated F_{ST} at the chromosome ends. This inverse pattern is generated by the sparsity of loci with FW-OC alternative haplotypes (Case 1 and Case 3 loci) but an increase in the number loci at which the ocean is more haplotype-rich (Case 4). A clear red-to-blue shift of haplotypes can be seen between marine and FW2 loci on LGVII. For comparison, the F_{ST} heat-map plots to the right are as in Fig 2.

For freshwater haplotypes identified in the island marine populations, the median and mean allele frequency spectra (AFS) at those loci were nearly identical across the three populations (Fig 8; Table S4; Fig S6). In fact, linear models of the pairwise comparison of freshwater allele frequencies across each of the three marine populations resulted in R^2 values of 0.69, 0.65 and 0.83 for the Da02:Mi17, Da02:Mi23 and Mi17:Mi23 regressions, respectively. Thus, even between marine populations on two different islands nearly 70% of the variance in allele frequencies of one population are predicted by the other population, and between Mi17 and Mi23 on different coasts of the same island the congruence goes up to nearly 85%. This similarity of freshwater allele frequencies across marine populations

Fig 7. Marine stickleback carry rare haplotypes that are found at high frequency in freshwater fish. Plotted for three geographically distant marine populations are all RAD loci that fit two criteria: the locus falls into Case 1 or Case 3 (as defined in Fig 5) in a comparison between the marine population and a Middleton Island freshwater (Mi11), and at the same locus at least one marine fish carries a FW majority haplotype that is alternative to majority haplotype/s in the sea. Rings are numbered from the center out. Loci are colored by which Case they belong to.

holds regardless of whether they were calculated for freshwater majority haplotype sets defined by freshwater populations that are geographically nearby or distant from each marine population (Fig 8; Fig S6; Fig S7). Although the AFS of the entire genome was similar across populations, there was appreciable variation among chromosomes within populations with regards to median and mean freshwater allele frequencies as well as AFS (Table S4; Fig S8), as might be expected given different patterns of divergence across the chromosomes.

The same allele frequencies can be generated if many marine fish each carry a few freshwater haplotypes or if a few marine fish carry a majority or “jackpot” numbers of such

Fig 8. Three marine populations have similar allele frequency spectra of FW majority haplotypes. Distributions of allele frequencies across all Case 1 through 3 loci (as described previously) in comparisons between each marine population and the MIFW1 population. Allele frequency histograms are presented as 2% bins.

haplotypes. To test whether freshwater haplotypes are primarily carried by only a fraction of marine individuals, and to assess the patterns of freshwater haplotypes across the genomes of marine fish, we examined haplotype states at all Case 1-3 loci that carry alternative freshwater majority haplotypes from each marine fish. For example, most of the marine fish in the OC group carried fewer than 10% of these freshwater haplotypes, most often at heterozygous loci, but some fish had much larger numbers of them (Fig S9). Loci bearing freshwater haplotypes are scattered diffusely among linkage groups in most individual marine fish, but in the fish that carry them at the highest proportions, these haplotypes often reside at loci densely concentrated on various subsets of the freshwater-oceanic most divergent linkage groups. For example, one Mi17 individual carries over 46% of the Case 1-3 freshwater majority haplotypes identified from LG VII, but carries less than 7% of those

identified from LG IV, and a Da02 fish harbors over 47% of the LG XII freshwater haplotypes but less than 8% of those from LG VII (Table 4).

	LGIV	LGVII	LGI	LGXX	LGXII	LGXI	LGIX	LGVIII	LGXXI	LGXIX	LGII	LGXIII	LGX	LGIII	LGVI	LGXVI	LGVI	LGXVIII	LGXVII	LGXIV	LGXV	Genome
	550	428	356	250	240	226	222	212	199	162	157	147	130	112	112	99	85	72	53	45	19	3876
	14.2%	11.0%	9.2%	6.5%	6.2%	5.8%	5.7%	5.5%	5.1%	4.2%	4.1%	3.8%	3.4%	2.9%	2.9%	2.6%	2.2%	1.9%	1.4%	1.2%	0.5%	
mi17_2188.49	91.1%	76.4%	90.7%	89.2%	64.6%	86.7%	97.3%	86.8%	83.4%	88.9%	72.0%	82.3%	87.7%	85.7%	55.4%	66.7%	62.4%	62.5%	73.6%	60.0%	63.2%	82.12%
da02_2197.20	60.9%	71.7%	45.2%	45.6%	42.1%	11.5%	26.1%	40.6%	76.9%	46.3%	16.6%	15.7%	13.1%	20.5%	42.9%	12.1%	22.4%	23.6%	26.4%	22.2%	31.6%	42.08%
da02_2197.44	62.0%	56.5%	12.1%	4.4%	44.2%	10.6%	17.6%	36.3%	65.8%	54.9%	38.9%	21.1%	7.7%	18.8%	18.8%	13.1%	16.5%	22.2%	35.9%	20.0%	36.6%	34.18%
da02_2197.31	4.9%	62.9%	9.0%	55.2%	46.3%	58.9%	17.1%	14.6%	8.0%	14.2%	52.9%	45.6%	29.2%	21.4%	41.1%	12.1%	17.7%	26.4%	41.5%	24.4%	31.6%	29.95%
da02_2197.05	6.4%	51.6%	15.2%	53.2%	8.3%	56.2%	19.8%	9.9%	68.3%	53.7%	39.5%	15.7%	9.2%	35.7%	18.8%	19.2%	18.8%	25.0%	28.3%	17.8%	26.3%	28.82%
da02_2197.18	5.3%	7.5%	46.4%	57.6%	47.5%	10.6%	22.5%	9.4%	7.0%	11.7%	10.8%	17.7%	30.8%	41.1%	40.2%	11.1%	24.7%	18.1%	9.4%	26.7%	21.1%	21.96%
mi17_2188.27	24.0%	5.8%	11.5%	7.2%	7.9%	36.3%	22.1%	9.9%	5.0%	12.4%	45.9%	15.7%	13.1%	25.0%	14.3%	21.2%	21.2%	16.7%	9.4%	20.0%	31.6%	16.62%
mi17_2188.20	6.4%	46.3%	11.2%	11.2%	4.2%	12.0%	15.3%	11.8%	7.0%	13.0%	10.2%	17.0%	16.9%	13.4%	11.6%	15.2%	21.2%	25.0%	18.9%	26.7%	26.3%	15.51%
da02_2197.07	18.7%	6.5%	11.5%	8.4%	7.9%	15.0%	22.1%	13.7%	5.0%	11.7%	26.8%	10.9%	10.0%	25.9%	11.6%	10.1%	21.2%	20.8%	5.7%	13.3%	10.5%	13.42%
da02_2197.38	9.1%	7.2%	8.4%	52.8%	6.3%	8.0%	13.5%	12.7%	10.1%	8.6%	8.3%	11.6%	14.6%	16.1%	12.5%	13.1%	24.7%	13.9%	15.1%	15.6%	21.1%	13.18%
da02_2197.27	5.1%	8.6%	9.3%	50.8%	9.6%	10.2%	12.2%	10.4%	7.0%	13.0%	12.1%	10.2%	10.8%	18.8%	13.4%	18.2%	21.2%	15.3%	15.1%	17.8%	15.8%	13.03%
...																						
mi17_2188.48	7.1%	5.4%	11.2%	8.4%	5.0%	11.1%	18.9%	9.9%	8.5%	13.6%	9.6%	19.7%	12.3%	21.4%	10.7%	14.1%	21.2%	29.2%	17.0%	11.1%	26.3%	11.09%
mi17_2188.13	5.6%	5.6%	12.1%	6.4%	5.4%	10.6%	23.9%	16.0%	9.6%	11.7%	7.0%	18.4%	10.0%	19.6%	8.9%	13.1%	17.7%	27.8%	28.3%	4.4%	26.3%	11.07%
mi17_2188.14	6.0%	7.7%	11.2%	9.2%	4.6%	10.2%	17.1%	11.3%	8.0%	13.0%	18.5%	14.3%	18.5%	14.3%	13.4%	7.1%	18.8%	22.2%	13.2%	17.8%	31.6%	11.02%
mi17_2188.28	5.6%	5.8%	10.4%	8.8%	8.3%	13.3%	19.8%	12.7%	6.5%	11.1%	10.8%	19.1%	8.5%	23.2%	13.4%	12.1%	20.0%	23.6%	18.9%	13.3%	5.3%	11.02%
da02_2197.50	6.9%	7.5%	9.3%	8.8%	4.6%	12.4%	15.3%	10.4%	7.5%	12.4%	12.7%	15.0%	14.6%	21.4%	18.8%	12.1%	18.8%	26.4%	7.6%	20.0%	26.3%	10.99%
mi23_2194.11	4.7%	8.4%	9.8%	9.2%	5.8%	13.7%	18.5%	12.3%	8.0%	13.6%	10.8%	12.2%	12.3%	18.8%	17.9%	17.2%	21.2%	16.7%	13.2%	13.3%	15.8%	10.96%
mi23_2194.21	3.1%	6.8%	15.7%	8.0%	8.3%	9.3%	16.2%	11.3%	5.5%	16.7%	9.6%	15.0%	14.6%	26.8%	9.8%	16.2%	18.8%	15.3%	22.6%	20.0%	15.8%	10.96%
mi23_2194.28	5.5%	5.8%	11.8%	9.2%	6.7%	14.6%	22.5%	12.3%	8.0%	9.9%	8.3%	16.3%	8.5%	18.8%	12.5%	11.1%	23.5%	19.4%	17.0%	8.9%	36.8%	10.96%
da02_2197.47	6.9%	6.1%	16.6%	8.0%	5.0%	10.6%	18.9%	9.0%	5.5%	11.1%	5.7%	11.6%	11.5%	35.7%	9.8%	17.2%	20.0%	20.8%	5.7%	17.8%	15.8%	10.94%
mi23_2194.01	5.1%	8.2%	9.3%	8.8%	9.6%	13.3%	17.6%	9.9%	9.6%	11.7%	10.8%	12.2%	12.3%	19.6%	15.2%	11.1%	16.5%	27.8%	13.2%	20.0%	21.1%	10.94%
...																						
mi17_2188.03	4.4%	5.4%	8.4%	8.4%	5.8%	12.0%	13.1%	9.9%	8.5%	8.6%	12.1%	19.7%	10.8%	18.8%	10.7%	13.1%	17.7%	13.9%	9.4%	20.0%	10.5%	9.52%
da02_2197.22	5.6%	7.0%	6.7%	6.4%	6.3%	9.3%	14.4%	9.4%	7.5%	11.7%	9.6%	12.9%	14.6%	18.8%	14.3%	11.1%	12.9%	20.8%	17.0%	13.3%	15.8%	9.49%
mi17_2188.12	5.8%	6.5%	7.6%	8.4%	6.3%	9.7%	16.7%	9.0%	8.0%	8.6%	10.2%	12.9%	10.0%	17.0%	11.6%	12.1%	15.3%	16.7%	15.1%	15.6%	21.1%	9.47%
mi17_2188.43	4.9%	4.2%	10.1%	5.2%	7.9%	10.2%	16.7%	10.4%	8.0%	6.8%	8.9%	15.0%	8.5%	12.5%	15.2%	14.1%	18.8%	25.0%	20.8%	11.1%	10.5%	9.44%
mi17_2188.30	5.1%	7.2%	8.4%	8.8%	3.3%	12.0%	16.7%	9.9%	5.5%	9.9%	5.1%	13.6%	8.5%	17.0%	11.6%	15.2%	15.3%	18.1%	9.4%	13.3%	15.8%	9.21%
da02_2197.37	4.7%	7.2%	11.0%	5.6%	5.4%	11.5%	10.8%	10.4%	6.0%	9.9%	7.6%	10.2%	10.0%	17.0%	15.2%	13.1%	17.7%	12.5%	17.0%	13.3%	26.3%	9.18%
mi23_2194.29	3.8%	5.6%	8.2%	7.6%	6.7%	8.9%	12.2%	10.4%	5.0%	11.7%	11.5%	8.8%	10.8%	19.6%	8.9%	14.1%	20.0%	27.8%	9.4%	13.3%	31.6%	9.08%
mi17_2188.11	4.9%	4.0%	9.3%	8.0%	7.5%	9.3%	13.5%	6.6%	6.5%	11.7%	10.2%	11.6%	10.0%	19.6%	8.9%	18.2%	22.4%	12.5%	15.1%	8.9%	10.5%	9.03%
da02_2197.30	5.1%	5.6%	9.8%	8.8%	4.6%	9.7%	15.8%	10.4%	8.0%	12.4%	10.2%	8.8%	10.8%	12.5%	3.6%	8.1%	21.2%	12.5%	7.6%	8.9%	31.6%	8.90%
mi23_2194.25	1.5%	2.8%	6.5%	3.6%	2.1%	2.2%	5.4%	4.7%	2.0%	4.3%	3.2%	8.2%	6.2%	8.0%	6.3%	5.1%	11.8%	11.1%	7.6%	8.9%	15.8%	4.39%

Table 4. Alternative FW majority haplotypes are carried in a variety of patterns by individual marine fish. Out of the total number of possible FW majority haplotypes identified at Case 1, 2 and 3 loci in the MiFW1 versus OC comparison, the proportion that each marine fish harbors is listed for each linkage group. The total number of possible FW haplotypes is listed first for each chromosome, which are ordered left to right from the highest to lowest number of these loci. Some chromosomes, such as LG XV, have relatively few loci with high FW-OC divergence, so there are few possible FW haplotypes on them. Others, such as LG IV and VII have very large numbers of these loci. Cells are heat map colored to highlight where individuals harbor a higher (red) or lower (yellow) proportion of the possible FW haplotypes on a given chromosome. Individuals are ranked from those that carry the most to the fewest FW haplotypes, and of the 131 marine fish sampled, representative blocks of fish at the top, middle and bottom of this range are shown.

Discussion

Heterogeneous genomic divergence occurs in just decades despite gene flow

The heterogeneous divergence across the stickleback genome that we documented in these very young populations is unlikely to have occurred simply by neutral demographic processes or by the action of background selection in genomic regions of reduced recombination (63-66). The stickleback genome is nearly dichotomous in these newly formed populations, with clustered regions of highly significant genetic differentiation on a background of no divergence. Furthermore, this genomic mosaic is replicated across independently derived populations, which is not expected under neutral processes or background selection alone. The most plausible explanation is that strong, alternative selection between the marine and freshwater habitats has rapidly changed allele frequencies in only a subset of genomic regions.

These so-called “genomic islands” (67-69) may be created by the action of local selection that sorts ancestral variation either in the absence of contemporary gene flow,

or with continuing gene flow via a process termed ‘divergence hitchhiking’ (9, 70, 71). Because of the very recent origin of the freshwater populations on the uplift islands, discriminating these two processes is difficult. However, some characteristics of these populations argue for the opportunity for gene flow during divergence since the 1964 earthquake. Marked reduction in body size and bony lateral armor plates is typical of freshwater stickleback (38), and we documented previously that many uplift ponds on Middleton house only small fish with minimal armor (35). Several low-lying ponds still allow influx of marine stickleback, however, and harbor both low-armored and well-armored fish as well as some phenotypic and genetic intermediates (35). Even here, despite cohabitation and evidence of hybridization, the same divergent genome segments seen in older, more isolated freshwater populations (46, 72) manifest when sympatric marine and freshwater morphological ecotypes are compared against each other (seen in F_{ST} in Fig 2). In addition, these regions of divergence align with defined blocks of linkage disequilibrium (LD) that stretch over megabases and mirror the LD pattern of an *in silico* “mock sympatric population” in which data from allopatric marine and freshwater samples are artificially combined (Fig S10). The LD pattern implies that such ponds are zones of contact between alternatively adapted fish.

Strong alternative and divergent selection in both marine and freshwater habitats

Historically, a central (but not exclusive; see (73, 74)) focus of stickleback phenotypic and population genomic studies has been on the action of natural selection in freshwater habitats as marine fish colonize them. However, the present study underscores the action of similarly strong divergent selection in marine habitats. The reuse of ancient haplotypes can be seen in comparisons among the independently colonized islands. In fact, we observed as much as 99% of freshwater majority haplotypes shared between freshwater populations are at RAD loci where marine and freshwater populations have strongly alternative haplotypes. Although it has been predicted that a pool of freshwater-adapted haplotypes must be rare but present in and dispersed by the marine population (e.g., (26, 75)), such haplotypes had so far not been quantified with a genome-wide approach. In a sample of only 131 phenotypically marine stickleback collected along the shores of the uplifted islands, we found a large proportion of freshwater majority haplotypes that had been identified at divergent loci in island pond populations (Fig 7). As would be predicted if many freshwater-adapted

haplotypes are selected against in the sea, we found them to be rare and most often carried heterozygously in marine stickleback (Fig S9). Nonetheless, the allele frequency spectrum of freshwater majority haplotypes borne by the marine populations (Fig 8) argues that continuous and substantial gene flow from myriad freshwater habitats maintains appreciable levels of these genotypes despite selection against them in the sea.

Together, our data and other recent studies in stickleback argue for a view of the stickleback species as a metapopulation, with strong divergent selection occurring in both marine and freshwater habitats despite bi-directional gene flow between them. This view contrasts with a more traditional, ‘raceme’ or ‘bottle-brush’ concept of stickleback, with a core of marine populations that harbor the majority of genetic diversity in this species, while freshwater populations are the bottle-brush ‘whiskers’, ephemeral in nature (so-called “evolutionary dead-ends”). In this asymmetric model, the freshwater populations draw on the genetic diversity stored in the large marine population, but themselves wink out of existence having had little influence on the genetic composition of the species as a whole. In contrast, our data argue strongly that both population types serve as reservoirs for maintaining genetic diversity through sustained bi-directional gene flow that is sorted by divergent selection in both habitats, in other words, through a long-term dynamic of migration-selection balance. Although the relative impact of any particular freshwater population may be slight, the aggregated millions of freshwater populations could have a significant impact on the pool of standing genetic variation in marine populations. This is particularly likely when, as has been documented, haplotypes that are identical by descent are selected repeatedly in independently derived freshwater populations.

A large proportion of the stickleback genome is affected by selection during divergence

By comparing patterns of freshwater-marine divergence in the young island populations against post-glacial mainland populations, we find that as much as 15% of the genome in common is affected by alternative adaptation to ponds and lakes versus the sea (Table 1; Table S2; Fig S2). By comparison, Jones *et al.* (62) sequenced 20 genomes in a geographically broad survey of marine and freshwater stickleback, and found that less than 0.5% of the genome fell in a phylogenetic dichotomy of freshwater and marine

alleles. The findings from these two studies are of course strikingly divergent, and several factors may explain this 30-fold difference between them in the inferred proportion of the genome involved in alternative adaptation. Importantly, differences in the methodologies employed between the studies in terms of geographic breadth versus population genomic depth of coverage suggest that the findings are likely compatible and provide different and complementary information on divergence of stickleback between marine and freshwater habitats. A specific subset of selected loci – those that are both presumably ancient and house strictly alternative haplotypes - can be revealed in a broad geographic survey of few individuals, whereas polygenic selection may involve more subtle changes in allele frequencies at multiple loci (76, 77) that can only be detected with deeper population genomic sampling. Similarly, the immediate effects of linked selection can be revealed by this kind of sampling, particularly in very young populations such as on these uplifted islands.

While we might expect to see broader regions of divergence in these young freshwater ponds because of the effects of linkage (Table 1), others investigating stickleback populations that could be of similar age to those we studied here reported much less total divergence across the genome. Terekhanova *et al.* (42) studied marine-freshwater genomic divergence along the shores of the White Sea (Russia) in populations that were of known or inferred age ranging from several decades to several centuries. Those of known age — in two flooded rock quarries — were artificially seeded with only two or forty individual stickleback from both marine and freshwater habitats, making it difficult to assess the relevance of the findings to the dynamics of natural populations. They also focused on four freshwater populations that were naturally formed but of less certain independence and age, inferred from projections of sea level rate of change. The proportion of the genome estimated to be affected by marine-freshwater divergence in these Russian populations is far less than what we report for young populations on the uplift islands or even for much older Alaskan populations (Fig S11; Table 1; Table S2). Regional differences between the White Sea and Alaskan stickleback could account for these broad differences. Alternatively, the small biological sample sizes from the White Sea populations, and subsequent analytical approaches that this under-sampling necessitated (42), could have led to a significant underestimate of the extent of linked selection there.

In contrast, because of the high level of biological sampling at the level of individuals and populations, we provide clear evidence that at least a quarter of the stickleback genome is initially affected by the action of strong divergent natural selection, and that this effect is replicated independently across populations within and among islands. A key, novel finding that emerges from such a sufficiently powered population genomic study is an unexpected change in haplotype diversities in each habitat. Regions of reduced haplotype diversity in the marine populations correspond to most of the genomic segments elevated in F_{ST} (Fig 4), a pattern that may have gone unrecognized in previous oceanic-freshwater comparisons because of the primary use of SNPs (43, 72) or sampling that does not allow accurate diversity estimation within populations (42, 48). For example, Terekhanova *et al.* (42) report that nucleotide diversity in the ocean is higher than in fresh water within regions of divergence. This important difference may be due to sampling (Fig S11) or to uncertainties in using nucleotide diversity as a proxy for haplotype diversity. Regions of reduced absolute and relative haplotype diversity in the sea highlights an under-appreciated likelihood of strong directional selection in both habitats, not just in fresh water.

That haplotype diversity is relatively higher in freshwater fish across the regions of divergence perhaps runs counter to an intuition that hard selective sweeps could reduce haplotype diversity in nascent freshwater populations relative to genetically diverse holarctic marine populations. One potential explanation for elevated diversity in these young populations is that freshwater-adapted haplotypes have not completely displaced marine haplotypes, as they are in a transitory state due to limited time since founding or they are in a stable equilibrium due to lack of isolation from the sea. A high magnitude of F_{ST} divergence, however, can also manifest through a relative increase in freshwater haplotype diversity *via* selection of an ancient allele embedded in a small number of syntenic genotypes, a process that would transiently increase the relative haplotype diversity in neighboring genomic regions in freshwater populations (78). The majority of selected genomic regions may comprise such old alleles - for example a region highlighted in the recurrent evolution of armor loss (55, 75, 79) - which drive the broadly parallel trait changes in freshwater fish, but these alleles exist on diverse genomic backgrounds.

An important extension of our findings is that linked selection across these large regions of divergence may drive significant correlated phenotypic responses in both maintenance of marine-adapted traits in the sea and alteration of traits following freshwater colonization, even if the overall proportion of genetic variants that are subject to direct selection in common across populations is relatively small. Our observation supports the role of extended linked selection affecting a large proportion of the stickleback genome, and accords with a documented pattern of greater phenotypic variation among geographically localized freshwater populations relative to the globally distributed marine population (36). This broader phenotypic range has been attributed to the diversity of limnological habitats (36). Findings from these Alaskan island populations suggest an additional hypothesis: that divergent phenotypic responses could result from direct selection on a subset of alleles and indirect selection on a diverse array of linked variants, leading to a correlated response in both mean and variance of traits. We studied only a dozen populations on three islands, but the geographic region uplifted in 1964 likely led to many additional freshwater stickleback populations that could provide a key resource for future studies.

The transporter hypothesis and the size and ordering of genomic changes in the first few generations after colonization

Migration-selection balance and the reuse of ancient haplotypes underlie the popularly cited "Transporter Hypothesis" to interpret parallel evolution in threespine stickleback (26). These mechanisms predict that the freshwater adapted genome should enter new habitats piecemeal from marine colonizers over a period of time and that populations in the earliest stages of colonization might therefore initially have captured only a subset of these ancient haplotypes. Though capped at a maximum age by the 1964 earthquake, the island populations studied here probably span a range of younger ages. Montague's SE ponds that are perched on a narrow peninsula less than two meters above sea level (Fig S12), and similarly, the only freshwater pond (Da04) on Danger Island sits only one meter above sea level. These ponds might house the youngest freshwater populations in this analysis, as they are vulnerable to periodic inundation from the sea, which could cause osmotic stress, local extinction of adapting freshwater stickleback, or an influx of marine stickleback leading to recolonization or genetic introgression. These populations show only a subset of the pattern of divergence seen in the other freshwater groups

(e.g., (51) and this study). For instance, an inversion-spanning region on LG I (62) alone stands out in Da04 (Fig 2; Fig 6). While Middleton populations and SW Montague have a sharp peak of F_{ST} coinciding with this small inversion, in Da04 the inversion sits in a broader window of divergence that includes reduced haplotype diversity, the clear signature of an ongoing selective sweep. The sparser pattern of genomic divergence in these precarious habitats may reflect initial random sampling of freshwater-adapted genotypes from the sea, and could also imply that these subsets of loci are sufficient for initial persistence of freshwater pioneers. Along those lines, others have remarked on the location at the LG I inversion of *atp1a1*, a gene encoding a sodium-potassium ATPase important for osmoregulation in fishes (80-82).

Though fewer chromosomes show divergence from marine fish in MoSE, it nonetheless has elevated F_{ST} across the full length of the massive region of parallel divergence on LG IV (Fig 2). This is perhaps a testament to a low recombination rate across this part of the chromosome (65), making it more likely for LG IV haplotypes alternative to majority marine haplotypes to enter new ponds as a unit. If true, this genomic architecture should speed the time that the parallel pattern of a diverged freshwater-adapted genome can be reassembled, from what might be a relatively small number of large genomic segments.

Most of the alternative freshwater majority haplotypes we identified from MiFW1, MiFW2 and MoSW were also present in marine stickleback, where in most individuals they occur in low numbers across the genome. However, a subset of the marine fish in our sample carried freshwater haplotypes densely on some linkage groups (Table 4). These individuals could be early generation hybrids between freshwater and oceanic stickleback. Some of the freshwater haplotypes may also be in regions that resist recombination, such as those that lie in inversions or across large parts of LG IV and LG VII (65). Discovery of these patterns of freshwater majority haplotypes in a relatively small sample of fish from the sea suggests that the rapid time of adaptation to new freshwater habitats could hinge on “jackpot” carriers, and that a freshwater adapted genome might effectively be reassembled by migrants in new habitats one or more chromosomes at a time. Together, our findings provide a potential answer to the question of how enough low frequency standing genetic variation can be reconstituted in small freshwater ponds even with few initial stickleback colonists: luck of the draw. A

small number of migrants may in total carry enough chromosome-scale freshwater haplotypes to colonize newly opened habitats. Even if the initial colonists do not have an entire set of variants, subsequent migrants may quickly complete the winning hand.

General lessons about migration-selection balance and the genomic architecture of rapid adaptation

Striking genomic patterns of strong differentiation on a background of virtually no divergence argue that changes during adaptation to fresh water are established promptly, and that mainland populations in much older lakes could have experienced most of their present-day genomic shifts, and associated phenotypic evolution, in just their first decades after colonization. The speed with which freshwater and marine populations diverge across a large number of genomic spans - on the order of decades - could be an outcome of the metapopulation structure particular to stickleback. On the other hand, the stickleback radiation may provide more general evolutionary genomic lessons because adaptation from standing genetic variation on such a short time frame may be common in the wild. This could be particularly true in spatially structured species that experience variable selection and gene flow that maintains ancient diversity, a scenario we argue is more likely to be the rule than the exception in natural populations of most species, as has been documented, for example, in the study of Darwin's finches (83). If common, then for many species standing genetic variation maintained and shaped by molecular evolutionary processes operating over thousands or millions of generations may permit population genomic changes and organismal evolution to new environments over just decades. Fully understanding rapid adaptive dynamics may often require a deep molecular evolutionary perspective of the haplotypes' changing frequencies.

The long-term association of the stickleback species with both oceanic and freshwater habitats leads to an intriguing hypothesis. Although each bout of adaptation may only require decades, the genomic regions underlying the change are a much older balanced polymorphism. If so, the molecular evolutionary history of locally adapted genomic regions across the metapopulation might actually be quite deep. Alternatively-adapted freshwater and marine haplotypes in genomic regions with lower recombination rates (65) may be thousands or millions of years old (75) and lead to increased absolute divergence even in the absence of contemporary gene flow. In fact, longer read RAD-

seq data from our laboratory on many fewer individuals align with this hypothesis. Coalescent analysis of these Sanger length sequences sampled across the genome demonstrates that in many of the genomic regions with the most significant relative divergence of allele frequencies, the haplotypes also show more ancient evolutionary divergence that is millions of years old (84). Therefore, a deep evolutionary history underlies the ability of stickleback to adapt so rapidly. Similar analyses of other species with populations experiencing divergent selection and gene flow could show whether this phenomenon is unique to stickleback, or a common theme in nature.

Materials and methods

Biological collections

Freshwater ponds and marine sites for stickleback collection (Fig S1) were chosen on the basis of maps and aerial imagery created before and after 1964. Fish were trapped during the summers of 2005 (Montague and Middleton Islands), 2010 (Danger and Middleton Islands) and 2011 (Montague and Middleton Islands), and processed as described in Lescak *et al.* (35). All research was approved by the IACUCs of the University of Alaska and the University of Oregon. Fish were collected under Alaska Department of Fish and Game permits SF-2005-020, SF-2010-029, and SF-2011-153, as well as Oregon Department of Fish and Wildlife scientific taking permits OR2007_3495 and 13920.

Restriction site Associated DNA sequencing library preparation and sequence analysis

RAD-seq libraries for 1057 fish were created as described in Lescak *et al.* (35) using endonuclease *SbfI*, and sequenced to 101 nucleotides (including a 6 nt inline barcode) on an Illumina HiSeq 2000 platform (Table S5). In addition to these data, RAD sequences from an additional 98 fish from Cushman Slough, Oregon (USA), which were first published in Catchen *et al.* (85), were included in analyses for a total of 1155 fish. Individual fish were represented by an average of $1,265,744 \pm 13,864$ SE sequences each, of which $1,225,729 \pm 13,744$ SE sequences (97%) were aligned to the reference genome, and these produced an average of $43,174 \pm 165.7$ SE loci with an average

depth of coverage of 27x using the software *Stacks* (86-88), version 1.46 (Table S6). Raw sequence data were demultiplexed according to barcode and filtered for quality using the `process_radtags` module (`-c`, `-q`, and `-r` parameters) in *Stacks*. Cleaned reads were aligned to the stickleback reference genome (version BROADS1, Ensembl release 86) using the `mem` module from the BWA aligner (89) with default settings. *Stacks* were assembled, SNPs were called and a population-level catalog was constructed by executing the `pstacks` (`-m 3`), `cstacks` (`-aligned`), and `sstacks` (`-aligned`) modules. Population-level corrections were made to the data by running `rxstacks` (`-model_type bounded`, `-bound_high 0.1`, `-prune_haplo`, and `-conf_lim 0.25` parameters), and then rebuilding and matching to the catalog with `cstacks` (`-aligned`) and `sstacks` (`-aligned`). Final filtering was done with the `populations` module, as well as the calculation of population genetic statistics and associated kernel smoothing.

Our full data set comprises 48,307 RAD markers encompassing a total of 266,902 SNPs. We applied two general filtering steps for most analyses, one when considering populations individually, one when considering Structure-defined population clusters. In the former, for a locus to be included in downstream analyses we required it to be present in at least 20 populations (`-p 20` parameter) and in 75% of individuals of each population (`-r 0.75`), and we applied a minor allele frequency (MAF) filter of 1% (`-maf 0.01`). This resulted in 29,911 RAD markers and 74,556 SNPs. In the latter analysis, a locus had to be present in 9 populations and in 75% of individuals, with a MAF of 1%. This resulted in 30,285 RAD markers and 75,476 SNPs. Sequence data are available in the NCBI Short Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under accession XXXXX.

To compare the Middleton Island samples against those found in Cook Inlet previously, we downloaded sequences used in Hohenlohe *et al.* (51) from the NCBI Sequence Read Archive using accession SRA010788.9. In addition, we added 54 more recently sequenced individuals from Rabbit Slough and 24 from High Ridge Lake. As these data stem from sequencing runs using different versions of Illumina technology, we retained 94 individuals that had read lengths of at least 49bp and an average depth of coverage of 10x (Table S7). These data were aligned to the stickleback reference

genome and processed with *Stacks*, in the same manner as the Middleton Island data. We grouped the individuals into freshwater and marine clusters, required loci to be present in both clusters and at a frequency of 75% among individuals in either separate cluster, and applied a MAF of 6%.

Statistical approach

The populations that were shown to be united into genetic groupings by Structure analysis in Lescak *et al.* (35) were treated together here as groups rather than as single populations. Also, the three geographically proximate Cook Inlet freshwater populations sampled in Hohenlohe *et al.* (51) are here treated as a group to increase sampling robustness for a regional comparison. Latitude and longitude, respectively, for these three populations are 61.6 and -149.75 (Bear Paw Lake), 61.72 and -150.12 (Boot Lake), and 61.93 and -150.97 (Mud Lake). See Fig S1 for information on island sampling locations and population groupings. Population genetic statistics were calculated using the populations program in *Stacks* (86-88). Calculations of haplotype diversity, Φ_{ST} and F_{ST}' are based on the state of each ~100bp RAD locus instead of individual SNPs. The combination of fixed and variable nucleotides that make up each locus are considered a set of haplotypes at the locus. While SNP-based statistics require bi-allelic nucleotide positions, the haplotype-based statistics can have multiple haplotypes per locus, with a distance measured between haplotypes, providing a richer set of states. Genome-wide circle plots (*aka*, Large Hadron Collider Plots) were generated by a Python program, `lhc_plot.py`, which is distributed with the *Stacks* package. Population genetic statistics were kernel smoothed along the genome using default smoothing parameters in *Stacks* software.

Calculating linkage disequilibrium (LD)

All SNPs output from *Stacks* were imported into Beagle (version 3.3.2) to phase SNPs across each chromosome of each population. We then transferred the resulting Beagle output into the phasedstacks program of *Stacks* to calculate D' , the measure of LD (90, 91), and values of D' were plotted as heat maps in Gnuplot 4.6 patchlevel 4 (<http://gnuplot.info>).

Calculating Gene and Haplotype Diversity

We calculated two separate diversity measures for each RAD locus in each population. A RAD locus contains one or more haplotypes, which are the physically phased combination of SNPs present at that locus across the population. The first measure, Gene Diversity (90, 91), can be thought of as the probability that two randomly chosen haplotypes at that locus are different. Given the number of samples n , and the i^{th} haplotype p_i , it is calculated as:

$$\hat{H} = \frac{n}{n-1} (1 - \sum_i p_i^2)$$

The second measure, Haplotype Diversity (90, 91), is similar to Gene Diversity, but it is scaled by the substitution distance, d , between the two randomly chosen haplotypes:

$$\hat{\pi} = \frac{n}{n-1} \sum_i \sum_j p_i p_j \hat{d}_{ij}$$

Haplotype diversity measures were kernel smoothed along the genome using default smoothing parameters in *Stacks* software. For haplotype diversity difference, the continuous smoothed values for the marine group were subtracted from those for the freshwater group, and these raw values were plotted on a nonlinear color scale (e.g., as in Fig 3E).

Segmenting the genome via Hidden Markov Model

We used the General Hidden Markov Model Library (GHMM (92), version 0.9-rc3) to segment the genome using haplotype-based F_{ST}' . In order to make the F_{ST}' measures discrete, we binned them into four categories; we took the F_{ST}' measures from LG IV of the FW2 versus marine data set and hierarchically clustered them in R (R Core Team, 2013). Using the Python bindings for GHMM, we designed a two-state HMM representing *diverged* and *not-diverged*. Initial emission and transition parameters were defined using F_{ST}' values in regions of marked divergence between freshwater and oceanic populations on LG IV and in regions of nearly zero divergence on LG XV, respectively. For each pairwise comparison (FW2 vs. marine, FW1 vs. marine, SW vs. marine, and Cook Inlet vs. marine) the HMM was trained on LG IV (diverged) and LG XV (not diverged) by executing the Baum-Welch algorithm of the GHMM library on the FW2 vs. marine F_{ST}' values. After training, each linkage group was segmented into the

diverged and *not-diverged* states using a custom program `segment.py`. Finally, we compared the set of states for each of the four data sets (FW2, FW1, SW, CI vs. marine) to find the diverged regions in common using our `shared_segments.py` program. Plots of the HMM data were generated using Gnuplot 4.6 patchlevel 4 (Fig 3; Fig S2).

Defining cases from haplotype data

Classification of cases was done using custom Python scripts. For each locus in an analysis, we summed the haplotypes in order from the haplotype with the largest frequency in a population to the smallest. The set of ordered haplotypes that occupied at least 60% of the haplotypes at that locus in that population were considered *majority haplotypes*. Loci were classified as *Case 1* when one or more majority haplotype/s occurred in a marine population, but the corresponding freshwater population had a single, alternative majority haplotype (i.e., one that is *not* a majority haplotype in the marine population). Loci were classified as *Case 2* when the marine population had one or more majority haplotype/s at a locus and the freshwater population had one alternative majority haplotype, and one or more haplotype/s also in the majority in the marine population. Loci were classified as *Case 3* when the marine population had one or more majority haplotype/s at a locus, but the corresponding freshwater population had two or more alternative majority haplotypes and no majority haplotypes in common with the marine population. Loci were classified as *Case 4* when the freshwater population had only a subset of majority haplotypes in the marine population. Circular genomic plots of cases were created with the `lhcp_plot.py` script while the linear plots of cases (e.g., Fig 6) were created with a custom Python script.

Acknowledgements

We thank M. S. Christy, S. A. Hatch, B. Lohman, V. M. Padula, L. Smayda, and K. Walton for assistance with fieldwork logistics and fish collection, as well as T. Wilson and M. Currey for technical help. We also thank P. Hohenlohe for discussions during early stages of the development of the project. We thank members of the W.A.C., P. C.

Phillips, and M. A. Streisfeld Laboratories at the University of Oregon, for discussions and critical comments.

References

1. Turner TL, Hahn MW, Nuzhdin SV. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 2005;3(9):e285. doi: 10.1371/journal.pbio.0030285. PubMed PMID: 16076241; PubMed Central PMCID: PMC1182689.
2. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. *Mol Ecol.* 2009;18(3):375-402. doi: 10.1111/j.1365-294X.2008.03946.x. PubMed PMID: 19143936.
3. Feder JL, Flaxman SM, Egan SP, Nosil P. Hybridization and the build-up of genomic divergence during speciation. *J Evol Biol.* 2013;26(2):261-6. doi: 10.1111/jeb.12009. PubMed PMID: 23324002.
4. Hey J. Isolation with migration models for more than two populations. *Mol Biol Evol.* 2010;27(4):905-20. doi: 10.1093/molbev/msp296. PubMed PMID: 19955477; PubMed Central PMCID: PMC2877539.
5. Feder JL, Egan SP, Nosil P. The genomics of speciation-with-gene-flow. *Trends Genet.* 2012;28(7):342-50. doi: 10.1016/j.tig.2012.03.009. PubMed PMID: 22520730.
6. Akerman A, Burger R. The consequences of gene flow for local adaptation and differentiation: a two-locus two-deme model. *J Math Biol.* 2014;68(5):1135-98. doi: 10.1007/s00285-013-0660-z. PubMed PMID: 23532261; PubMed Central PMCID: PMC3948587.
7. Yeaman S, Otto SP. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution.* 2011;65(7):2123-9. doi: 10.1111/j.1558-5646.2011.01277.x. PubMed PMID: 21729066.
8. Yeaman S, Whitlock MC. The genetic architecture of adaptation under migration-selection balance. *Evolution.* 2011;65(7):1897-911. doi: 10.1111/j.1558-5646.2011.01269.x. PubMed PMID: 21729046.

9. Feder JL, Nosil P. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*. 2010;64(6):1729-47. doi: 10.1111/j.1558-5646.2010.00943.x. PubMed PMID: 20624183.
10. Levin D. Metapopulations: an arena for local speciation. *Journal of Evolutionary Biology*. 1995. doi: 10.1046/j.1420-9101.1995.8050635.x/pdf.
11. Barrett RD, Schluter D. Adaptation from standing genetic variation. *Trends Ecol Evol*. 2008;23(1):38-44. doi: 10.1016/j.tree.2007.09.008. PubMed PMID: 18006185.
12. Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol*. 2013;28(11):659-69. doi: 10.1016/j.tree.2013.08.003. PubMed PMID: 24075201; PubMed Central PMCID: PMC3834262.
13. Hendry AP, Kinnison MT. Perspective: The pace of modern life: Measuring rates of contemporary microevolution. *Evolution*. 1999;53(6):1637-53. doi: Doi 10.2307/2640428. PubMed PMID: WOS:000084703100001.
14. Hendry AP, Wenburg JK, Bentzen P, Volk EC, Quinn TP. Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. *Science*. 2000;290(5491):516-9. doi: 10.1126/science.290.5491.516. PubMed PMID: 11039932.
15. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol*. 2010;20(4):R208-15. doi: 10.1016/j.cub.2009.11.055. PubMed PMID: 20178769; PubMed Central PMCID: PMC2994553.
16. Burke MK, Liti G, Long AD. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. *Mol Biol Evol*. 2014;31(12):3228-39. doi: 10.1093/molbev/msu256. PubMed PMID: 25172959; PubMed Central PMCID: PMC4245818.
17. Thompson JN. Rapid evolution as an ecological process. *Trends in Ecology & Evolution*. 1998;13(8):329-32. doi: 10.1016/s0169-5347(98)01378-0.
18. Hendry AP, Kinnison MT. An introduction to microevolution: rate, pattern, process. *Genetica*. 2001;112-113:1-8. PubMed PMID: 11838760.

19. Bell MA, Aguirre WE, Buck NJ. Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution*. 2004;58(4):814-24. doi: 10.2307/3449278. PubMed PMID: 15154557.
20. Lotterhos KE, Schaal SM. Genome scans for the contemporary response to selection in quantitative traits. *Mol Ecol*. 2014;23(18):4435-7. doi: 10.1111/mec.12853. PubMed PMID: 25208503.
21. Lucek K, Lemoine M, Seehausen O. Contemporary ecotypic divergence during a recent range expansion was facilitated by adaptive introgression. *J Evol Biol*. 2014;27(10):2233-48. doi: 10.1111/jeb.12475. PubMed PMID: 25228272.
22. Kimura M, Ota T. Theoretical aspects of population genetics. *Monogr Popul Biol*. 1971;4:1-219. PubMed PMID: 5162676.
23. Orr HA. The genetic theory of adaptation: a brief history. *Nat Rev Genet*. 2005;6(2):119-27. doi: 10.1038/nrg1523. PubMed PMID: 15716908.
24. Pritchard JK, Di Rienzo A. Adaptation - not by sweeps alone. *Nat Rev Genet*. 2010;11(10):665-7. doi: 10.1038/nrg2880. PubMed PMID: 20838407; PubMed Central PMCID: PMC4652788.
25. Radwan J, Babik W. The genomics of adaptation. *Proc Biol Sci*. 2012;279(1749):5024-8. doi: 10.1098/rspb.2012.2322. PubMed PMID: 23097510; PubMed Central PMCID: PMC3497254.
26. Schluter D, Conte GL. Genetics and ecological speciation. *Proc Natl Acad Sci U S A*. 2009;106 Suppl 1(Supplement_1):9955-62. doi: 10.1073/pnas.0901264106. PubMed PMID: 19528639; PubMed Central PMCID: PMC4652788.
27. Hahn MW. Toward a selection theory of molecular evolution. *Evolution*. 2008;62(2):255-65. doi: 10.1111/j.1558-5646.2007.00308.x. PubMed PMID: 18302709.
28. Slatkin M, Wiehe T. Genetic hitch-hiking in a subdivided population. *Genet Res*. 1998;71(2):155-60. PubMed PMID: 9717437.

29. Wood TE, Burke JM, Rieseberg LH. Parallel genotypic adaptation: when evolution repeats itself. *Genetica*. 2005;123(1-2):157-70. PubMed PMID: 15881688; PubMed Central PMCID: PMCPMC2442917.
30. Flaxman SM, Feder JL, Nosil P. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*. 2013;67(9):2577-91. doi: 10.1111/evo.12055. PubMed PMID: 24033168.
31. Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc Natl Acad Sci U S A*. 2013;110(19):E1743-51. doi: 10.1073/pnas.1219381110. PubMed PMID: 23610436; PubMed Central PMCID: PMCPMC3651494.
32. Bell MA. Evolutionary phenetics and genetics. The threespine stickleback, *Gasterosteus aculeatus*, and related species. In: Turner BJ, editor. *Evolutionary Genetics of Fishes* 1984. p. 431-527.
33. Bell MA, Foster SA. *The evolutionary biology of the threespine stickleback*. Oxford ; New York: Oxford University Press; 1994. xii, 571 p.
34. Cresko WA, McGuigan KL, Phillips PC, Postlethwait JH. Studies of threespine stickleback developmental evolution: progress and promise. *Genetica*. 2007;129(1):105-26. doi: 10.1007/s10709-006-0036-z. PubMed PMID: 16897450.
35. Lescak EA, Bassham SL, Catchen J, Gelmond O, Sherbick ML, von Hippel FA, et al. Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proc Natl Acad Sci U S A*. 2015;112(52):E7204-12. doi: 10.1073/pnas.1512020112. PubMed PMID: 26668399; PubMed Central PMCID: PMCPMC4702987.
36. Walker JA, Bell MA. Net evolutionary trajectories of body shape evolution within a microgeographic radiation of threespine sticklebacks (*Gasterosteus aculeatus*). *J Zool*. 2000;252:293-302. PubMed PMID: WOS:000165318600003.
37. Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver MD, et al. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell*. 2007;131(6):1179-89. doi: 10.1016/j.cell.2007.10.055. PubMed PMID: 18083106; PubMed Central PMCID: PMC2900316.

38. Albert AY, Sawaya S, Vines TH, Knecht AK, Miller CT, Summers BR, et al. The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution*. 2008;62(1):76-85. doi: 10.1111/j.1558-5646.2007.00259.x. PubMed PMID: 18005154.
39. Miller CT, Glazer AM, Summers BR, Blackman BK, Norman AR, Shapiro MD, et al. Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait Loci. *Genetics*. 2014;197(1):405-20. doi: 10.1534/genetics.114.162420. PubMed PMID: 24652999; PubMed Central PMCID: PMC4012497.
40. Milligan-Myhre K, Small CM, Mittge EK, Agarwal M, Currey M, Cresko WA, et al. Innate immune responses to gut microbiota differ between oceanic and freshwater threespine stickleback populations. *Dis Model Mech*. 2016;9(2):187-98. doi: 10.1242/dmm.021881. PubMed PMID: 26681746; PubMed Central PMCID: PMCPMC4770144.
41. Thompson CE, Taylor EB, McPhail JD. Parallel evolution of lake-stream pairs of threespine sticklebacks (*Gasterosteus*) inferred from mitochondrial dna variation. *Evolution*. 1997;51(6):1955-65. doi: Doi 10.2307/2411016. PubMed PMID: WOS:000071652200023.
42. Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, Schluter D, et al. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol*. 2004;2(5):E109. doi: 10.1371/journal.pbio.0020109. PubMed PMID: 15069472; PubMed Central PMCID: PMCPMC385219.
43. Cresko WA, Amores A, Wilson C, Murphy J, Currey M, Phillips P, et al. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc Natl Acad Sci U S A*. 2004;101(16):6050-5. doi: 10.1073/pnas.0308479101. PubMed PMID: 15069186; PubMed Central PMCID: PMC395921.
44. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, et al. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*. 2004;428(6984):717-23. doi: 10.1038/nature02415. PubMed PMID: 15085123.
45. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Jr., Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin

- alleles. *Science*. 2005;307(5717):1928-33. doi: 10.1126/science.1107239. PubMed PMID: 15790847.
46. Shapiro MD, Bell MA, Kingsley DM. Parallel genetic origins of pelvic reduction in vertebrates. *Proc Natl Acad Sci U S A*. 2006;103(37):13753-8. doi: 10.1073/pnas.0604706103. PubMed PMID: 16945911; PubMed Central PMCID: PMC1564237.
47. Shapiro MD, Summers BR, Balabhadra S, Aldenhoven JT, Miller AL, Cunningham CB, et al. The genetic architecture of skeletal convergence and sex determination in ninespine sticklebacks. *Curr Biol*. 2009;19(13):1140-5. doi: 10.1016/j.cub.2009.05.029. PubMed PMID: 19500990; PubMed Central PMCID: PMC1564237.
48. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 2010;6(2):e1000862. doi: 10.1371/journal.pgen.1000862. PubMed PMID: 20195501; PubMed Central PMCID: PMC2829049.
49. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*. 2012;484(7392):55-61. doi: 10.1038/nature10944. PubMed PMID: 22481358; PubMed Central PMCID: PMC3322419.
50. Feulner PG, Chain FJ, Panchal M, Eizaguirre C, Kalbe M, Lenz TL, et al. Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Mol Ecol*. 2013;22(3):635-49. doi: 10.1111/j.1365-294X.2012.05680.x. PubMed PMID: 22747593.
51. Terekhanova NV, Logacheva MD, Penin AA, Neretina TV, Barmintseva AE, Bazykin GA, et al. Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genet*. 2014;10(10):e1004696. doi: 10.1371/journal.pgen.1004696. PubMed PMID: 25299485; PubMed Central PMCID: PMC4191950.
52. Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, et al. Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. *PLoS Genet*.

2016;12(2):e1005887. doi: 10.1371/journal.pgen.1005887. PubMed PMID: 26925837; PubMed Central PMCID: PMC4771382.

53. Schluter D, Clifford EA, Nemethy M, McKinnon JS. Parallel evolution and inheritance of quantitative traits. *Am Nat.* 2004;163(6):809-22. doi: 10.1086/383621. PubMed PMID: 15266380.

54. Boughman JW, Rundle HD, Schluter D. Parallel evolution of sexual isolation in sticklebacks. *Evolution.* 2005;59(2):361-73. doi: 10.2307/3448926. PubMed PMID: 15807421.

55. DeFaveri J, Shikano T, Shimada Y, Goto A, Merila J. Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*). *Evolution.* 2011;65(6):1800-7. doi: 10.1111/j.1558-5646.2011.01247.x. PubMed PMID: 21644964.

56. Kaeuffer R, Peichel CL, Bolnick DI, Hendry AP. Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution.* 2012;66(2):402-18. doi: 10.1111/j.1558-5646.2011.01440.x. PubMed PMID: 22276537; PubMed Central PMCID: PMC4499166.

57. Hirase S, Ozaki H, Iwasaki W. Parallel selection on gene copy number variations through evolution of three-spined stickleback genomes. *BMC Genomics.* 2014;15:735. doi: 10.1186/1471-2164-15-735. PubMed PMID: 25168270; PubMed Central PMCID: PMC4159527.

58. Hohenlohe P, Bassham S, Etter P, Stiffler N, Johnson E, Cresko W. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 2010;6(2):e1000862. doi: 10.1371/journal.pgen.1000862. PubMed PMID: 20195501.

59. Plafker G. The Alaska Earthquake, March 27, 1964: regional effects – the tectonics of the March 27, 1964 Alaska Earthquake. 1969;543-I:1-73.

60. Plafker G, Rubin M. Uplift history and earthquake recurrence as deduced from marine terraces on Middleton Island, Alaska. 1978;78(943):687-721.

61. Gelmond O, von Hippel FA, Christy MS. Rapid ecological speciation in three-spined stickleback *Gasterosteus aculeatus* from Middleton Island, Alaska: the roles of selection and geographic isolation. *J Fish Biol.* 2009;75(8):2037-51. doi: 10.1111/j.1095-8649.2009.02417.x. PubMed PMID: 20738670.
62. Kimmel CB, Ullmann B, Walker C, Wilson C, Currey M, Phillips PC, et al. Evolution and development of facial bone morphology in threespine sticklebacks. *Proc Natl Acad Sci U S A.* 2005;102(16):5791-6. doi: 10.1073/pnas.0408533102. PubMed PMID: 15824312; PubMed Central PMCID: PMC556121.
63. Kimmel CB, Cresko WA, Phillips PC, Ullmann B, Currey M, von Hippel F, et al. Independent axes of genetic variation and parallel evolutionary divergence of opercle bone shape in threespine stickleback. *Evolution.* 2012;66(2):419-34. doi: 10.1111/j.1558-5646.2011.01441.x. PubMed PMID: 22276538; PubMed Central PMCID: PMC4039416.
64. Kimmel CB, Hohenlohe PA, Ullmann B, Currey M, Cresko WA. Developmental dissociation in morphological evolution of the stickleback opercle. *Evol Dev.* 2012;14(4):326-37. doi: 10.1111/j.1525-142X.2012.00551.x. PubMed PMID: 22765204; PubMed Central PMCID: PMC3394458.
65. Charlesworth B. Measures of divergence between populations and the effect of forces that reduce variability. *Mol Biol Evol.* 1998;15(5):538-43. PubMed PMID: 9580982.
66. Holsinger KE, Weir BS. Genetics in geographically structure populations: defining, estimating and interpreting F_{ST} . *Nat Rev Genet.* 2009;10:639-50.
67. Jakobsson M, Edge MD, Rosenberg NA. The relationship between F_{ST} and the frequency of the most frequent allele. *Genetics.* 2013;193(2):515-28. doi: 10.1534/genetics.112.144758. PubMed PMID: 23172852; PubMed Central PMCID: PMCPMC3567740.
68. Jones F, Grabherr M, Chan Y, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature.* 2012;484(7392):55-61. doi: 10.1038/nature10944. PubMed PMID: 22481358.

69. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics*. 1993;134(4):1289-303. PubMed PMID: 8375663; PubMed Central PMCID: PMCPMC1205596.
70. Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, Kawakami T, et al. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*. 2012;491(7426):756-60. doi: 10.1038/nature11584. PubMed PMID: 23103876.
71. Roesti M, Moser D, Berner D. Recombination in the threespine stickleback genome-- patterns and consequences. *Mol Ecol*. 2013;22(11):3014-27. doi: 10.1111/mec.12322. PubMed PMID: 23601112.
72. Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res*. 2015;25(11):1656-65. doi: 10.1101/gr.196485.115. PubMed PMID: 26355005; PubMed Central PMCID: PMCPMC4617962.
73. Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, et al. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc Lond B Biol Sci*. 2012;367(1587):343-53. doi: 10.1098/rstb.2011.0198. PubMed PMID: 22201164; PubMed Central PMCID: PMCPMC3233711.
74. Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*. 2014;344(6190):1410-4. doi: 10.1126/science.1253226. PubMed PMID: 24948738.
75. Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, et al. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*. 2015;350(6267):1493-8. doi: 10.1126/science.aac9927. PubMed PMID: 26680190; PubMed Central PMCID: PMCPMC4700518.
76. Via S, West J. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol*. 2008;17(19):4334-45. doi: 10.1111/j.1365-294X.2008.03921.x. PubMed PMID: 18986504.

77. Via S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos Trans R Soc Lond B Biol Sci.* 2012;367(1587):451-60. doi: 10.1098/rstb.2011.0260. PubMed PMID: 22201174; PubMed Central PMCID: PMCPMC3233719.
78. Roesti M, Gavrillets S, Hendry AP, Salzburger W, Berner D. The genomic signature of parallel adaptation from shared genetic variation. *Mol Ecol.* 2014;23(16):3944-56. doi: 10.1111/mec.12720. PubMed PMID: 24635356; PubMed Central PMCID: PMC4122612.
79. Guo B, DeFaveri J, Sotelo G, Nair A, Merila J. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol.* 2015;13:19. Epub 2015/04/11. doi: 10.1186/s12915-015-0130-8. PubMed PMID: 25857931; PubMed Central PMCID: PMCPMC4410466.
80. Defaveri J, Merila J. Evidence for adaptive phenotypic differentiation in Baltic Sea sticklebacks. *J Evol Biol.* 2013;26(8):1700-15. Epub 2013/07/19. doi: 10.1111/jeb.12168. PubMed PMID: 23859314.
81. Mather K. Polygenic inheritance and natural selection. *Biological Reviews.* 1943;18(1):32-64.
82. Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature.* 2010;467(7315):587-90. doi: 10.1038/nature09352. PubMed PMID: 20844486.
83. Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res.* 1997;70(2):155-74. PubMed PMID: 9449192.
84. McCairns RJ, Bernatchez L. Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution.* 2010;64(4):1029-47. doi: 10.1111/j.1558-5646.2009.00886.x. PubMed PMID: 19895556.
85. Jampol LM, Epstein FH. Sodium- potassium-activated adenosine triphosphatase and osmotic regulation by fishes. *Am J Physiol.* 1970;218(2):607-11. Epub 1970/02/01. PubMed PMID: 4244054.

86. Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, et al. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr Biol.* 2012;22(1):83-90. doi: 10.1016/j.cub.2011.11.045. PubMed PMID: 22197244; PubMed Central PMCID: PMC3319444.
87. Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* 2017;27(6):1004-15. Epub 2017/04/27. doi: 10.1101/gr.212522.116. PubMed PMID: 28442558; PubMed Central PMCID: PMC5453315.
88. Nelson TC, Cresko WA. Ancient genomic variation underlies recent and repeated ecological adaptation. *bioRxiv.* 2017.
89. Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, et al. The population structure and recent colonization history of Oregon threespine stickleback determined using RAD-seq. *Molecular Ecology.* 2012.
90. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda).* 2011;1(3):171-82. doi: 10.1534/g3.111.000240. PubMed PMID: 22384329; PubMed Central PMCID: PMC3276136.
91. Hohenlohe PA, Catchen J, Cresko WA. Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. *Methods Mol Biol.* 2012;888:235-60. doi: 10.1007/978-1-61779-870-2_14. PubMed PMID: 22665285.
92. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol.* 2013;22(11):3124-40. doi: 10.1111/mec.12354. PubMed PMID: 23701397; PubMed Central PMCID: PMC3936987.
93. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv e-prints [Internet].* 2013 March 1, 2013; 1303. Available from: <http://adsabs.harvard.edu/abs/2013arXiv1303.3997L>.
94. Nei M. *Molecular Evolutionary Genetics.* New York, NY: Columbia University Press; 1987.

95. Weir BS. Genetic Data Analysis II. Sunderland, MA.: Sinauer Associates; 1996.
96. Schliep A, Costa IG, Steinhoff C, Schonhuth A. Analyzing gene expression time-courses. IEEE/ACM Trans Comput Biol Bioinform. 2005;2(3):179-93. doi: 10.1109/TCBB.2005.31. PubMed PMID: 17044182.