

Genomic changes underlying repeated niche shifts in an adaptive radiation

David A. Marques,^{1,2,3,4,5}  Felicity C. Jones,^{6,7,8}  Federica Di Palma,^{9,10}  David M. Kingsley,^{6,7} 
 and Thomas E. Reimchen¹ 

¹Department of Biology, University of Victoria, Victoria, BC V8W 3N5, Canada

²Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern CH-3012, Switzerland

³Department of Fish Ecology and Evolution, Centre for Ecology, Evolution, and Biogeochemistry, Swiss Federal Institute of Aquatic Science and Technology (EAWAG), Eawag - Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum CH-6047, Switzerland

⁴Natural History Museum Basel, Basel CH-4051, Switzerland

⁵E-mail: david.marques@bs.ch

⁶Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA

⁷Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA

⁸Friedrich Miescher Laboratory of the Max Planck Society, Tübingen 72076, Germany

⁹Earlham Institute, Norwich NR4 7UZ, United Kingdom

¹⁰Department of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom

Received September 6, 2021

Accepted March 9, 2022

In adaptive radiations, single lineages rapidly diversify by adapting to many new niches. Little is known yet about the genomic mechanisms involved, that is, the source of genetic variation or genomic architecture facilitating or constraining adaptive radiation. Here, we investigate genomic changes associated with repeated invasion of many different freshwater niches by threespine stickleback in the Haida Gwaii archipelago, Canada, by resequencing single genomes from one marine and 28 freshwater populations. We find 89 likely targets of parallel selection in the genome that are enriched for old standing genetic variation. In contrast to theoretical expectations, their genomic architecture is highly dispersed with little clustering. Candidate genes and genotype-environment correlations match the three major environmental axes predation regime, light environment, and ecosystem size. In a niche space with these three dimensions, we find that the more divergent a new niche from the ancestral marine habitat, the more loci show signatures of parallel selection. Our findings suggest that the genomic architecture of parallel adaptation in adaptive radiation depends on the steepness of ecological gradients and the dimensionality of the niche space.

KEY WORDS: Adaptive radiation, genomics, Haida Gwaii, niche shift, niche space, threespine stickleback.

Adaptive radiations, when single lineages rapidly evolve phenotypic and ecological diversity, are an important source of biodiversity (Simpson 1953; Schluter 2000). Often, similar niches are occupied in replicated fashion via convergent evolution (Kocher 2004; Arendt and Reznick 2008), offering an outstanding oppor-

tunity to understand genomic mechanisms underlying adaptation to a niche (Martin and Richards 2019).

Theory predicts a simple genetic architecture for traits under selection in an adaptive radiation (Gavrilets and Vose 2005), a small total number of loci involved (Gavrilets 2004; Gavrilets and Vose 2005, 2009; Gavrilets and Losos 2009), and arrangement in clusters (Kirkpatrick and Barton 2006; Yeaman 2013). Repeated shifts between two niches have indeed often been

[Correction added on 13 MAY 2022, after first online publication: CSAL funding statement has been added.]

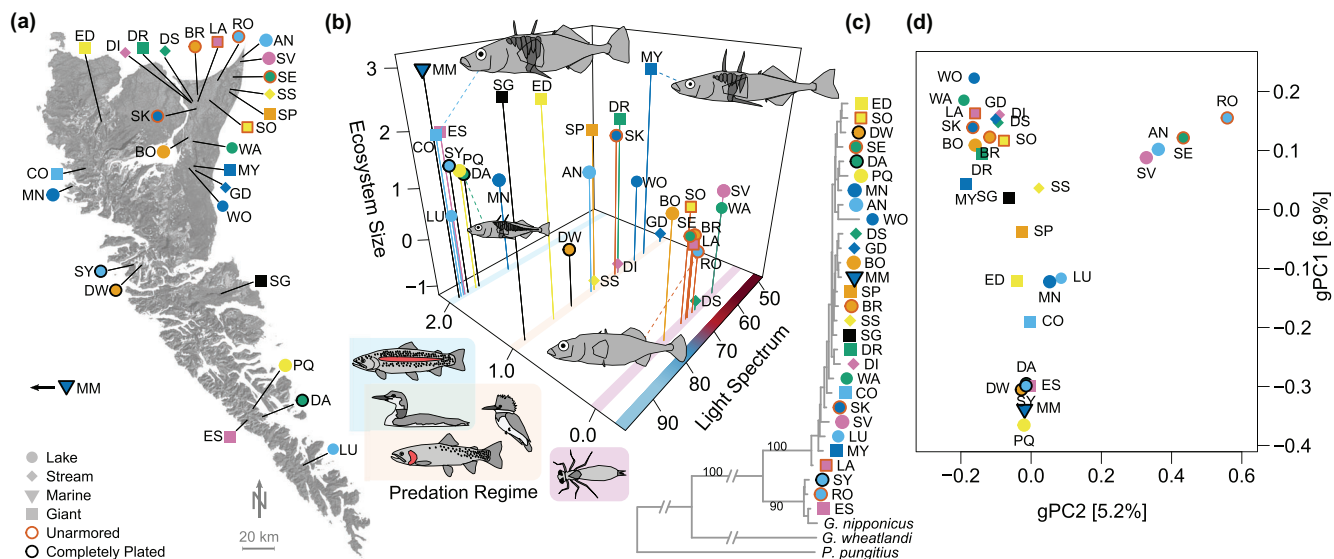


Figure 1. Adaptive radiation of threespine stickleback on Haida Gwaii, Canada. (a) Geography of the 28 freshwater populations, with habitat, gigantism, and plate morph indicated by symbol shapes and border colors. (b) Niche occupation of 28 freshwater and one marine stickleback populations along three major ecological axes: predation regime, light spectrum, and ecosystem size (Reimchen et al. 2013). Similar niches in this multidimensional niche space were colonized independently in different watersheds. Predation regime: invertebrate (0: purple) or vertebrate dominated, with main predators either being cutthroat trout *Oncorhynchus clarkii* (1: orange) or rainbow trout *O. mykiss* (2: blue). Stickleback drawings depict typical morphologies of selected populations, representing extremes in morphospace (Reimchen et al. 2013). Axis units: light spectrum = percent light transmission at 400 nm wavelength, ecosystem size = \log_{10} -transformed lake area in hectares. (c) Whole mitochondria maximum-likelihood phylogeny of Haida Gwaii stickleback reveals the presence of Japan Sea stickleback (*Gasterosteus nipponicus*) derived haplotypes in three freshwater populations. Outgroups are blackspotted stickleback (*G. wheatlandi*) and nine-spined stickleback (*Pungitius pungitius*). Branch labels: bootstrap support (%) shown if >80% support. “Broken” branches are long branches shortened for better visualization of the Haida Gwaii haplotypes. (d) Genomic variation mirrors geography along a SW-NE gradient in a principal component analysis of genome-wide autosomal SNPs (gPC). Square brackets contain percentages variance explained by each principal component. See Table S1 for full population names.

associated with simple genetic architectures of traits (Schemske and Bradshaw 1999; Peichel et al. 2001; Chan et al. 2010; Reed et al. 2011; Nadeau et al. 2012; Wright et al. 2013; Lamichhane et al. 2015; Nadeau et al. 2016; Sheehan et al. 2016; Kratochwil et al. 2018; Nosil et al. 2018) or clustering of loci in recombination suppressed genomic regions such as in inversions (Feder et al. 2003; Hoffmann et al. 2004; Joron et al. 2006; Lowry and Willis 2010; Samuk et al. 2017). However, these expectations contrast with most traits having a complex genetic architecture with many loci of mostly small effect (Orr 1998; Flint and Mackay 2009; Sella and Barton 2019).

The source and amount of genetic variation available may matter for a radiating lineage. Adaptive de novo mutations arriving in a lineage are limited by population size, time, and mutation rate, especially in organisms with comparatively small population sizes (Smith 1976; Lanfear et al. 2014; Rousselle et al. 2020). “Old” genetic variation present as standing genetic variation or derived from admixture is often involved in rapid habitat shifts and speciation in adaptive radiations (Feder et al. 2003; Rieseberg et al. 2003; Lamichhane et al. 2015; Meier et al. 2017; Richards and Martin 2017; Nelson and Cresko 2018; Marques

et al. 2019). But most theoretical and empirical studies have investigated adaptation to a single new niche or a two-niche contrast. It thus remains an open question how genomic architecture and the source of genetic variation may promote or constrain adaptive radiation into a complex, multidimensional niche space (but see, e.g., Kautt et al. 2020).

In this study, we investigate adaptive radiation of threespine stickleback into a multidimensional niche space made of diverse freshwater habitats on the Haida Gwaii archipelago off the Canadian west coast (Reimchen et al. 2013) (Fig. 1a,b). As ice sheets retreated ~12,000 years ago, a diversity of freshwater habitats emerged, ranging from large lakes to small ponds and streams, from deeply stained blackwater to clear lake habitats with divergent predator faunas, each colonized by threespine stickleback (Moodie and Reimchen 1976). Today, the Haida Gwaii adaptive radiation shows the largest phenotypic disparity known within the threespine stickleback species, with parallel phenotype-environment associations (Reimchen et al. 1985; Reimchen 1994; Reimchen 1995; Deagle et al. 1996; Reimchen and Nosil 2002, 2004, 2006; Reimchen et al. 2013; Reimchen et al. 2019), trait utility experiments (Reimchen 1980,

1983, 1992; McDonald et al. 1995; Reimchen 2000; Flamarique et al. 2013), an evolution experiment (Leaver and Reimchen 2012; Marques et al. 2018), and known genetic basis of involved traits (Marques et al. 2017a; Peichel and Marques 2017) demonstrating that adaptation along three major axes—predation regime, light spectrum, and ecosystem size—explains most phenotypic variation (Reimchen et al. 2013). A reconstruction of the phylogeographic history of threespine stickleback in the Haida Gwaii archipelago (Deagle et al. 2013) has shown that similar habitats have been invaded in replicate in different watersheds and similar morphologies have evolved convergently in those habitats (Fig. 1).

Large lakes, small lakes, ponds, and streams, varying from highly stained (blackwater) through to clear water, are the main niches occupied by stickleback on Haida Gwaii (Fig. 1b). In large lakes, “giant” stickleback with long dorsal spines and few lateral body plates evolved. Gigantism and long spines protect from gape-limited predators, whereas few lateral plates are beneficial where avian predators causing compression injuries dominate (Reimchen 1983, 2000; Reimchen et al. 2013). Where fish predators causing puncture wounds dominate, stickleback either show few lateral plates where predator evasion into the dark of blackwater through a fast start is possible (Reimchen 1992; Law and Blake 1996; Reimchen 2000; Bergstrom 2002; Leinonen et al. 2011), whereas in clearwater lakes dominated by fish predators stickleback retained a full cover of lateral plates as in the ancestral marine habitat (Reimchen 2000; Reimchen et al. 2013). In small, shallow, blackwater ponds dominated by grappling invertebrate predators, stickleback repeatedly evolved reduced body armor up to the complete loss of lateral plates, pelvic girdle, and spines facilitating their postcapture escape (Reimchen 1980; Reimchen et al. 1985; Reimchen and Nosil 2002, 2004). Feeding ecology and morphology also evolved along these axes, with stickleback feeding on zooplankton in large lakes or on benthic invertebrates in shallow lakes and streams and respective parallel evolution of body shape, gape size, and gill raker number (Moodie and Reimchen 1976; Reimchen et al. 1985; Spoljaric and Reimchen 2007). Sexual selection and habitat properties also interact, for example, blackwater stickleback have repeatedly replaced red nuptial color with black pigmentation as sexual signal (Reimchen 1989; McDonald et al. 1995; Flamarique et al. 2013). Furthermore, highly divergent stream ecotypes have evolved in parapatry with lake populations in multiple watersheds (Reimchen et al. 1985; Deagle et al. 2012).

An experiment previously explored phenotypic and genomic responses to a single shift between extreme niches, by moving giant stickleback from a large dystrophic blackwater lake into a small fishless eutrophic clearwater pond, revealing rapid evolution into the predicted direction in morphology (Leaver and Reimchen 2012) and a highly dispersed genomic architecture of

adaptation even in a single niche shift across a multidimensional niche space (Marques et al. 2018). The generality of this result is, however, unclear, that is, whether niche shifts along fewer axes or along shallower ecological gradients may involve simpler genomic architectures of adaptation.

Here, we quantify the genomic architecture of parallel adaptation associated with divergent niche shifts in the Haida Gwaii adaptive radiation, using single genomes from 28 freshwater populations and a marine population. We identify genomic targets of parallel selection across the radiation with haplotype-based selection statistics and a neutral demographic control. We assess whether old standing genetic variation was involved in parallel adaptation and use this system to explore how the genomic architecture of parallel adaptation covaries with the nature and extent of niche differentiation.

Methods

SAMPLING SCHEME, PERMITS, AND GENOME SEQUENCING

Stickleback were sampled between 1993 and 2012 on Haida Gwaii (Reimchen et al. 2013) with previously published genomes, one per population, from 24 lakes, four streams, and one marine population (Marques et al. 2017a, 2018) included in this study (Table S1). Collection was conducted with minnow traps, euthanization with MS-222 under Ministry of Environment of British Columbia permits SM09-51584 and SM10-62059, University of Victoria aquatic unit standard operating procedure OA2003, following British Columbia’s guidelines for scientific fish collection (Reimchen et al. 2013). We extracted DNA from fin samples following Peichel et al. (2001), generated fragment libraries following Jones et al. (2012b), and paired-end sequenced libraries on an Illumina GAII machine (Illumina Inc., San Diego CA, USA) resulting in 77 bp reads and a mean sequencing depth of $6.42\times$ (Table S1). We aligned reads with BWA version 0.5.9 (Li and Durbin 2009) and parameters “-q 5 -l 32 -k 2 -o 1” to the threespine stickleback Broad S1 reference genome (Jones et al. 2012b), recalibrated base qualities with GATK version 1.4 tools “CountCovariates” and “TableRecalibration” (McKenna et al. 2010), and called variants and genotypes with GATK’s “UnifiedGenotyper” with default SNP and indel calling parameters. We filtered variants with GATK’s “VariantFiltration” based on read position rank sum test statistic ≥ -20 , allele-specific strand bias test statistic ≤ 200 , or quality normalized by depth value ≥ 2 and recalibrated variants with GATK’s “VariantRecalibrator” and “ApplyRecalibration” using a VQSRLD cutoff of 98.5%. With BCFTOOLS version 1.3.1 (Li et al. 2009), we removed indels, variants with quality < 45 , with < 4 reads per allele, with > 2 alleles, and with mean sequencing depth ≥ 9.51 ($=$ mean depth $+ 1.5\times$ interquartile range depth

distribution). We lifted biallelic SNPs to an improved reference (Glazer et al. 2015) using PICARD version 2.2.1 (Broad Institute 2018) and also realigned recalibrated reads with bwa version 0.7.2. We used VCFTOOLS version 0.1.15 (Danecek et al. 2011) to remove genotypes with <4 reads and SNPs with >50% missing data, resulting in a final SNP set with 19.4% missing genotypes across 29 individuals, and computed its transition to transversion ratio with BCFTOOLS version 1.7. For selection and demography reconstruction, we phased the final SNP dataset with read-backed phasing and genotype imputation implemented in SHAPEIT version 2.r790 (Delaneau et al. 2013), using molecular phase information for 7.5% of the heterozygote genotypes and phase-informative reads covering 30.9% of all heterozygous sites. Genome coordinates reported correspond to Glazer et al. (2015).

ANALYSIS OF GENOME-WIDE VARIATION

We investigated population structure in a principal components (PC) analysis. We converted genotype likelihoods of autosomal SNPs with <10% missing data to the BEAGLE PL format with VCFTOOLS version 0.1.14, computed the site allele frequency spectrum in ANGSD version 0.9.15 (Nielsen et al. 2012; Fumagalli et al. 2013; Korneliussen et al. 2014), the sample covariance matrix using NGSCOVAR (Fumagalli et al. 2013), and performed eigenvalue decomposition in R version 3.3.1 (R Development Core Team 2015). We genotyped three previously described inversions (Jones et al. 2012b) by repeating PC analysis for these genomic regions only. We reconstructed a mitochondrial phylogeny using all individuals and the complete mitochondrial sequence of blackspotted stickleback (*Gasterosteus wheatlandi*) and Japan Sea stickleback (*G. nipponicus*) as outgroups (GenBank accessions AB445130.1 and AB445129.1, Kawahara et al. 2009). We called consensus mitochondrial haplotypes from BAM files with SAMTOOLS, BCFTOOLS, and “vcfutils.pl,” all part of the SAMTOOLS version 1.7 toolkit (Li et al. 2009) and aligned them with outgroups by hand in BIOEDIT version 7.0.5.3 (Hall 1999). We used mitochondrial gene coordinates (Jones et al. 2012b) and PARTITIONFINDER2 version 2.1.1 (Lanfear et al. 2017) to partition the mitochondrion by rates of molecular evolution and to identify the best substitution model. We used RAXML version 8.2.4 (Stockwell et al. 2003) with the “GTR + I + G” model and the partition file to reconstruct a maximum likelihood phylogenetic tree and its robustness from 100 bootstrap replicates.

IDENTIFYING SIGNATURES OF PARALLEL SELECTION

We identified genomic signatures indicative of parallel selection across the Haida Gwaii adaptive radiation from haplotype-based selection statistics iHS (Voight et al. 2006) and H12 (Garud et al. 2015) computed on the whole adaptive radiation—an approach previously shown to successfully recover targets of repeated se-

lection in this radiation (Marques et al. 2017a). These statistics detect incomplete hard or soft selective sweeps based on haplotype length decay (iHS) or haplotype frequency spectra (H12) in a sample (Voight et al. 2006; Garud et al. 2015). They are thus sensitive to unusually long and common haplotypes that may have swept to high frequency in multiple populations of the adaptive radiation due to parallel selection on the same haplotype. We use a demographic control to identify the likely targets of parallel selection. We first reconstructed the demography of the 29 stickleback populations, simulated neutral whole genome datasets under the inferred demographic model to obtain neutral distributions for iHS and H12, and identified observed iHS and H12 values exceeding neutral expectations. We designated genomic regions with an enrichment of both iHS and H12 outliers as candidate targets of parallel selection in the Haida Gwaii adaptive radiation.

We used the pairwise sequentially Markovian coalescent (PSMC) to estimate population sizes of the 29 populations each (Li and Durbin 2011), as it allows single genome reconstruction and correction for low sequencing depth. We aggregated individual genomes into 100-bp bins with heterozygote presence/absence information, masking genotypes with depth <4, and depth >20. We ran PSMC with default settings except a custom time segmentation pattern of “1*6 + 7*2 + 1*4 + 1*6” to minimize uncertainty in the first time segments. We corrected effective population size (N_e) estimates with a false negative rate (FNR_{het}) for detecting heterozygotes as recommended for low sequencing depth data (Li and Durbin 2011). We estimated FNR_{het} from genotype depth distributions: $FNR_{het} = \sum_{i=4}^{i=20} n_i * (\frac{1}{2})^{(i-1)}$, with n_i being the number of genotypes with depth i in an individual. We converted generations into years assuming a generation time of 1.5 years (Marques et al. 2018) and a mutation rate of 1.7×10^{-8} originating from gene alignment phylogenetic estimation (Feulner et al. 2015). We combined N_e trajectories from the 29 populations into a demographic model assuming a simultaneous split of all freshwater populations from the marine population 12,000 years or 8000 generations ago (Reimchen et al. 2013) and the marine N_e trajectory prior to that time.

We generated 1000 neutral whole genome datasets under this demographic model with FASTSIMCOAL2 version 2.6.0.3 (Excoffier et al. 2013) and the “FTC” recombination map (Glazer et al. 2015), by dividing the 21 stickleback chromosomes into 34 recombination rate bins (10 bins between 0 and 0.1 cM/Mb, 9 between 0.1 and 1 cM/Mb, 15 between 1 and 16 cM/Mb) with average recombination rates from bin boundaries. A custom python script was used to convert FASTSIMCOAL2 output into VCF format (“arp2vcf.py,” see DATA ARCHIVING).

We computed iHS and H12 for phased SNPs with a minor allele frequency >5% in both observed and simulated datasets with HAPBIN version 1.3.0 (Macleán et al. 2015) and H12 scripts by Garud et al. (2015). We estimated H12 in 201-SNP windows

to obtain a window size of approximately 10 kb as in Garud et al. (2015). We computed 99.9% quantiles for simulated absolute iHS ($|iHS|$) and H12 values in sliding windows, using the R-function “quantile” with window and step sizes 1 Mb and 250 kb for $|iHS|$ and 200 kb and 50 kb for H12. For observed $|iHS|$ and H12 values, we assessed whether they exceeded neutral expectations using the R-function “smooth.spline” with “spar = 0.5” to predict local neutral 99.9% quantiles. We then tested for an enrichment of outlier SNPs in 50-kb sliding windows with step size 25 kb using one-sided binomial tests. After Bonferroni correction for multiple testing, we merged significantly enriched windows at $\alpha < 0.001$ within <50 kb distance into outlier regions, when $|iHS|$ and H12 outliers were aligned and <5 kb apart.

In each outlier region, we identified the SNP with the highest H12 value and used these SNPs to compute linkage disequilibrium (LD) r^2 between outlier regions in VCFTOOLS. To generate a null expectation for interchromosomal LD, we randomly picked 2046 SNPs >200 kb apart across all autosomes and the sex chromosome using VCFTOOLS. To compare mean interchromosomal LD between random and outlier region SNPs, we randomly subsampled 89 of these SNPs (\sim outlier region number) in 10,000 permutations. We clustered outlier regions into groups using the R-function “hclust,” a dissimilarity matrix (values: $1 - r^2$), and the R-function “cutree” with parameter “ $h = 0.75$.”

We tested whether potential phasing error due to statistical phasing could lead to false positive iHS and H12 outliers exceeding the expected neutral false positive rate of 0.1%. We removed phase information from 10 of the 1000 simulated whole genome datasets and phased them statistically with shapeit as described above but lacking phase-informative read information, computed iHS and H12, and identified outliers exceeding the local neutral 99.9% quantile across the genome as outlined above, for both statistically phased iHS and H12 estimates and those based on known phase from the simulated datasets. We computed Pearson’s correlation coefficient between statistically phased and known phase $|iHS|$ and H12 estimates for the same loci and the proportion of outliers in the 10 neutral datasets, once for statistically phased and known phase $|iHS|$ and H12 estimates to assess their rates of false positive outliers against the neutral distributions from all 1000 simulated whole genome datasets.

IDENTIFYING DIVERGENT GENOMIC BLOCKS

We identified genomic regions with old genetic variation from the range of absolute divergence (d_{XY}) (Nei 1987) between pairs of individuals in nonoverlapping 100-kb windows (Nelson and Cresko 2018; Marques et al. 2019). A large d_{XY} range among all comparisons of 29 populations indicates strong divergence between some but not other segregating haplotypes, while accounting for mutation rate variation that leads to correlated d_{XY} values between all pairs independent of haplotype age (Fig. 2a).

We computed the unfolded two-dimensional site frequency spectra (2D-SFS) for each pair in 100-kb windows with ANGSD (see above), but for windows specified with the “-r” option.

d_{XY} is the average number of pairwise differences between all haplotypes, or all alleles in the case of a single site and of two populations (Nei 1987), and can thus be computed from the 2D-SFS as follows. For a pairwise comparison with one individual per population, there are four haplotype comparisons between populations, with zero pairwise differences for the nonvariant site coordinates [0,0] and [2,2] in the 2D-SFS, two pairwise differences out of four comparisons for the coordinates [1,1], [1,2], [2,1], [0,1], [1,0], and four pairwise differences out of four comparisons for the coordinates [0,2] and [2,0] in the 2D-SFS. d_{XY} can be computed by multiplying genotype counts in each SFS coordinate with the number of these pairwise differences (zero, two, or four), dividing the product by the number of comparisons (four), and then dividing the sum of these quotients by the sum of all SFS entries (total number of variant and nonvariant sites sequenced, e.g., in a window). We wrote and used a custom script (“dxy_wsfs.py,” see DATA ARCHIVING) to convert 2D-SFS to d_{XY} estimates using this logic and applied it to our 100-kb window 2D-SFS estimates.

We assigned empirical quantiles of the d_{XY} -range distribution to each window using the “ecdf” function in R and tested whether iHS/H12 outlier regions are enriched for overlapping top 5% or top 1% quantile d_{XY} range windows using a chi-square test. We also tested whether top 5% quantile d_{XY} range windows were enriched for overlapping genes by permuting window positions across the genome 10,000 times. We overlapped outlier regions and top 5% d_{XY} range windows with previously published genomic regions differentiated between marine and freshwater, lake and stream, benthic and limnetic, and different freshwater stickleback ecotypes (Hohenlohe et al. 2010; Deagle et al. 2012; Jones et al. 2012a,b; Feulner et al. 2015; Roesti et al. 2015; Marques et al. 2016). We used chi-square tests to assess top 5% d_{XY} -range regions for enrichment of such ecotype categories.

CONNECTING TARGETS AND SOURCES OF PARALLEL SELECTION

We designated the longest haplotype at top H12-peak SNPs in each outlier region as “sweep haplotype” and used linear models to test whether their number or sharing can be explained by the ecology or genetic properties of populations. As ecological predictors, we used the three major environmental axes light spectrum (S), ecosystem size (E), and predation regime (P, see definitions in Fig. 1b), niche divergence from the ancestral niche (marine habitat) in a three-dimensional niche space made of these axes (ANC, Fig. 1b), niche divergence between population pairs (DIV), and isolation in niche space (ISO, median distance from all other populations in niche space). Associations with

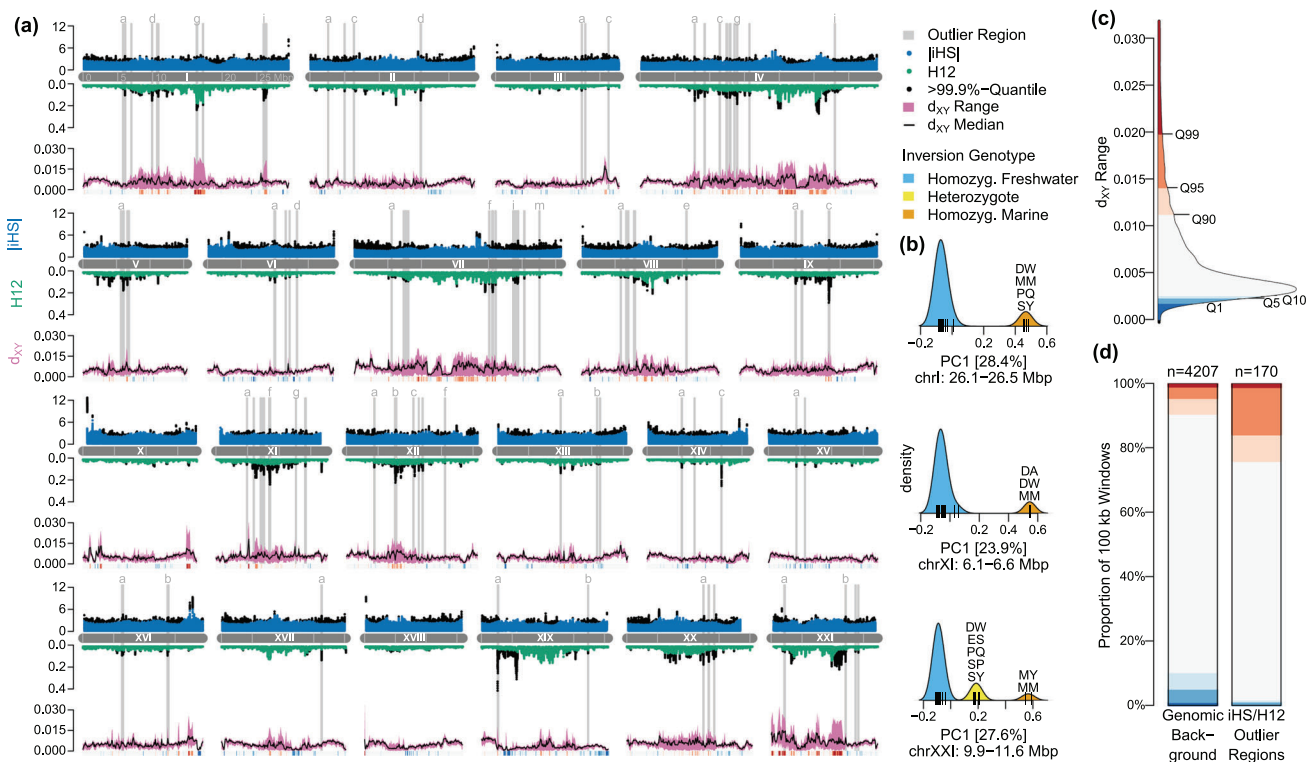


Figure 2. Dispersed genomic architecture of parallel adaptation and enrichment for old genetic variation. (a) Signatures of selective sweeps across the stickleback adaptive radiation on Haida Gwaii are distributed throughout the genome. Vertical gray bars indicate outlier regions enriched for both iHS (blue) and H12 (green) outliers (black points) at Bonferroni-corrected $\alpha < 0.001$, with iHS and H12 outliers exceeding 99.9% quantiles of neutral demographic expectations. Several blocks of old standing variation, indicated by a large range in absolute divergence (d_{XY}) between population pairs (purple area), overlap such outlier regions. Color codes below each chromosome indicate d_{XY} range quantiles shown in panel c. Roman numerals are chromosome names; letters correspond to outlier regions (see Table S2). (b) Genotype distribution of Haida Gwaii stickleback for known large inversions: “marine” haplotypes are found in several freshwater populations, in particular in fully plated freshwater populations (DA, DW, SY). First principal components of SNP genotypes in the indicated genomic interval are shown, with the percentage of variation explained in brackets. (c) Distribution of d_{XY} ranges in 100-kb windows based on all pairwise d_{XY} values between the 29 individuals, with quantiles indicated by colors. (d) Outlier regions are enriched for old standing genetic variation: 16% of the outlier regions fall into genomic regions containing old genetic variation.

S, E, and P, ANC, and DIV could indicate that the genomic architecture of parallel adaptation is influenced by each or multiple axes, and thus by the degree of niche differentiation or the dimensionality of a niche space, respectively. Association with ISO could indicate a potential detection bias, as signatures of parallel selection might have a higher detectability in populations clustered in niche space (= less isolated) experiencing similar selection regimes. As genetic predictors, we used the most recent effective population size (POP) estimated by PSMC (see above, Table S1), autosomal genomic PC1 and PC2 as proxy for geography (gPC1, gPC2), and pairwise distances in genomic PC1/PC2 space (gPC, Fig. 1d) as proxy for genetic relatedness between populations. A positive association with population size could indicate constraints to adaptive evolution in small populations (Ohta 1992). An association with geography or relatedness between populations might reflect either (a) a higher likelihood of parallel evolution due to shared standing genetic

variation (Conte et al. 2012), (b) a geographic sampling bias due to isolation by distance reducing the number and length of shared haplotypes between more distant populations, or (c) a spurious association due to spatial autocorrelation of niches known from the Haida Gwaii archipelago (Reimchen et al. 2013), whereas niche differentiation and niche space dimensionality are the drivers of genomic architecture. We included two more genetic predictors, mean sequencing depth (DEP), and the proportion of missing SNP genotypes in each individual, to test whether sequencing depth might have biased our ability to detect signatures of parallel selection. All ecological and genetic predictors were fixed effects, with the number of sweep haplotypes or number of shared sweep haplotypes as response variable, and bidirectional stepwise regression was used to maximize model likelihood in MASS version 7.3-47 (Venables and Ripley 2002). To corroborate the importance of niche space dimensionality with uncorrelated environmental axes, we repeated above analyses

with an alternative niche space defined by three ecological PC axes (ePC1-3) built from S, E, and P, and with recomputed ANC, DIV, and ISO, to remove correlation between ecological axes using the R-function “prcomp”. All predictors were z -transformed to mean zero and variance one before inclusion into linear models. Predictor P -values were false discovery rate adjusted to control for multiple testing (Benjamini and Hochberg 1995).

We identified genes overlapping each outlier region’s [iHS] and H12 selection signature and retrieved functional annotation and expression information for those from zebrafish, mouse, rat, and human databases (Howe et al. 2013; Shimoyama et al. 2015; Blake et al. 2017). We tested them for gene ontology term enrichment using the STRING database version 10 (Szklarczyk et al. 2015). We assessed the overlap of outlier regions with targets of selection in a selection experiment on Haida Gwaii (Marques et al. 2018) and outlier regions between marine-freshwater (Hohenlohe et al. 2010; Jones et al. 2012a; Jones et al. 2012b), lake-stream (Deagle et al. 2012; Feulner et al. 2015; Roesti et al. 2015; Marques et al. 2016), and limnetic-benthic (Jones et al. 2012a) ecotypes, using either exact regions when reported or outlier SNP positions ± 100 kb buffers (Hohenlohe et al. 2010; Deagle et al. 2012; Jones et al. 2012a; Roesti et al. 2015).

We assessed potential associations of outlier regions (genotypes at top H12-peak SNPs) with geography, environment, phenotype, and mtDNA haplotype with three approaches: genome-wide association (GWAS), random forest, and SAGUARO (Zamani et al. 2013). We used the same 16 untransformed response variables in all association tests (see Table S3) and replaced missing values with overall means or a reasonable replacement (“lake” depth: 1 m for streams, 100 m for the marine habitat; “lake” area: 0.1 ha for streams, 1000 ha for the marine habitat).

For GWAS, we converted autosomal SNPs to PLINK binary format using VCFTOOLS and PLINK version 1.9 (Purcell et al. 2007; Chang et al. 2015), computed a relatedness matrix on autosomal SNPs to correct for population structure, and ran univariate linear mixed models in GEMMA version 0.94.1 (Zhou and Stephens 2012). We repeated these steps for the sex chromosome XIX with the only male (see Table S1) excluded. We identified likely associations in 50-kb sliding windows with 12.5 kb step size where $-\log_{10}(P\text{-value})$ means < 2 .

We built random forests using the “randomForest” R-package version 4.6-14 (Breiman 2001) with outlier region top H12-peak SNP genotypes as predictors. We assessed variable importance stability for “ntree” values between 500 and 10,000 and found strong correlations (Spearman’s $\rho > 0.95$) between mean decrease in accuracy (MDA) and mean decrease in Gini node impurity index (Gini) at “ntree = 8000” or higher for all response variables. We optimized “mtry” with the function “tuneRF” at “ntree = 8000” for each variable, choosing “mtry”

values leading to the lowest out-of-bag (OBB) error for categorical and lowest proportion of variance explained (PVE) for continuous response variables. We used stratified random sampling for categorical variables with unequal sample sizes per group, with three (plate morph, blackwater, stream, mtDNA haplotype), six (gigantism, melanism), or seven (predation regime) individuals sampled in each iteration. We retained SNPs with a standardized MDA > 20 as potential associations, if OBB was $> 25\%$ for categorical variables or PVE was $> 20\%$ for continuous variables.

SAGUARO (Zamani et al. 2013) generated 21 local phylogenetic trees from all autosomal SNPs and produced distance matrices and SNP assignments corresponding to these 21 topologies. We generated distance matrices for all response variables and used Mantel tests to test for associations with the 21 topology distance matrices with P -values adjusted for false discovery rate (Benjamini and Hochberg 1995).

Results

GENOMES MIRROR GEOGRAPHY AND REVEAL MITONUCLEAR DISCORDANCE

Our dataset of 29 genomes from 28 Haida Gwaii freshwater and one marine population contains 6,564,500 high-quality SNPs with a transition to transversion ratio (Ts/Tv) 1.31 (Table S1, see *Methods*). Genomic variation mirrors geography of the archipelago: populations from the South-West and North-East spread along a first PC axis (gPC1) apart from the isolated populations LU and ED, whereas populations from the North-East spread along gPC2 (Fig. 1d), recapitulating earlier findings based on SNP chip data (Deagle et al. 2013). In three genomic regions on chromosomes I, XI, and XXI, where large inversions have been previously shown to be under divergent selection between marine and freshwater habitats (Jones et al. 2012b), local PCs separate individuals into two or three discontinuous clusters corresponding to inversion genotypes (Fig. 2b). Haida Gwaii freshwater populations are thus polymorphic in all three inversions, with marine-associated haplotypes notably present in freshwater populations with the marine-like full lateral plate cover phenotype (ST, DA, DW, Fig. 2b).

Mitochondria reveal a separation into two highly divergent clades (Fig. 1c), noted earlier as unusual mitochondrial haplotype diversity in Haida Gwaii populations (Oreilly et al. 1993; Orti et al. 1994; Deagle et al. 1996). One mitochondrial clade present in three Haida Gwaii lake populations (ST, RO, ES) shares a most recent common ancestor with the ~ 1 million years divergent Japan Sea stickleback (Ravinet et al. 2018) before coalescing with other threespine stickleback populations (Fig. 1c). Such mitonuclear discordance and inversions represent

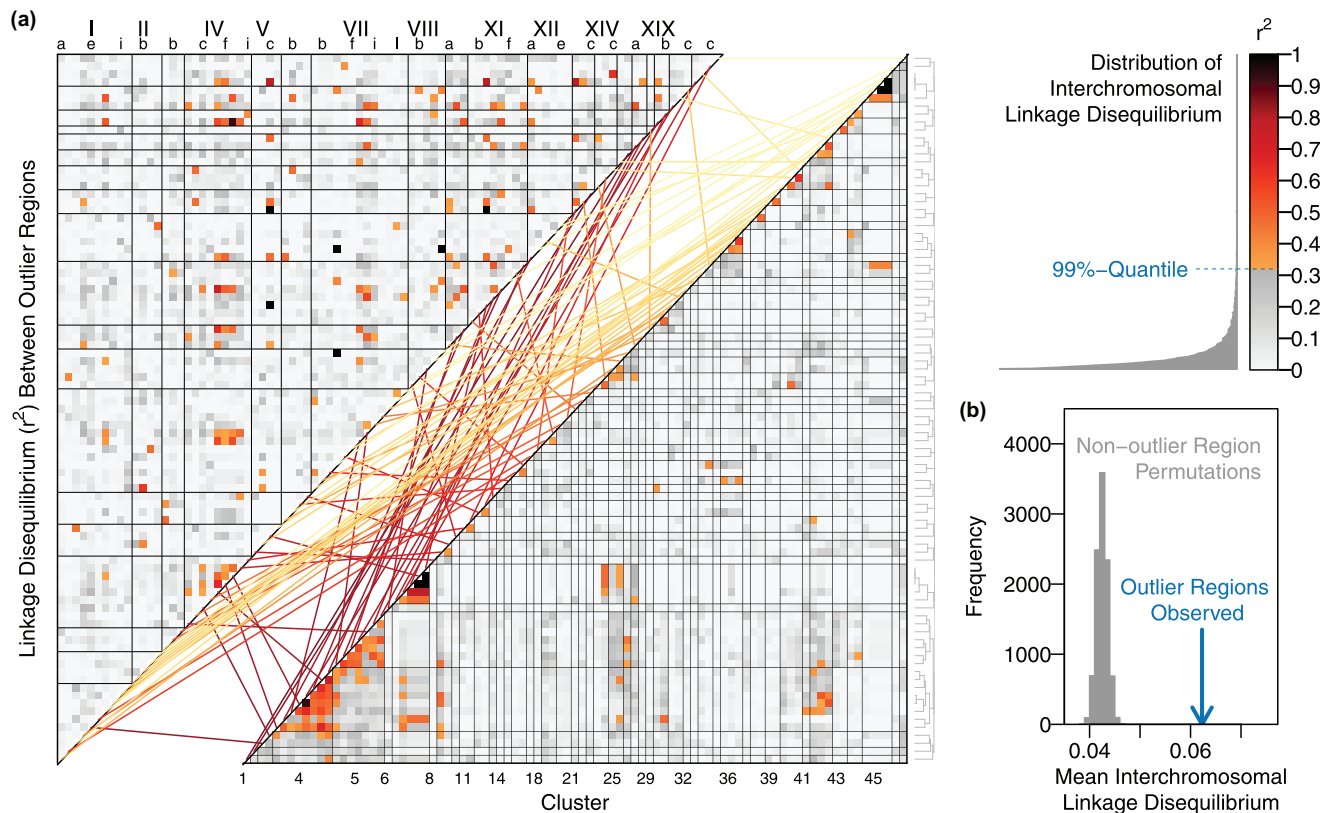


Figure 3. Linkage disequilibrium (LD) between outlier regions indicate correlated evolution between physically unlinked outlier regions. (a) LD between outlier regions sorted by chromosome and physical position (above diagonal) and into 47 clusters in a hierarchical cluster analysis (below diagonal, see cluster tree on the right), with lines connecting outlier regions and colored lines corresponding to the 47 groups. Colored r^2 values exceed the top 1% LD values among random SNPs on different chromosomes (gray distribution). (b) Observed mean interchromosomal LD between 89 outlier regions exceeds interchromosomal LD between random SNPs estimated from 10,000 permutations.

old genetic variation within Haida Gwaii freshwater threespine stickleback.

DISPERSED GENOMIC ARCHITECTURE OF PARALLEL ADAPTATION

We scanned the genome for likely signatures of parallel selection in the adaptive radiation, using haplotype-based selection statistics *iHS* and *H12* computed across all 29 populations, and controlled for demography by reconstructing the demographic history of the 29 populations (Fig. S1) and obtaining neutral expectations for these statistics with 1000 whole genome simulations under the reconstructed demographic model (see *Methods*). We found 43,947 *iHS* outliers (0.92%) and 94,626 *H12* outliers (1.44%) exceeding neutral expectations at the 99.9% quantile (Fig. 2a), with 4894 SNPs being outliers for both *iHS* and *H12* (0.10%). Thus, we found five to 10 times more outliers for each statistic than expected under neutrality due to demographic stochasticity (Fig. S2, see *Methods*). Potential phasing errors due to statistical phasing did not increase the false positive rate to detect outliers (Fig. S2).

We found 89 outlier regions in the genome that are enriched for both *iHS* and *H12* outliers and are thus likely targets of parallel selection (Fig. 2a; Table S2). These outlier regions are dispersed across all but two chromosomes, with an accumulation on chromosomes XI and VII (Fig. S3). They span 8.7 Mb or 1.99% of the genome, with a median size of 75 kb and ranging from 50 to 250 kb (Table S2; Figs. S4–S8). One outlier region overlaps a known inversion on chromosome I (Fig. 2a,b; Table S2). The genomic architecture of parallel adaptation inferred from these outlier regions is thus rather dispersed with clustering on two chromosomes and in one inversion only.

LD between the 89 outlier regions across all 29 populations suggests that the number of independently evolving genomic regions in the Haida Gwaii stickleback radiation is smaller than 89 (Fig. 3). Although LD is expected for physically linked outlier regions, for example, on chromosome IV, we find elevated interchromosomal LD between outlier regions, exceeding expectations from population structure (Fig. 3). Hierarchical clustering revealed 47 independently evolving LD clusters of outlier regions (Fig. 3), with correlated evolution between some physically

unlinked outlier regions showing near perfect LD, for example, XI.c, XIII.a, and V.c, XIX.a and IV.g, or VIII.e, XII.b, and VII.d (Fig. 3).

ADAPTATION FROM OLD GENOMIC BLOCKS

We scanned the genome for old segregating variation by examining absolute divergence d_{XY} between pairs of individuals (Fig. 2a). Absolute divergence varies considerably across the genome with local mutation rate variation, but old segregating variation stands out by a large range in d_{XY} values observed between different pairs of individuals (Nelson and Cresko 2018; Marques et al. 2019), exemplified by the three inversions on chromosomes I, XI, and XXI (Fig. 2a,c,d). We found that outlier regions were enriched for such blocks of old segregating genetic variation: 24% of the outlier regions fall into the top 5% d_{XY} range windows (Q95%- d_{XY} range, $\chi^2_1 = 38.4$, $P = 5.7 \times 10^{-10}$) and 16% into the top 1% windows (Q99%- d_{XY} range, $\chi^2_1 = 43.4$, $P = 4.4 \times 10^{-11}$; Fig. 2d; Table S2). However, blocks of old variation were not enriched for gene number (permutation test, $P = 0.19$; Fig. S10).

We found that 45 of the 89 outlier regions (Table S2) overlap previously identified divergent regions between stickleback ecotypes (Hohenlohe et al. 2010; Deagle et al. 2012; Jones et al. 2012a; Jones et al. 2012b; Feulner et al. 2015; Roesti et al. 2015; Marques et al. 2016). Twenty-five outlier regions overlap marine versus freshwater ecotype differentiated regions, including seven of 13 regions identified by Hohenlohe et al. (2010) in Alaska and 21 of 102 regions by Jones et al. (2012a) and 62 of 322 regions by Jones et al. (2012b) across the Northern Hemisphere. Thirty-four outlier regions overlap lake versus stream ecotype differentiated regions, including three of 42 regions identified by Deagle et al. (2012) on Haida Gwaii, 48 of 839 regions by Feulner et al. (2015) in the East Pacific and East Atlantic as well as two of 47 regions by Roesti et al. (2015) and six of 37 regions by Marques et al. (2016) in Central Europe. Eight outlier regions overlap eight benthic versus limnetic ecotype divergent regions among the 46 regions identified by Jones et al. (2012a) in British Columbia. However, only four outlier regions overlap four of the 77 regions under selection previously identified in a selection experiment between extreme Haida Gwaii freshwater niches (Marques et al. 2018). Blocks of old standing genetic variation (Q95%- d_{XY} range windows) are enriched for marine-freshwater ($\chi^2_1 = 10.2$, $P = 1.4 \times 10^{-3}$) and limnetic-benthic outlier regions ($\chi^2_1 = 4.2$, $P = 4.1 \times 10^{-2}$) but not for other ecotype contrasts (Table S2). The signatures of parallel selection in the adaptive radiation of stickleback on Haida Gwaii are thus enriched for old genetic variation, in particular variation involved in repeated marine-freshwater and limnetic-benthic ecotype divergence.

NICHE DIVERGENCE PREDICTS GENOMIC ARCHITECTURE

We investigated how the genomic architecture of parallel adaptation changes across the multidimensional niche space of Haida Gwaii freshwater habitats (Figs. 4a,c, S11a,c). We tested what ecological or genetic predictors (see *Methods*) best explain one aspect of genomic architecture: the number of selection targets in the genome, with the number of sweep haplotypes in a population or shared between populations as proxies. Niche divergence from the ancestral marine habitat had the strongest effect on the number of sweep haplotypes in a population, followed by geography, ecosystem size, predation regime, light spectrum, and population size in order of decreasing effect size, whereas niche isolation had a negligible and sequencing depth and the proportion of missing data no detectable effects (Table 1; linear model, $F_{10,18} = 109.5$, $P = 5.1 \times 10^{-14}$, adjusted $r^2 = 0.97$; Fig. 4b). In a niche space defined by ecological principal components (Fig. S11), niche divergence from the ancestral habitat also had the strongest effect, followed by ecological PC1, geography, population size, and ecological PC2, whereas isolation in niche space, sequencing depth, and missing data had no detectable effect (Table 1; linear model, $F_{9,19} = 121$, $P = 7.3 \times 10^{-15}$, adjusted $r^2 = 0.97$; Fig. S11b). Similarly, the number of shared sweep haplotypes was best predicted by niche distance from the ancestral habitat, followed by genetic relatedness, population size, pairwise niche differentiation, and ecosystem size, whereas predation regime and light spectrum were excluded from the model and niche isolation, sequencing depth, and the proportion of missing data had no detectable effect (Table 1; linear model, $F_{7,398} = 151$, $P = 2.2 \times 10^{-16}$, adjusted $r^2 = 0.72$; Fig. 4d). In the principal component niche space, effect sizes were again similar, but with ecological PC2 having a detectable effect on the sharing of sweep haplotypes (Table 1; linear model, $F_{7,398} = 121$, $P = 2.2 \times 10^{-16}$, adjusted $r^2 = 0.72$; Fig. S11d). Taken together, our results show that populations further away from the ancestral niche and diverging along multiple ecological axes contain more total and more shared sweep haplotypes. Likewise, populations North-East in the Haida Gwaii archipelago and populations that are more closely related to each other contain more total and more shared haplotypes under parallel selection. Also, populations with smaller sizes and population pairs whose niches are more similar to each other show more and more shared sweep haplotypes (Figs. 4, S11).

CONNECTING TARGETS AND SOURCES OF PARALLEL SELECTION

The 89 outlier regions overlap 293 genes without gene ontology term enrichment but functions relevant to the three main ecological axes (Figs. S4-S9; Table S4). Functions related to predation regime include seven bone development genes potentially

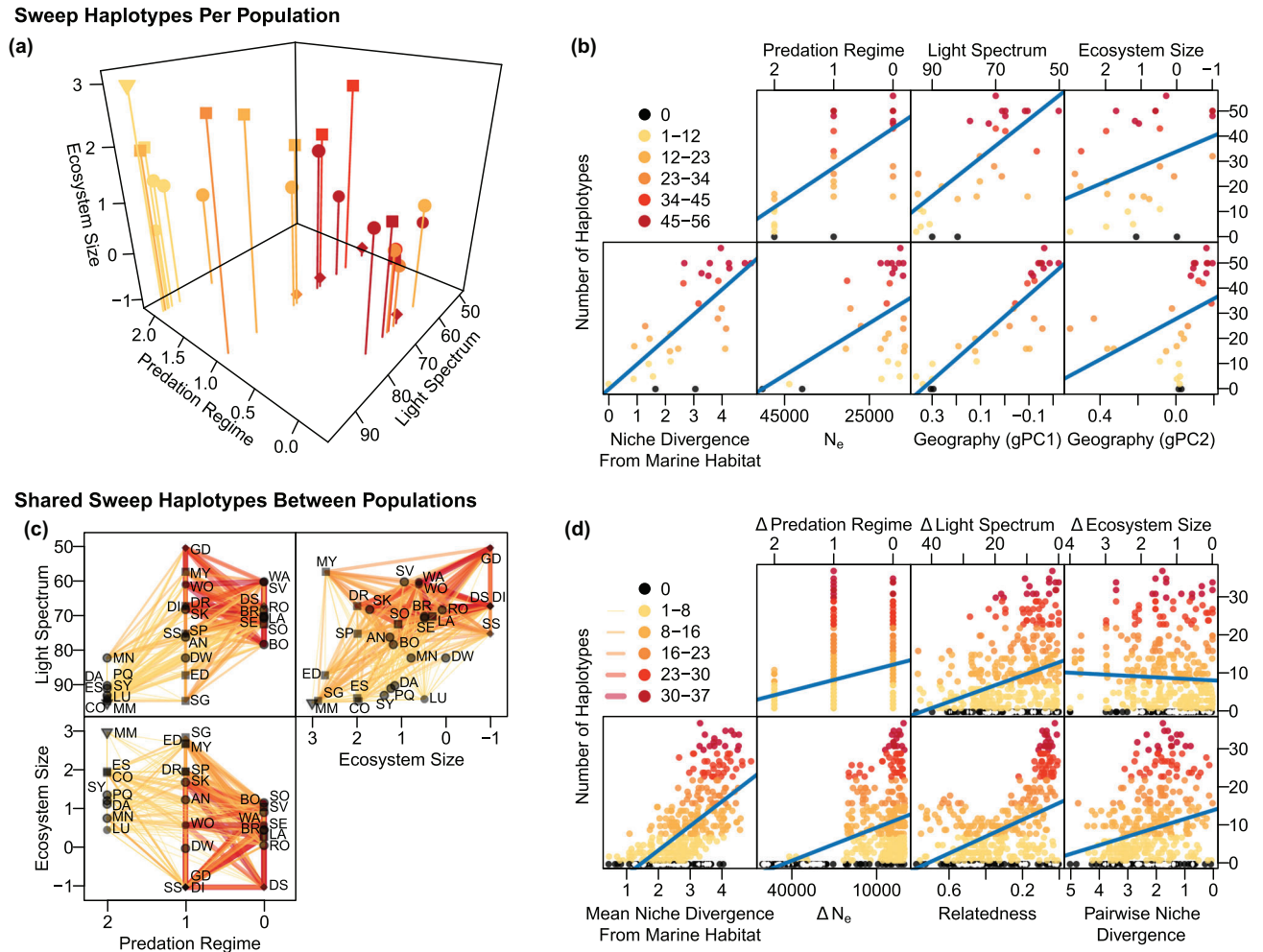


Figure 4. Niche divergence from the ancestral habitat and geography best predict the genomic architecture of parallel adaptation. (a) Number of sweep haplotypes in a population shown by color code in the three-dimensional niche space of Haida Gwaii archipelago. (b) Niche divergence from the ancestral, marine habitat has the strongest effect (Table 1) on the number of sweep haplotypes in a population, followed by geography (gPC1, gPC2), likely due to spatial autocorrelation of the niche space in the Haida Gwaii archipelago (Reimchen et al. 2013). Population size N_e is negatively associated with haplotype number, as the marine and large freshwater populations with marine-like phenotypes contain few sweep haplotypes, contrary to expectations from population size limiting the efficiency of selection. (c) Number of shared sweep haplotypes between populations shown by color code in the niche space of the Haida Gwaii archipelago. (d) Mean niche divergence from the marine habitat has also the strongest effect on the number of shared sweep haplotypes between population, followed by genetic relatedness of populations and differences in population size ΔN_e and pairwise niche divergence (Table 1).

related to bony predation defense traits (Reimchen et al. 2013). Furthermore, two muscle development genes and four genes associated with fear-related behavior in mice or autism in humans may potentially be relevant for adaptations in predator evasion behavior (Reimchen 1991). Two genes involved in growth regulation/body size and wound healing might potentially be relevant for escaping gape-limited predators or surviving unsuccessful predation attempts (Reimchen 1991). Gene functions related to light spectrum across the Haida Gwaii populations include 20 visual perception and five pigmentation genes. Visual perception genes include the *short-wave-sensitive opsin (SWS2)* color

vision gene demonstrated previously to be under selection (Marques et al. 2017a), *cnrgala* controlling the last step in the phototransduction pathway in rod cells (Shuart et al. 2011), *dnmt1* being part of rod cells (Tittle et al. 2011), *gucy2d* related to cone-rod dystrophy (Bradford et al. 2011), and several genes involved in corneal dystrophy, *gnb2* involved in retinal signal transduction (Dhingra et al. 2012), or *prom2* involved in pigmentation of the retina (Zerbino et al. 2018). Five genes involved in melanocyte development may be relevant to melanism associated with the light spectrum and sexual selection imposed by females across the radiation (Reimchen 1989; McDonald et al. 1995; Flamarique

Table 1. Associations between ecological and genetic predictors and the genomic architecture of parallel adaptation. Shown are the predictor effect sizes for four separate linear models using either predictor values computed in a niche space consisting of three main ecological axes (P, S, E) or of three ecological principal components (ePC) and either the number of sweep haplotypes per population or the number of shared haplotypes as response variable. β = parameter estimate/effect size, SE = standard error for the parameter, t = test statistic for single predictors, P -value = false discovery rate adjusted P -value.

Niche Space	Response Variable	Number of Sweep Haplotypes				Number of Shared Sweep Haplotypes			
	Predictor Variable	β	SE	t	P -value	B	SE	t	P -value
x: P y: S z: E	Predation Regime (P)	-8.3	2.0	-4.0	0.0049	-	-	-	-
	Light Spectrum (S)	-7.9	2.6	-3.1	0.0254	-	-	-	-
	Ecosystem Size (E)	-10.4	2.6	-3.9	0.0052	1.0	0.3	3.1	0.0119
	Ancestral Niche Divergence (ANC)	-17.2	5.2	-3.3	0.0172	5.2	0.2	21.0	<0.0001
	Pairwise Niche Divergence (DIV)	-	-	-	-	-1.2	0.3	-3.4	0.0052
	Niche Isolation (ISO)	1.6	0.7	2.1	0.1616	-0.4	0.3	-1.5	0.5198
	Population Size (POP)	-3.8	0.7	-5.4	0.0003	-2.4	0.2	-9.6	<0.0001
	Geography (gPC1)	-11.7	1.4	-8.5	<0.0001	-	-	-	-
	Geography (gPC2)	-8.4	0.6	-14.4	<0.0001	-	-	-	-
	Relatedness (gPC)	-	-	-	-	-3.9	0.3	-14.4	<0.0001
	Sequencing Depth (DEP)	-2.2	1.9	-1.2	0.7362	-	-	-	-
	Missing Data (MIS)	-3.0	1.8	-1.7	0.3602	0.4	0.2	1.6	0.4826
	x: ePC1 y: ePC2 z: ePC3	Ecological PC1 (ePC1)	15.6	3.8	4.1	0.0042	-0.7	0.3	-2.6
Ecological PC2 (ePC2)		3.2	1.1	2.9	0.0459	-0.9	0.2	-3.6	0.0027
Ecological PC3 (ePC3)		-	-	-	-	0.6	0.2	2.2	0.1199
Ancestral Niche Divergence (ANC)		-17.8	5.1	-3.5	0.0141	5.3	0.3	20.6	<0.0001
Pairwise Niche Divergence (DIV)		-	-	-	-	-	-	-	-
Niche Isolation (ISO)		1.4	0.7	2.0	0.2144	-	-	-	-
Population Size (POP)		-3.9	0.7	-5.5	0.0002	-2.3	0.2	-9.5	<0.0001
Geography (gPC1)		-11.5	1.4	-8.4	<0.0001	-	-	-	-
Geography (gPC2)		-8.3	0.6	-14.4	<0.0001	-	-	-	-
Relatedness (gPC)		-	-	-	-	-4.1	0.3	-15.5	<0.0001
Sequencing Depth (DEP)		-2.9	1.7	-1.7	0.3670	-	-	-	-
Missing Data (MIS)		-3.9	1.6	-2.4	0.1155	-	-	-	-

et al. 2013). Gene functions related to ecosystem size include 58 genes with metabolic functions ($n = 23$) or affecting facial or tooth morphology ($n = 3$) putatively related to zooplankton or benthic invertebrate diets (Moodie and Reimchen

1976; Reimchen et al. 1985; Spoljaric and Reimchen 2007), 17 immunity genes potentially related to divergent parasite faunas (Buckland-Nicks et al. 1990; Reimchen and Buckland-Nicks 1990; Buckland-Nicks and Reimchen 1995; Buckland-Nicks

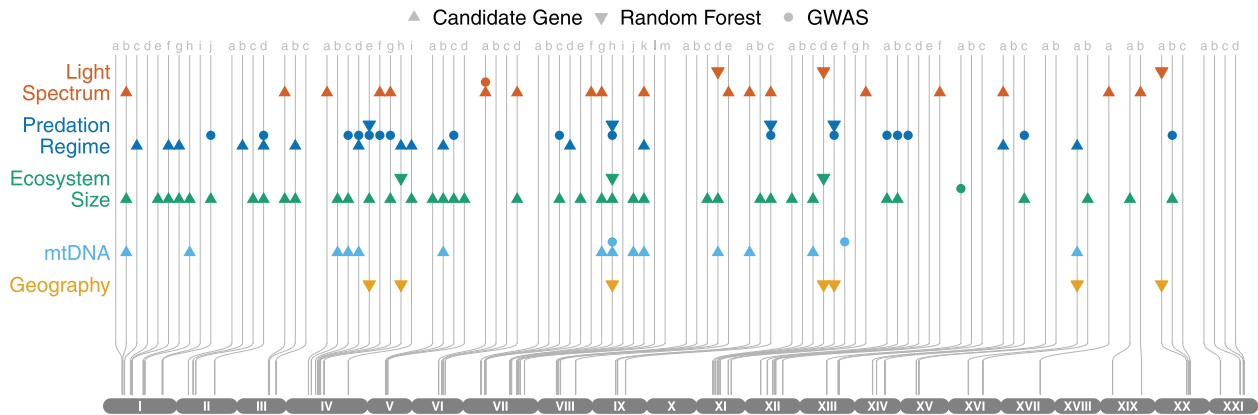


Figure 5. Ecology-associated outlier regions and their distribution across the genome. Associations between outlier regions (vertical gray bars) and the three major ecological axes are widely distributed across the genome. Shown are putative candidate gene targets of selection (upward pointing triangle), GWAS associations of peak H12-SNPs with 50-kb sliding window averaged P -values <0.01 (circle), and random forest associations for SNPs with variable importance mean decrease in accuracy of >20 (downward pointing triangle). See *Methods* for grouping of ecological properties and phenotypic traits into associations with light spectrum, predation regime, and ecosystem size.

et al. 1997), and genes with roles in reproduction ($n = 12$), oxidative stress ($n = 2$), or fin development ($n = 1$) potentially related to other ecosystem adaptations, for example, to the abiotic environment or environment-dependent sexual selection (Reimchen 1989; McDonald et al. 1995; Bergstrom 2002; Flamarique et al. 2013). Twenty-two nuclear genes interacting with the mitochondrion may be associated with mitonuclear incompatibilities as additional source of selection.

We tested whether outlier region genotypes are associated with 16 ecological and phenotypic variables, mtDNA haplotypes, and geographic structure (Table S3), using GWAS, random forest, and SAGUARO (Zamani et al. 2013) (Figs. 4, S12–S14; Table S1). We found several associations between outlier regions and ecological and phenotypic metrics reflecting the three major ecological axes, for example, lateral plate morph (predation), light spectrum, and gill raker length (ecosystem size; Figs. 5, S12–S14). For the six variables plate morph, lateral plate number, light spectrum, gill raker length, mtDNA haplotype, and geography, random forests showed appreciable predictive power, whereas for other variables predictive power was low (Fig. S12). Predictive power was greatest for geography, followed by plate morph and light spectrum, with multiple outlier regions across several chromosomes showing associations with these variables (Figs. 5, S12). Of 21 local phylogenetic trees assembled by SAGUARO (see *Methods*), 17 showed associations with environmental or phenotypic variables, of which 11 also showed simultaneous association with geography (Fig. S13). Covariance of genotype-ecology associations with geography was common. GWAS results, taking population structure and thus geography into account, suggest some associations between outlier regions and predominantly predation-related traits such as plate morph

(Fig. 5, S14), but are limited in power with only 29 populations in our dataset. Taken together, candidate gene functions suggest many putative links between sources and targets of parallel selection.

Discussion

The adaptive radiation of threespine stickleback in diverse freshwater habitats on the Haida Gwaii archipelago is characterized by a genomic architecture of 89 loci putatively under parallel selection dispersed throughout the genome with little clustering, enriched for old genetic variation, and containing candidate genes and genotype-environment correlations related to predation regime, ecosystem size, and light environment. Population sizes appear to have been sufficiently large for parallel adaptation to overcome drift (Ohta 1992), as we found more signatures of parallel adaptation in smaller populations (Figs. 4b, S11b) and few signatures in the largest marine and freshwater populations with marine-like phenotype (e.g., full lateral plate cover). Importantly, we found that the number and sharing of parallel selection signatures depend most strongly on the degree of niche differentiation away from the ancestral niche and the dimensionality of the niche space, followed only by geography, genetic relatedness, and population size. Thus, our results suggest that the genomic architecture of parallel adaptation depends on the length and trajectory of an adaptive walk across a niche space.

Niches show spatial autocorrelation on Haida Gwaii, as the most divergent freshwater habitats tend to cluster geographically in the North-East of the archipelago (Reimchen et al. 2013), where we found the most complex and shared genomic architectures of parallel adaptation. Although this finding fits

expectations from niche differentiation and niche space dimensionality being important, two alternative hypotheses need to be considered. First, closely related populations share more standing genetic variation, making parallel adaptation more likely (Conte et al. 2012), which is in line with the positive correlation we observe between genetic relatedness and the number of shared sweep signatures identified (Figs. 4d, S11d). Second, due to isolation by distance, haplotypes between geographically distant populations are expected to be more dissimilar. Denser population sampling in the North-East of the archipelago (Fig. 1a) could thus have led to a detection bias in haplotype-based signatures of selection, as suggested by most signatures of parallel adaptation being found in the North-East (Fig. 4b). Although both alternative hypotheses may have contributed to our findings, they cannot explain the larger effect size of niche divergence compared to genetic relatedness and geography in explaining the number of sweep signatures in a population or shared between populations (Table 1). We thus conclude that niche divergence and niche space dimensionality—the length and complexity of an adaptive walk—must have played an important role in shaping the genomic architecture of parallel adaptation in the Haida Gwaii threespine stickleback adaptive radiation.

The genomic architecture of parallel adaptation we find contrasts with theoretical predictions of a simplified or clustered genetic architecture of adaptation associated with adaptive radiations into multiple niches (Gavrilets 2004; Gavrilets and Vose 2005, 2009; Gavrilets and Losos 2009), but lends empirical support to an important role of old genetic variation in adaptive radiations (Feder et al. 2003; Rieseberg et al. 2003; Lamichhaney et al. 2015; Meier et al. 2017; Richards and Martin 2017; Nelson and Cresko 2018; Marques et al. 2019). Theory and empirical studies of the genomic architecture of adaptation for both two-habitat problems (Kirkpatrick and Barton 2006; Yeaman and Whitlock 2011; Jones et al. 2012b; Renaud et al. 2012; Yeaman 2013; Malinsky et al. 2018; Meier et al. 2018; Van Belleghem et al. 2018; van Rijssel et al. 2018; Roberts Kingman et al. 2021) and multidimensional niche spaces (Gavrilets and Vose 2005; Gavrilets and Losos 2009; Martin et al. 2013; Arnegard et al. 2014; Richards and Martin 2017; Kautt et al. 2020; Magalhaes et al. 2021) predict simple genetic architectures with few and linked loci underlying adaptation to new niches. Although some empirical studies have found such simplified genomic architectures (Feder et al. 2003; Lowry and Willis 2010; Renaud et al. 2012; Nosil et al. 2018; Van Belleghem et al. 2018), many others uncovered more complexity with many regions under selection (Jones et al. 2012b; Martin et al. 2013; Arnegard et al. 2014; Malinsky et al. 2015; Marques et al. 2016, 2018; Richards and Martin 2017; Meier et al. 2018; Miller et al. 2019; Therkildsen et al. 2019), and an association between complex genomic architecture and the likelihood of sympatric speciation (Kautt et al. 2020).

The discrepancy between theory and empirical findings may be due to most traits having a more complex genetic architecture (Orr 1998; Flint and Mackay 2009; Peichel and Marques 2017; Sella and Barton 2019), due to correlational selection on multiple traits affecting a large array of underlying loci (Svensson et al. 2021), or, as our study suggests, due to variation in the steepness of ecological gradients and a higher niche space dimensionality in those empirical studies than in theoretical models. In the few theoretical adaptive radiation studies considering multidimensional niche spaces (Gavrilets and Vose 2005), each niche dimension affects one trait only, whereas in reality multiple niche axes may affect the same and more than one trait, which could explain more complex empirical genomic architectures of adaptation. In the Haida Gwaii stickleback adaptive radiation, convergent phenotypic divergence in multiple traits along the three niche axes predation regime, light spectra, and ecosystem size suggests such complexity in phenotypic adaptation (Reimchen et al. 2013). The resulting complexity in the genomic architecture might thus be a product of many traits being involved, several of them being complex traits, correlational selection, variation in ecological gradient steepness, and niche space dimensionality.

The dispersed genomic architecture of parallel adaptation we find in the Haida Gwaii stickleback radiation is comparable to other stickleback ecotypes, such as marine and freshwater (Hohenlohe et al. 2010; Jones et al. 2012b; Bassham et al. 2018; Fang et al. 2020; Roberts Kingman et al. 2021) or lake and stream ecotypes (Deagle et al. 2012; Roesti et al. 2012; Marques et al. 2017b; Rennison et al. 2019a). Despite being single habitat contrasts, multiple environmental axes are involved in each, for example, abiotic conditions (Kusakabe et al. 2017; Stuart et al. 2017), food resources (Moser et al. 2016; Stuart et al. 2017; Ishikawa et al. 2019), predation regimes (Colosimo et al. 2005), parasite faunas (Karvonen et al. 2015; Weber et al. 2016), or environment-based sexual selection (Jones et al. 2008; Seear et al. 2015). Reduced recombination and inversions effectively reduced the high number of loci to fewer independently segregating blocks between marine and freshwater ecotypes (Jones et al. 2012b; Samuk et al. 2017; Nelson et al. 2019). In contrast to marine-freshwater and lake-stream contrasts, however, environmental axes in Haida Gwaii freshwater niche space are not all parallel and allowed us to explore the effects of niche space dimensionality and niche differentiation on parallel genomic adaptation. It makes intuitive sense that we found niche differentiation, the steepness of an ecological gradient, to be the best predictor of the number of loci putatively under parallel selection (Fig. 4)—the longer and more complex an adaptive walk to a new optimum, the more steps or traits may be involved. Fisher's geometric model (Fisher 1930) and other empirical results (Rogers et al. 2012) would predict that more large effect

size variants should be involved with longer adaptive walks, a prediction that should be tested in future work quantifying the effect sizes of loci putatively under parallel selection.

Our finding of the involvement of old genetic variation in the adaptive radiation of Haida Gwaii stickleback is in line with recent findings across multiple adaptive radiations (Feder et al. 2003; Rieseberg et al. 2003; Lamichhaney et al. 2015; Meier et al. 2017; Richards and Martin 2017; Nelson and Cresko 2018; Marques et al. 2019). Old genetic variation involved in stickleback adaptation is traditionally thought of as standing genetic variation maintained in a large marine gene pool (Schluter and Conte 2009) shared within an ocean and to a smaller degree between oceans (Rennison et al. 2019b; Fang et al. 2020; Magalhaes et al. 2021). The mitonuclear discordance we find between Japan Sea and threespine stickleback suggests that some of this standing genetic variation may have been derived from admixture between ~ 1 million years divergent lineages (Ravinet et al. 2018) in secondary contact in the Pacific between Japan Sea and threespine stickleback. This would be in line with a recent demonstration that several haplotypes instrumental in repeated marine and freshwater divergence (Colosimo et al. 2004, 2005; Jones et al. 2012b; Terekhanova et al. 2014; Bassham et al. 2018) are several million years divergent (Nelson and Cresko 2018; Roberts Kingman et al. 2021). Our results suggest that the Haida Gwaii radiation of threespine stickleback descends from a hybrid lineage containing Japan Sea stickleback alleles and thus would be consistent with the hypothesis of hybridization being a catalyst of adaptive radiation (Seehausen 2004, 2013). This also demonstrates that marine-freshwater divergent haplotypes, despite being enriched among old variants likely under parallel selection in the Haida Gwaii radiation, are not the sole source of ancient variation, nor are divergent freshwater populations a mere reassembly of old marine-freshwater divergent haplotypes. This calls for reassessing the sources of selection that originally gave rise to old adaptive variants, for example, for inversions previously associated with salinity differences between sea and fresh water (Jones et al. 2012b; Kusakabe et al. 2017) that appear to be associated with a largely marine-like phenotype in freshwater populations on Haida Gwaii instead (Fig. 2b). Further study is needed to establish whether this association in the Haida Gwaii radiation might be a by-product of correlated evolution of outlier regions, for example, correlation of marine alleles for inversions with the marine allele of the *Eda* gene controlling plate morph (Peichel et al. 2001; Colosimo et al. 2004, 2005), if other sources of selection than salinity are acting on these inversions, or whether associations are a consequence of recent colonization and incomplete adaptation (Bassham et al. 2018).

Correlated evolution among physically unlinked outlier regions in our data of 28 Haida Gwaii freshwater populations is

an interesting feature that could be explained in several ways. Correlational selection, selection on trait combinations, may explain the correlated response of many genomic targets observed in our study (Lande and Arnold 1983; Sinervo and Svensson 2002; Svensson et al. 2021). The independent evolution of convergent phenotypes and trait combinations (Reimchen et al. 2013) in response to similar niches on Haida Gwaii supports such an explanation. Alternatively, epistatic interactions between incompatibilities, for example, between old, divergent haplotypes, could constrain the number of possible allelic combinations. On the other hand, correlated evolution might also be a joint product of population structure and spatial autocorrelation of the niche space on Haida Gwaii, or historical selection and population structure prior to the colonization of postglacial freshwater habitats on Haida Gwaii (Bierne et al. 2013). Future analysis of selection and LD on a population level for key Haida Gwaii populations may help disentangle these potential drivers of correlated genomic evolution.

Future population-level analyses may further help quantify the relative importance of old variants and newly arising mutations. Although de novo variants have been shown to play a major role in recurrent evolution of some stickleback phenotypes (Chan et al. 2010; Xie et al. 2019), selection on them would not lead to parallel selection signatures as different mutations and associated haplotypes would be favored in different populations. Although we would have missed such regions in our current analysis of single individuals surveyed from many different populations, dense genotype information from multiple individuals within populations will allow for their quantification (Chan et al. 2010; Hohenlohe et al. 2010). Our focus on signatures of parallel selection between populations in our current study may also explain why we found only a minor overlap of outlier regions with the 77 regions under selection identified in a selection experiment recreating a major niche shift between freshwater habitats on Haida Gwaii (Marques et al. 2018). In the latter analysis, 36 of the regions under selection showed genotype versus environment or phenotype associations across the adaptive radiation (Marques et al. 2018), but only four of those regions showed parallel selective sweep signatures across the adaptive radiation here. A discrepancy between signatures of parallel evolution and population-specific responses to selection is to be expected, as population-specific responses might be contingent on the availability of alleles in a founding population (Conte et al. 2012), on the spatial and temporal dynamics selection (Reimchen 1995; Grant and Grant 2002; Reimchen and Nosil 2002, 2004; Nosil et al. 2018), and in addition the degree of contingency might depend on the genetic architecture of traits, with highly polygenic traits less likely leading to parallel genomic outcomes (Peichel and Marques 2017).

Conclusions

In conclusion, the radiation of threespine stickleback into a multi-dimensional niche space on Haida Gwaii demonstrates that niche differentiation and niche space dimensionality may be important predictors of the genomic architecture of parallel adaptation. Furthermore, enrichment of old genetic variation derived from other ecotypes and admixture between divergent lineages underlines its important role in facilitating adaptive radiation. A role of niche differentiation and niche space dimensionality in the genomic architecture of adaptation may have consequences for the evolution of reproductive isolation in adaptive radiations and thereby for the likelihood of sympatric speciation (Bolnick 2011; Kautt et al. 2020), range expansion, return into, and persistence in sympatry (Price 2008), and the build-up of sympatric species diversity as seen in the most iconic adaptive radiations (Gillespie et al. 2020). With similar information on the sources of genetic variation and genomic architecture of adaptation emerging from other adaptive radiations, the Haida Gwaii stickleback radiation adds a comparative puzzle piece to how genomic mechanisms facilitate or constrain adaptive radiation once a lineage gets the opportunity of filling a diversity of empty niches.

AUTHOR CONTRIBUTIONS

TER conceived the sampling design and conducted the collection of fish, ecological, and morphological data. Sequencing and genotype calling was performed by DMK, FCJ, and FDP. DAM conducted the data analysis and wrote this article, with contributions of all co-authors.

ACKNOWLEDGMENTS

We would like to thank C. B. Lowe, S. D. Brady, B. Deagle, J. Turner, K. Lindblad-Toh, and the Broad Genomics Platforms for assistance with samples and sequencing; B. Moa for bioinformatics support; and M. Muschick for discussion. This work was funded by the National Research Council Canada grant NRC2354 to TER and National Institute of Health grants 3P50HG002568-09S1 ARRA and 3P50HG002568 to DMK; DAM was also supported through the Swiss National Science Foundation grant 31003A_163338 to O. Seehausen and L. Excoffier.

Open Access Funding provided by Universitat Bern.

DATA ARCHIVING

All genomic data of threespine stickleback collected on Haida Gwaii have been deposited on the NCBI Sequence Read Archive under accession SRP100209. Mitochondrial sequences of blackspotted stickleback (*Gasterosteus wheatlandi*) and Japan Sea stickleback (*Gasterosteus nipponicus*) used in this study were obtained from GenBank accessions AB445130.1 and AB445129.1 (Kawahara et al. 2009). Ecological and phenotypic data used in this study are published in Table S1. All custom scripts used in this study (“arp2vcf.py,” “dxy_wsfs.py”) are accessible on <https://github.com/marquedada> and have been deposited on Dryad <https://doi.org/10.5061/dryad.1ns1rn8wk> alongside code and parameter files to replicate the simulated whole genome data. Specimens and associated materials are stored at the University of Victoria, Canada, in the collection of TER, and are accessible for examination on site upon request.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

LITERATURE CITED

- Arendt, J. & Reznick, D. (2008) Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.*, 23, 26–32.
- Arnegard, M.E., McGee, M.D., Matthews, B., Marchinko, K.B., Conte, G.L., Kabir, S., Bedford, N., Berggek, S., Chan, Y.F., Jones, F.C., et al. (2014) Genetics of ecological divergence during speciation. *Nature*, 511, 307–311.
- Bassham, S., Catchen, J., Lescak, E., vonHippel, F.A. & Cresko, W.A. (2018) Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. *Genetics*, 209, 921–939.
- Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.*, 57, 289–300.
- Bergstrom, C.A. (2002) Fast-start swimming performance and reduction in lateral plate number in threespine stickleback. *Can. J. Zool.*, 80, 207–213.
- Bierne, N., Gagnaire, P.A. & David, P. (2013) The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Curr. Zool.*, 59, 72–86.
- Blake, J.A., Eppig, J.T., Kadin, J.A., Richardson, J.E., Smith, C.L., Bult, C.J., & the Mouse Genome Database Group (2017) Mouse Genome Database (MGD): community knowledge resource for the laboratory mouse. *Nucleic Acids Res.*, 45, D723–D729.
- Bolnick, D.I. (2011) Sympatric speciation in threespine stickleback: why not? *Int. J. Ecol.*, 2011, 1–15.
- Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Howe, D.G., Knight, J., Mani, P., Martin, R., Moxon, S.A., et al. (2011) ZFIN: enhancements and updates to the zebrafish model organism database. *Nucleic Acids Res.*, 39, D822–D829.
- Breiman, L. (2001) Random forests. *Mach. Learn.*, 45, 5–32.
- Broad Institute (2018) Picard tools. Broad Institute, Cambridge, MA.
- Buckland-Nicks, J. & Reimchen, T. (1995) A novel association between an endemic stickleback and a parasitic Dinoflagellate. 3. Details of the life-cycle. *Arch. Protistenkd.*, 145, 165–175.
- Buckland-Nicks, J., Reimchen, T.E. & Garbary, D.J. (1997) *Haidadinium ichthyophilum* gen.nov. et sp.nov. (Phytophycinales, Dinophyceae), a freshwater ectoparasite on stickleback (*Gasterosteus aculeatus*) from the Queen Charlotte Islands, Canada. *Can. J. Bot.*, 75, 1936–1940.
- Buckland-Nicks, J.A., Reimchen, T.E. & Taylor, M.F.J.R. (1990) A novel association between an endemic stickleback and a parasitic Dinoflagellate. 2. Morphology and life-cycle. *J. Phycol.*, 26, 539–548.
- Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G.Jr., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science*, 327, 302–305.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4, 7.
- Colosimo, P.F., Peichel, C.L., Nereng, K., Blackman, B.K., Shapiro, M.D., Schluter, D. & Kingsley, D.M. (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *Plos Biol.*, 2, E109.
- Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G.Jr., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D. & Kingsley, D.M. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science*, 307, 1928–1933.

- Conte, G.L., Arnegard, M.E., Peichel, C.L. & Schluter, D. (2012) The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. B Biol. Sci.*, 279, 5039–5047.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.
- Deagle, B.E., Reimchen, T.E. & Levin, D.B. (1996) Origins of endemic stickleback from the Queen Charlotte Islands: mitochondrial and morphological evidence. *Can. J. Zool.*, 74, 1045–1056.
- Deagle, B.E., Jones, F.C., Chan, Y.F., Absher, D.M., Kingsley, D.M. & Reimchen, T.E. (2012) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proc. R. Soc. B Biol. Sci.*, 279, 1277–1286.
- Deagle, B.E., Jones, F.C., Absher, D.M., Kingsley, D.M. & Reimchen, T.E. (2013) Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Mol. Ecol.*, 22, 1917–1932.
- Delaneau, O., Howie, B., Cox, A.J., Zagury, J.F. & Marchini, J. (2013) Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.*, 93, 687–696.
- Dhingra, A., Ramakrishnan, H., Neinstein, A., Fina, M.E., Xu, Y., Li, J., Chung, D.C., Lyubarsky, A. & Vardi, N. (2012) *Gbeta3* is required for normal light ON responses and synaptic maintenance. *J. Neurosci.*, 32, 11343–11355.
- Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V.C. & Foll, M. (2013) Robust demographic inference from genomic and SNP data. *Plos Genet.*, 9:e1003905.
- Fang, B., Kempainen, P., Momigliano, P., Feng, X. & Merilä, J. (2020) On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat. Ecol. Evol.*, 4, 1105–1115.
- Feder, J.L., Berlocher, S.H., Roethele, J.B., Dambroski, H., Smith, J.J., Perry, W.L., Gavrilovic, V., Filchak, K.E., Rull, J. & Aluja, M. (2003) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci. USA*, 100, 10314–10319.
- Feulner, P.G., Chain, F.J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., Lenz, T.L., Samonte, I.E., Stoll, M., Bornberg-Bauer, E., et al. (2015) Genomics of divergence along a continuum of parapatric population differentiation. *Plos Genet.*, 11:e1004966.
- Fisher, R.A. (1930) The genetical theory of natural selection. The Clarendon Press, Oxford, U.K.
- Flamarique, I.N., Bergstrom, C., Cheng, C.L. & Reimchen, T.E. (2013) Role of the iridescent eye in stickleback female mate choice. *J. Exp. Biol.*, 216, 2806–2812.
- Flint, J. & Mackay, T.F. (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.*, 19, 723–733.
- Fumagalli, M., Vieira, F.G., Korneliussen, T.S., Linderoth, T., Huerta-Sanchez, E., Albrechtsen, A. & Nielsen, R. (2013) Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*, 195, 979–992.
- Garud, N.R., Messer, P.W., Buzbas, E.O. & Petrov, D.A. (2015) Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *Plos Genet.*, 11:e1005004.
- Gavrilets, S. (2004) Fitness landscapes and the origin of species. Princeton Univ. Press, Princeton, NJ.
- Gavrilets, S. & Vose, A. (2005) Dynamic patterns of adaptive radiation. *Proc. Natl. Acad. Sci. USA*, 102, 18040–18045.
- . (2009) Dynamic patterns of adaptive radiation: evolution of mating preferences. Pp. 102–126 in D. Schluter, J. Bridle, & R. Butlin, eds. Speciation and patterns of diversity. Cambridge Univ. Press, Cambridge, U.K.
- Gavrilets, S. & Losos, J.B. (2009) Adaptive radiation: contrasting theory with data. *Science*, 323, 732–737.
- Gillespie, R.G., Bennett, G.M., De Meester, L., Feder, J.L., Fleischer, R.C., Harmon, L.J., Hendry, A.P., Knope, M.L., Mallet, J., Martin, C., et al. (2020) Comparing adaptive radiations across space, time, and taxa. *J. Hered.*, 111, 1–20.
- Glazer, A.M., Killingbeck, E.E., Mitros, T., Rokhsar, D.S. & Miller, C.T. (2015) Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3*, 5, 1463–1472.
- Grant, P.R. & Grant, B.R. (2002) Unpredictable evolution in a 30-year study of Darwin's finches. *Science*, 296, 707–711.
- Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids S.*, 41, 95–98.
- Hoffmann, A.A., Sgro, C.M. & Weeks, A.R. (2004) Chromosomal inversion polymorphisms and adaptation. *Trends Ecol. Evolution*, 19, 482–488.
- Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genet.*, 6:e1000862.
- Howe, D.G., Bradford, Y.M., Conlin, T., Eagle, A.E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S.A., et al. (2013) ZFIN, the zebrafish model organism database: increased support for mutants and transgenics. *Nucleic Acids Res. earch*, 41, D854–D860.
- Ishikawa, A., Kabeya, N., Ikeya, K., Kakioka, R., Cech, J.N., Osada, N., Leal, M.C., Inoue, J., Kume, M., Toyoda, A., et al. (2019) A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science*, 364, 886–889.
- Jones, F.C., Brown, C. & Braithwaite, V.A. (2008) Lack of assortative mating between incipient species of stickleback from a hybrid zone. *Behaviour*, 145, 463–484.
- Jones, F.C., Chan, Y.F., Schmutz, J., Grimwood, J., Brady, S.D., Southwick, A.M., Absher, D.M., Myers, R.M., Reimchen, T.E., Deagle, B.E., et al. (2012a) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.*, 22, 83–90.
- Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., et al. (2012b) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484, 55–61.
- Joron, M., Papa, R., Beltran, M., Chamberlain, N., Mavarez, J., Baxter, S., Abanto, M., Bermingham, E., Humphray, S.J., Rogers, J., et al. (2006) A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *Plos Biol.*, 4, 1831–1840.
- Karvonen, A., Lucek, K., Marques, D.A. & Seehausen, O. (2015) Divergent macroparasite infections in parapatric Swiss lake-stream pairs of threespine stickleback (*Gasterosteus aculeatus*). *Plos One*, 10:e0130579.
- Kautt, A.F., Kratochwil, C.F., Nater, A., Machado-Schiaffino, G., Olave, M., Henning, F., Torres-Dowdall, J., Harer, A., Hulsey, C.D., Franchini, P., et al. (2020) Contrasting signatures of genomic divergence during sympatric speciation. *Nature*, 588, 106–111.
- Kawahara, R., Miya, M., Mabuchi, K., Near, T.J. & Nishida, M. (2009) Stickleback phylogenies resolved: evidence from mitochondrial genomes and 11 nuclear genes. *Mol. Phyl. Evol.*, 50, 401–404.
- Kirkpatrick, M. & Barton, N. (2006) Chromosome inversions, local adaptation and speciation. *Genetics*, 173, 419–434.
- Kocher, T.D. (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.*, 5, 288–298.
- Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. (2014) ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, 15, 356.

- Kratochwil, C.F., Liang, Y., Gerwin, J., Woltering, J.M., Urban, S., Henning, F., Machado-Schiaffino, G., Hulsey, C.D. & Meyer, A. (2018) *Agouti-related peptide 2* facilitates convergent evolution of stripe patterns across cichlid fish radiations. *Science*, 362, 457–460.
- Kusakabe, M., Ishikawa, A., Ravinet, M., Yoshida, K., Makino, T., Toyoda, A., Fujiyama, A. & Kitano, J. (2017) Genetic basis for variation in salinity tolerance between stickleback ecotypes. *Mol. Ecol.*, 26, 304–319.
- Lamichhaney, S., Berglund, J., Almen, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerova, M., Rubin, C.J., Wang, C., Zamani, N., et al. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518, 371–375.
- Lande, R. & Arnold, S.J. (1983) The measurement of selection on correlated characters. *Evolution*, 37, 1210–1226.
- Lanfear, R., Kokko, H. & Eyre-Walker, A. (2014) Population size and the rate of evolution. *Trends Ecol. Evol.*, 29, 33–41.
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evolution*, 34, 772–773.
- Law, T. & Blake, R. (1996) Comparison of the fast-start performances of closely related, morphologically distinct threespine sticklebacks (*Gasterosteus* spp.). *J. of Exp. Biol.*, 199, 2595–2604.
- Leaver, S.D. & Reimchen, T.E. (2012) Abrupt changes in defence and trophic morphology of the giant threespine stickleback (*Gasterosteus* sp.) following colonization of a vacant habitat. *Biol. J. Linn. Soc. iety*, 107, 494–509.
- Leinonen, T., Herczeg, G., Cano, J.M. & Merilä, J. (2011) Predation-imposed selection on threespine stickleback (*Gasterosteus aculeatus*) morphology: a test of the refuge use hypothesis. *Evolution*, 65, 2916–2926.
- Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.
- . (2011) Inference of human population history from individual whole-genome sequences. *Nature*, 475, 493–484.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & S. Genome Project Data Processing (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
- Lowry, D.B. & Willis, J.H. (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *Plos Biol.*, 8:e1000500.
- Maclean, C.A., Chue Hong, N.P. & Prendergast, J.G. (2015) hapbin: an efficient program for performing haplotype-based scans for positive selection in large genomic datasets. *Mol. Biol. Evol.*, 32, 3027–3029.
- Magalhaes, I.S., Whiting, J.R., D'Agostino, D., Hohenlohe, P.A., Mahmud, M., Bell, M.A., Skúlason, S. & MacColl, A.D.C. (2021) Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nat. Ecol. Evol.*, 5, 251–261.
- Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terai, Y., Ngatunga, B.P., Miska, E.A., Durbin, R., Genner, M.J. & Turner, G.F. (2015) Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, 350, 1493–1498.
- Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F. & Durbin, R. (2018) Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.*, 2, 1940–1955.
- Marques, D.A., Lucek, K., Meier, J.I., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. (2016) Genomics of rapid incipient speciation in sympatric threespine stickleback. *Plos Genet.*, 12:e1005887.
- Marques, D.A., Taylor, J.S., Jones, F.C., Di Palma, F., Kingsley, D.M. & Reimchen, T.E. (2017a) Convergent evolution of *SWS2* opsin facilitates adaptive radiation of threespine stickleback into different light environments. *Plos Biol.*, 15:e2001627.
- Marques, D.A., Lucek, K., Haesler, M.P., Feller, A.F., Meier, J.I., Wagner, C.E., Excoffier, L. & Seehausen, O. (2017b) Genomic landscape of early ecological speciation initiated by selection on nuptial colour. *Mol. Ecology*, 26, 7–24.
- Marques, D.A., Jones, F.C., Di Palma, F., Kingsley, D.M. & Reimchen, T.E. (2018) Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nat. Ecol. Evol.*, 2, 1130–1140.
- Marques, D.A., Meier, J.I. & Seehausen, O. (2019) A combinatorial view on speciation and adaptive radiation. *Trends in Ecol. Evol.*, 34, 531–544.
- Martin, C.H. & Richards, E.J. (2019) The paradox behind the pattern of rapid adaptive radiation: how can the speciation process sustain itself through an early burst? *Ann. Rev. Ecol., Evol. S.*, 50, 569–593.
- Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J. & Jiggins, C.D. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.*, 23, 1817–1828.
- McDonald, C.G., Reimchen, T.E. & Hawryshyn, C.W. (1995) Nuptial colour loss and signal masking in *Gasterosteus*: an analysis using video imaging. *Behaviour*, 132, 963–977.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, 20, 1297–1303.
- Meier, J.I., Marques, D.A., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. (2017) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Comm.unication*, 8:14363.
- Meier, J.I., Marques, D.A., Wagner, C.E., Excoffier, L. & Seehausen, O. (2018) Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evolution*, 35, 1489–1506.
- Miller, S.E., Roesti, M. & Schluter, D. (2019) A single interacting species leads to widespread parallel evolution of the stickleback genome. *Curr. Biol. ogy*, 29, 530.e6–537.e6.
- Moodie, G.E.E. & Reimchen, T.E. (1976) Phenetic variation and habitat differences in *Gasterosteus* populations of Queen Charlotte Islands. *Syst. Zool.*, 25, 49–61.
- Moser, D., Frey, A. & Berner, D. (2016) Fitness differences between parapatric lake and stream stickleback revealed by a field transplant. *J. Evol. Biology*, 29, 711–719.
- Nadeau, N.J., Whibley, A., Jones, R.T., Davey, J.W., Dasmahapatra, K.K., Baxter, S.W., Quail, M.A., Joron, M., ffrench-Constant, R.H., Blaxter, M.L., et al. (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Phil. Transactions of the Royal Society of London. Series B: Biological Sciences*, 367, 343–353.
- Nadeau, N.J., Pardo-Diaz, C., Whibley, A., Supple, M.A., Saenko, S.V., Wallbank, R.W., Wu, G.C., Maroja, L., Ferguson, L., Hanly, J.J., et al. (2016) The gene *cortex* controls mimicry and crypsis in butterflies and moths. *Nature*, 534, 106–110.
- Nei, M. (1987) *Molecular evolutionary genetics*. Columbia Univ. Press, New York.
- Nelson, T.C. & Cresko, W.A. (2018) Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.*, 2, 9–21.
- Nelson, T.C., Crandall, J.G., Ituarte, C.M., Catchen, J.M. & Cresko, W.A. (2019) Selection, linkage, and population structure interact to shape genetic variation among threespine stickleback genomes. *Genetics*, 212, 1367–1382.

- Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. & Wang, J. (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *Plos One*, 7:e37558.
- Nosil, P., Villoutreix, R., deCarvalho, C.F., Farkas, T.E., Soria-Carrasco, V., Feder, J.L., Crespi, B.J. & Gompert, Z. (2018) Natural selection and the predictability of evolution in *Timema* stick insects. *Science*, 359, 765–770.
- Ohta, T. (1992) The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.*, 23, 263–286.
- O'Reilly, P., Reimchen, T.E., Beech, R. & Strobeck, C. (1993) Mitochondrial-DNA in *Gasterosteus* and pleistocene glacial refugium on the Queen-Charlotte-Islands, British-Columbia. *Evolution*, 47, 678–684.
- Orr, H.A. (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52, 935–949.
- Orti, G., Bell, M.A., Reimchen, T.E. & Meyer, A. (1994) Global survey of mitochondrial-DNA sequences in the threespine stickleback - evidence for recent migrations. *Evolution*, 48, 608–622.
- Peichel, C.L. & Marques, D.A. (2017) The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. T. Roy. Soc. B*, 372:20150486.
- Peichel, C.L., Nereng, K.S., Ohgi, K.A., Cole, B.L., Colosimo, P.F., Buerkle, C.A., Schluter, D. & Kingsley, D.M. (2001) The genetic architecture of divergence between threespine stickleback species. *Nature*, 414, 901–905.
- Price, T. (2008) Speciation in birds. Roberts and Co, Greenwood Village, CO.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., deBakker, P.I., Daly, M.J., et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, 81, 559–575.
- R Development Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A. & Kitano, J. (2018) The genomic landscape at a late stage of stickleback speciation: high genomic divergence interspersed by small localized regions of introgression. *Plos Genet.*, 14:e1007358.
- Reed, R.D., Papa, R., Martin, A., Hines, H.M., Counterman, B.A., Pardo-Diaz, C., Jiggins, C.D., Chamberlain, N.L., Kronforst, M.R., Chen, R., et al. (2011) *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, 333, 1137–1141.
- Reimchen, T.E. (1980) Spine deficiency and polymorphism in a population of *Gasterosteus aculeatus* - an adaptation to predators. *Can. J. Zool.*, 58, 1232–1244.
- . (1983) Structural relationships between spines and lateral plates in threespine stickleback (*Gasterosteus aculeatus*). *Evolution*, 37, 931–946.
- . (1989) Loss of nuptial dolor in threespine sticklebacks (*Gasterosteus aculeatus*). *Evolution*, 43, 450–460.
- . (1991) Trout foraging failures and the evolution of body size in stickleback. *Copeia*, 1991, 1098–1104.
- . (1992) Injuries on stickleback from attacks by a toothed predator (*Oncorhynchus*) and implications for the evolution of lateral plates. *Evolution*, 46, 1224–1230.
- . (1994) Predators and morphological evolution in threespine stickleback. Pp. 240–276 in M. A. Bell, & S. A. Foster, eds. The evolutionary biology of the threespine stickleback. Oxford Univ. Press, Oxford, NY.
- . (1995) Predator-induced cyclical changes in lateral plate frequencies of *Gasterosteus*. *Behaviour*, 132, 1079–1094.
- . (2000) Predator handling failures of lateral plate morphs in *Gasterosteus aculeatus*: functional implications for the ancestral plate condition. *Behaviour*, 137, 1081–1096.
- Reimchen, T.E. & Buckland-Nicks, J. (1990) A novel association between an endemic stickleback and a parasitic Dinoflagellate. 1. Seasonal cycle and host response. *Can. J. Zool.*, 68, 667–671.
- Reimchen, T.E. & Nosil, P. (2002) Temporal variation in divergent selection on spine number in threespine stickleback. *Evolution*, 56, 2472–2483.
- . (2004) Variable predation regimes predict the evolution of sexual dimorphism in a population of threespine stickleback. *Evolution*, 58, 1274–1281.
- . (2006) Replicated ecological landscapes and the evolution of morphological diversity among *Gasterosteus* populations from an archipelago on the west coast of Canada. *Can. J. Zool.*, 84, 643–654.
- Reimchen, T.E., Stinson, E.M. & Nelson, J.S. (1985) Multivariate differentiation of parapatric and allopatric populations of threespine stickleback in the Sangan river watershed, Queen-Charlotte islands. *Can. J. Zool.*, 63, 2944–2951.
- Reimchen, T.E., Bergstrom, C. & Nosil, P. (2013) Natural selection and the adaptive radiation of Haida Gwaii stickleback. *Evol. Ecol. Res.*, 15, 241–269.
- Reimchen, T.E., Frey, S., Brady, S.D. & Kingsley, D.M. (2019) Predictive covariation among trophic, isotopic, and genomic traits is consistent with intrapopulation diversifying selection. *Evol. Ecol. Res.*, 20, 231–245.
- Renaut, S., Mailliet, N., Normandeau, E., Sauvage, C., Derome, N., Rogers, S.M. & Bernatchez, L. (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philos. T. Roy. Soc. B*, 367, 354–363.
- Rennison, D.J., Stuart, Y.E., Bolnick, D.I. & Peichel, C.L. (2019a) Ecological factors and morphological traits are associated with repeated genomic differentiation between lake and stream stickleback. *Philos. T. Roy. Soc. B*, 374:20180241.
- Rennison, D.J., Delmore, K.E., Samuk, K., Owens, G.L. & Miller, S.E. (2019b) Shared patterns of genome-wide differentiation are more strongly predicted by geography than by ecology. *Am. Nat.*, 195, 192–200.
- Richards, E.J. & Martin, C.H. (2017) Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *Plos Genet.*, 13:e1006919.
- Rieseberg, L.H., Raymond, O., Rosenthal, D.M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J.L., Schwarzbach, A.E., Donovan, L.A. & Lexer, C. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, 301, 1211–1216.
- Roberts Kingman, G.A., Vyas, D.N., Jones, F.C., Brady, S.D., Chen, H.I., Reid, K., Milhaven, M., Bertino, T.S., Aguirre, W.E., Heins, D.C., et al. (2021) Predicting future from past: the genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv.*, 7:eabg5285.
- Roesti, M., Hendry, A.P., Salzburger, W. & Berner, D. (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol. Ecol.*, 21, 2852–2862.
- Roesti, M., Kueng, B., Moser, D. & Berner, D. (2015) The genomics of ecological vicariance in threespine stickleback fish. *Nat. Comm.*, 6, 8767.
- Rogers, S.M., Tamkee, P., Summers, B., Balabhadra, S., Marks, M., Kingsley, D.M. & Schluter, D. (2012) Genetic signature of adaptive peak shift in threespine stickleback. *Evolution*, 66, 2439–2450.
- Rousselle, M., Simion, P., Tilak, M.-K., Figue, E., Nabholz, B. & Galtier, N. (2020) Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *Plos Genet.*, 16:e1008668.
- Samuk, K., Owens, G.L., Delmore, K.E., Miller, S.E., Rennison, D.J. & Schluter, D. (2017) Gene flow and selection interact to promote adaptive

- divergence in regions of low recombination. *Mol. Ecology*, 26, 4378–4390.
- Schemske, D.W. & Bradshaw, H.D.Jr. (1999) Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl. Acad. Sci. USA*, 96, 11910–11915.
- Schluter, D. (2000) The ecology of adaptive radiation. Oxford Univ. Press, Oxford, U.K.
- Schluter, D. & Conte, G.L. (2009) Genetics and ecological speciation. *Proc. Natl. Acad. Sci. USA*, 106, 9955–9962.
- Seear, P.J., Rosato, E., Goodall-Copestake, W.P. & Barber, I. (2015) The molecular evolution of spiggin nesting glue in sticklebacks. *Mol. Ecol.*, 24, 4474–4488.
- Seehausen, O. (2004) Hybridization and adaptive radiation. *Trends Ecol. Evol.*, 19, 198–207.
- . (2013) Conditions when hybridization might predispose populations for adaptive radiation. *J. Evol. Biol.*, 26, 279–281.
- Sella, G. & Barton, N.H. (2019) Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annu. Rev. Genom. Hum. Genetics*, 20, 461–493.
- Sheehan, H., Moser, M., Klahre, U., Esfeld, K., Dell’Olivo, A., Mandel, T., Metzger, S., Vandenbussche, M., Freitas, L. & Kuhlemeier, C. (2016) *MYB-FL* controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. *Nat. Genet.*, 48, 159–166.
- Shimoyama, M., De Pons, J., Hayman, G.T., Laulederkind, S.J., Liu, W., Nigam, R., Petri, V., Smith, J.R., Tutaj, M., Wang, S.J., et al. (2015) The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res. earch*, 43, D743–D750.
- Shuart, N.G., Haitin, Y., Camp, S.S., Black, K.D. & Zagotta, W.N. (2011) Molecular mechanism for 3:1 subunit stoichiometry of rod cyclic nucleotide-gated ion channels. *Nat. Comm.*, 2, 457.
- Simpson, G.G. (1953) The major features of evolution. Columbia Univ. Press, New York.
- Sinervo, B. & Svensson, E. (2002) Correlational selection and the evolution of genomic architecture. *Heredity*, 89, 329–338.
- Smith, J.M. (1976) What determines the rate of evolution? *Am. Nat.*, 110, 331–338.
- Spoljaric, M.A. & Reimchen, T.E. (2007) 10 000 years later: evolution of body shape in Haida Gwaii three-spined stickleback. *J. Fish Biol.*, 70, 1484–1503.
- Stockwell, C.A., Hendry, A.P. & Kinnison, M.T. (2003) Contemporary evolution meets conservation biology. *Trends Ecol. Evol.*, 18, 94–101.
- Stuart, Y.E., Veen, T., Weber, J.N., Hanson, D., Ravinet, M., Lohman, B.K., Thompson, C.J., Tasneem, T., Doggett, A., Izen, R., et al. (2017) Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat. Ecol. Evolution*, 1, 158.
- Svensson, E.I., Arnold, S.J., Bürger, R., Csilléry, K., Draghi, J., Henshaw, J.M., Jones, A.G., De Lisle, S., Marques, D.A., McGuigan, K., et al. (2021) Correlational selection in the age of genomics. *Nat. Ecol. Evol.*, 5, 562–573.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, 43, D447–D452.
- Terekhanova, N.V., Logacheva, M.D., Penin, A.A., Neretina, T.V., Barmintseva, A.E., Bazykin, G.A., Kondrashov, A.S. & Mugue, N.S. (2014) Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *Plos Genet.*, 10:e1004696.
- Therkildsen, N.O., Wilder, A.P., Conover, D.O., Munch, S.B., Baumann, H. & Palumbi, S.R. (2019) Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science*, 365, 487–490.
- Tittle, R.K., Sze, R., Ng, A., Nuckels, R.J., Swartz, M.E., Anderson, R.M., Bosch, J., Stainier, D.Y., Eberhart, J.K. & Gross, J.M. (2011) *Uhrf1* and *Dnmt1* are required for development and maintenance of the zebrafish lens. *Dev. Biol.*, 350, 50–63.
- Van Belleghem, S.M., Vangestel, C., De Wolf, K., De Corte, Z., Möst, M., Rastas, P., De Meester, L. & Hendrickx, F. (2018) Evolution at two time frames: polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *Plos Genet.*, 14:e1007796.
- van Rijssel, J.C., Moser, F.N., Frei, D. & Seehausen, O. (2018) Prevalence of disruptive selection predicts extent of species differentiation in Lake Victoria cichlids. *Proc. R. Soc. B Biol. Sci.*, 285:20172630.
- Venables, W.N. & Ripley, B.D. (2002) Modern applied statistics with S. Springer, New York.
- Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. (2006) A map of recent positive selection in the human genome. *Plos Biol.*, 4:e72.
- Weber, J.N., Kalbe, M., Shim, K.C., Erin, N.I., Steinel, N.C., Ma, L. & Bolnick, D.I. (2016) Resist globally, infect locally: a transcontinental test of adaptation by stickleback and their tapeworm parasite. *Am. Nat.*, 189, 43–57.
- Wright, K.M., Lloyd, D., Lowry, D.B., Macnair, M.R. & Willis, J.H. (2013) Indirect evolution of hybrid lethality due to linkage with selected locus in *Mimulus guttatus*. *Plos Biol.*, 11:e1001497.
- Xie, K.T., Wang, G., Thompson, A.C., Wucherpfennig, J.I., Reimchen, T.E., MacColl, A.D.C., Schluter, D., Bell, M.A., Vasquez, K.M. & Kingsley, D.M. (2019) DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science*, 363, 81–84.
- Yeaman, S. (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci. USA*, 110, E1743–E1751.
- Yeaman, S. & Whitlock, M.C. (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65, 1897–1911.
- Zamani, N., Russell, P., Lantz, H., Hoepfner, M.P., Meadows, J.R., Vijay, N., Mauceli, E., di Palma, F., Lindblad-Toh, K., Jern, P., et al. (2013) Unsupervised genome-wide recognition of local relationship patterns. *BMC Genomics*, 14, 347.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G., et al. (2018) Ensembl 2018. *Nucleic Acids Res.*, 46, D754–D761.
- Zhou, X. & Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.*, 44, 821–824.

Associate Editor: M. Maan
 Handling Editor: A. McAdam

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

SUPPORTING INFORMATION
 SUPPORTING INFORMATION