EVOLUTIONARY BIOLOGY

Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution

Garrett A. Roberts Kingman[1], Deven N. Vyas[2], Felicity C. Jones[3], Shannon D. Brady[1], Heidi I. Chen[1], Kerry Reid[2], Mark Milhaven[2,4], Thomas S. Bertino[2], Windsor E. Aguirre[5], David C. Heins[6], Frank A. von Hippel[7], Peter J. Park[8], Melanie Kirch[3], Devin M. Absher[9], Richard M. Myers[9], Federica Di Palma[10], Michael A. Bell[11]*, David M. Kingsley[1,12]*, Krishna R. Veeramah[2]*

Similar forms often evolve repeatedly in nature, raising long-standing questions about the underlying mechanisms. Here, we use repeated evolution in stickleback to identify a large set of genomic loci that change recurrently during colonization of freshwater habitats by marine fish. The same loci used repeatedly in extant populations also show rapid allele frequency changes when new freshwater populations are experimentally established from marine ancestors. Marked genotypic and phenotypic changes arise within 5 years, facilitated by standing genetic variation and linkage between adaptive regions. Both the speed and location of changes can be predicted using empirical observations of recurrence in natural populations or fundamental genomic features like allelic age, recombination rates, density of divergent loci, and overlap with mapped traits. A composite model trained on these stickleback features can also predict the location of key evolutionary loci in Darwin's finches, suggesting that similar features are important for evolution across diverse taxa.

INTRODUCTION

Can evolutionary outcomes be predicted? Biologists have long been fascinated with this question, including Darwin and Wallace's anticipation of the existence of Morgan's sphinx moth based on orchid morphology (*1*, *2*), Vavilov's prediction of the types of morphological variants likely to occur in plants (*3*), and Gould's gedankenexperiment about replaying the tape of life (*4*). Natural examples of recurrent evolution provide a particularly favorable opportunity to study the mechanisms that influence evolutionary predictability, including molecular patterns (*5*, *6*).

Although the predictability of evolution may appear to be in conflict with the unpredictability of historical contingency, understanding the past can yield important insights into future evolution. For example, vertebrate populations frequently harbor large reservoirs of standing genetic variation (SGV) (*7*) that give independent populations access to similar raw genetic material to respond to environmental challenges, as observed in diverse species including songbirds, cichlid fishes, and the threespine stickleback (*Gasterosteus aculeatus*) (*8*–*11*). SGV is often apparent in divergent species or populations where it is pretested by natural selection and then distributed by hybridization to related populations. Thus filtered and capable of leaping up fitness landscapes, SGV can also drive rapid evolution (*12*), helping address a very real practical challenge to testing evolutionary predictions: time.

Longitudinal studies of evolving populations have been used to estimate the tempo and strength of selection on a variety of traits in different species (*13*–*18*). Rapid phenotypic evolution over contemporary time scales has enabled hypothesis testing against detailed observations at every step in the process. There is an increasing and impressive body of research examining the genomic consequences of these phenotypic changes in microbial, invertebrate, and vertebrate systems (*19*–*26*).

Stickleback fish provide an outstanding system for further study of the genomic basis of recurrent evolution. At the end of the last Ice Age, threespine stickleback, including anadromous populations that migrate from the ocean to freshwater environments to breed, colonized and adapted to countless newly exposed freshwater environments created in the wake of retreating glaciers around the northern hemisphere (*27*, *28*). This massively parallel adaptive radiation was facilitated by natural selection acting on extensive ancient SGV (*8*, *11*). Under the "transporter" hypothesis, these variants are maintained at low frequencies in the marine populations by low levels of gene flow from freshwater populations (*29*). Reuse of ancient standing variants has enabled identification of genomewide sets of loci that are repeatedly differentiated among long-established stickleback populations (*8*, *30*–*35*). In addition, SGV enables new freshwater stickleback populations to evolve markedly within decades (*17*, *36*–*38*), including conspicuous phenotypic changes in armor plates (*17*) and body shape (*39*).

The rapidity of stickleback evolution has made it possible to begin characterizing genomic and allele frequency changes seen in very young or newly established populations under intense directional selection on multiple traits (*18*, *36*–*38*, *40*–*43*). Here, we identify key molecular features that underlie repeated and rapid evolution of freshwater stickleback by comparing genomes from diverse extant populations with the earliest generation-by-generation changes in a detailed genomic time series from three newly founded populations. We identify several basic genomic and genetic features that

[1]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305-5329, USA. [2]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA. [3]Friedrich Miescher Laboratory of the Max Planck Society, Max-Planck-Ring, Tübingen, Germany. [4]School of Life Sciences, Arizona State University, Tempe, AZ 85281, USA. [5]Department of Biological Sciences, DePaul University, Chicago, IL 60614-3207, USA. [6]Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA 70118, USA. [7]Department of Community, Environment and Policy, Mel & Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ 85724, USA. [8]Department of Biology, Farmingdale State College, Farmingdale, NY 11735-1021, USA. [9]HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35806, USA. [10]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA. [11]University of California Museum of Paleontology, University of California, Berkeley, Berkeley, CA 94720, USA. [12]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA. *Corresponding author. Email: krishna.veeramah@stonybrook.edu (K.R.V.); sticklemack@gmail.com (M.A.B.); kingsley@stanford.edu (D.M.K.)

can be used to predict evolutionary outcomes in stickleback and show that they can predict genomic responses to selection in distantly related cichlids and Darwin's finches.

RESULTS

Global resequencing and EcoPeak identification

Previous whole-genome sequencing (WGS) of threespine stickleback identified 174 loci covering 1.2 Mb with alleles shared by common descent repeatedly selected in freshwater populations around the world (*8*). Just as human genetic diversity is greatest in Africa, where *Homo sapiens* arose (*44*), we hypothesized that the north Pacific region where stickleback originated (*27*) may contain a particularly rich pool of ancient adaptive alleles. To test this hypothesis, we generated whole-genome sequence data with 76–base pair (bp) paired-end Illumina reads for 38 new marine and 110 new freshwater stickleback, respectively (mean coverage of 5.5×) (sections S2, S4, S6, and S7). Combined with previous stickleback sequencing (*8*, *41*), our dataset includes 227 individual genomes: 135 genomes from 70 northeast Pacific populations in Alaska, Haida Gwaii, British Columbia, and Washington and 92 genomes from 62 populations in California, Japan, and the Atlantic coasts of North America, Iceland, and northern Europe (Fig. 1A and section S8).

We used two methods to identify loci repeatedly differentiated in freshwater populations, both based on the expectation that variants recurrently selected from SGV will be more similar among geographically separated freshwater populations than neutral loci (section S9). First, we used a genetic distance–based approach within overlapping 2500-bp windows tiled across the genome [as in the study by Jones *et al.* (*8*)]. While statistically powerful, this approach may miss younger loci with few differences between alleles and exhibits spatial resolution dependent on window size. Second, we analyzed the distribution of variants at individual bases across the genome, which has base pair–level resolution and less bias against younger loci, though at the cost of statistical power. After calling *P* value–based peaks of ecotypic (freshwater- or marine-associated) differentiation using both methods, we accepted calls at two stringency levels, either requiring agreement between the two analyses at 1% false discovery rate (FDR) (specific) or support from either at 5% FDR (sensitive). We refer to these peaks of ecotypic differentiation as EcoPeaks. We called EcoPeaks for different geographic sets of samples to find alleles that were either shared globally, within the northeast Pacific, or within other geographic regions.

Although results of the global analysis largely matched a previous report [79 of 81 most stringent calls from Jones *et al.* (*8*) in sensitive EcoPeaks ($P = 4.2 \times 10^{-21}$; table S3)], both the sensitive and specific call sets identified approximately five times as many Pacific EcoPeaks as global EcoPeaks, spanning sevenfold more of the genome (Fig. 1, E and F, and Table 1). In addition, many northeast Pacific EcoPeaks not overlapping the globally shared regions identified by Jones *et al.* (*8*) exhibit even more consistent ecotypic differentiation (assessed by *P* values) than others shared around the world (Fig. 1, B and C). Much smaller sets of non-global EcoPeaks were identified in the North Atlantic, subglacial Pacific, and supraglacial geographic regions (fig. S5), consistent with other reports (*8*, *35*).

As theoretical studies indicate that SGV is immediately available for evolution and may show an increased likelihood of large-effect alleles being advantageous compared to de novo mutations (*12*, *45*), the rich genetic reservoir observed in the northeast Pacific provides a favorable system for studying the dynamics and predictability of rapid evolutionary change (section S10). Previous studies suggest that stickleback in the northeast Pacific can adapt to freshwater environments within decades (*36*). However, thus far, studies have lacked temporal resolution of genome evolution in the critical early years of adaptation.

Rapid contemporary evolution and TempoPeak identification

To characterize the earliest stages of evolution after the establishment of new freshwater populations, we analyzed annual samples from populations that were recently founded by anadromous stickleback in three lakes in Alaska (Fig. 2A and section S1). In 1982, stickleback in Loberg Lake (LB) were exterminated to improve recreational fishing (*17*). Sometime between 1983 and 1988, LB was invaded by completely plated (~33 plates per side) anadromous stickleback [most likely from neighboring Rabbit Slough (RS)]. The characteristic freshwater, armor-reduced phenotype increased rapidly from ~16% in 1991 to ~50% by 1995 and to ~95% by 2017 (Fig. 2B) (*17*), with similarly rapid changes in overall body shape (*39*) and reproductive patterns (*46*). So as to more systematically examine even earlier generations of freshwater adaptation, Bell *et al.* (*47*) introduced ~3000 anadromous RS fish into each of two other Cook Inlet lakes without outlets that had been similarly treated to exterminate fish: Cheney Lake (CH) in 2009 and Scout Lake (SC) in 2011. Low–armor-plated (~5 to 7 plates per side) stickleback began to appear in the second and third generation after founding in CH and SC respectively, and, by 2017, they had increased to 20 to 30% (Fig. 2B).

To obtain genomewide allele frequencies across our time series, we performed pooled WGS (pool-seq) on all seven available annual samples from CH and SC since founding and eight from LB distributed between 1999 and 2017 (Fig. 2A and sections S3, S4, S7, and S13). Each freshwater pool-seq experiment consisted of 100 individuals (with three exceptions), with mean coverage of 223× per pool. In addition, we resequenced a pool of 200 anadromous RS individuals used to found the CH population in 2009 (RS2009) to 585×.

We identified single-nucleotide polymorphisms (SNPs) with significant allele frequency changes, indicating directional selection, using a modified Cochran-Mantel-Haenszel (CMH) test optimized for pool-seq data (*48*), followed by an approach analogous to our EcoPeak analysis to define both a permissive "sensitive" and a stringent "specific" set of loci that we term TempoPeaks (sections S16 to S18). Combining all three populations into a single CMH analysis (CH + SC + LB) and using RS2009 as a proxy for the founders of LB, we identified 524 sensitive and 344 specific TempoPeaks. Despite operating over very different time spans, the visual correspondence between the Pacific EcoPeaks in long-established populations and the TempoPeaks in recently established populations is notable, particularly for the specific TempoPeaks, of which 323 of 344 (94%) overlap with the sensitive Pacific EcoPeaks (Fig. 2D and section S18). In contrast, even the most lenient set of global EcoPeaks and regions from Jones *et al.* (*8*) overlap only 96 of 344 (28%) and 47 of 344 (14%) specific TempoPeaks, respectively (tables S9 and S10), emphasizing the importance of understanding the locally available SGV. Even analyzing only CH + SC (thus focusing on <10 years of freshwater adaptation), we identified 271 sensitive and 86 specific TempoPeaks, 73% and 99% of which, respectively, overlap the sensitive Pacific EcoPeaks. This marked congruity strongly suggests that the ancient SGV represented by Pacific EcoPeaks is the primary

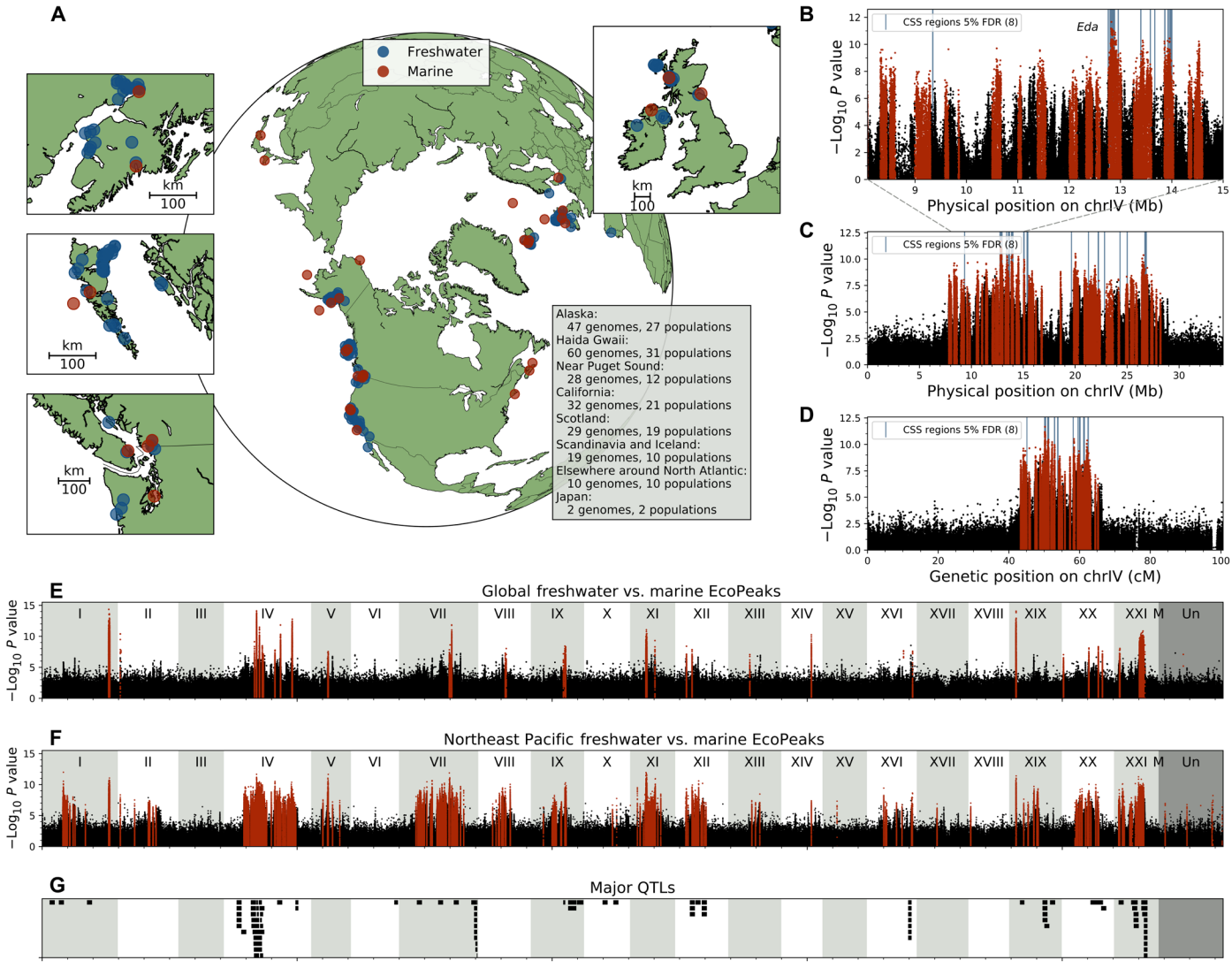

Fig. 1. Recurrent peaks of ecological sequence differentiation between marine and freshwater stickleback from different regions of world. (**A**) Marine (red) and freshwater (blue) stickleback from the locations shown were used for various analyses (table S2). (**B**) Detail of part of chrIV for single-nucleotide polymorphism (SNP)–based analysis of differential allele distribution between marine and freshwater ecotypes in the northeast Pacific basin. SNPs within specific-threshold EcoPeaks are red. A subset of regions overlap the globally shared peaks of marine-freshwater differentiation indicated by blue-colored bars [cluster separation score (CSS), 5% false discovery rate (FDR) identified by Jones *et al*. (*8*)]. (**C**) As in (B), but for the whole chromosome [dashed lines from (B) to (C)]. (**D**) Same whole chromosome as in (C), but with genetic (not physical) distance along the *x* axis. (**E** and **F**) Genomewide SNP divergence between marine and freshwater ecotypes globally and in the northeastern Pacific basin, with specific-threshold EcoPeaks in red. (**G**) Many differentiated regions overlap the location of major quantitative trait loci (QTLs) controlling various morphological, physiological, and behavioral traits in previous genetic crosses [percent variance explained (PVE) > 20, interval < 5 Mb from Peichel and Marques (*53*)].

Table 1. Overview of EcoPeaks and TempoPeaks. The comparisons by Jones *et al*. (*8*) are with the cluster separation score 5% FDR set (*8*).

	Global EcoPeaks (specific)	Pacific EcoPeaks (specific)	Pacific EcoPeaks (sensitive)	Cheney + Scout	All young
No. of regions	39	209	212	86	344
Total bases (%)	3.7 Mb (0.78%)	27.4 Mb (5.82%)	91.9 Mb (19.53%)	3.3 Mb (6.95%)	17.57 Mb (3.73%)
Median size	21.4 kb	80.2 kb	122.9 kb	21.7 kb	27.3 kb
Recovery of regions identified by Jones *et al.* (*8*)	86/174 (49.4%)	112/174 (64.3%)	158/174 (90.8%)	47/174 (27.0%)	98/174 (56.3%)
Fraction in regions identified by Jones *et al.* (*8*)	18/39 (46.2%)	29/209 (13.9%)	33/212 (15.6%)	10/86 (11.6%)	24/344 (7.0%)

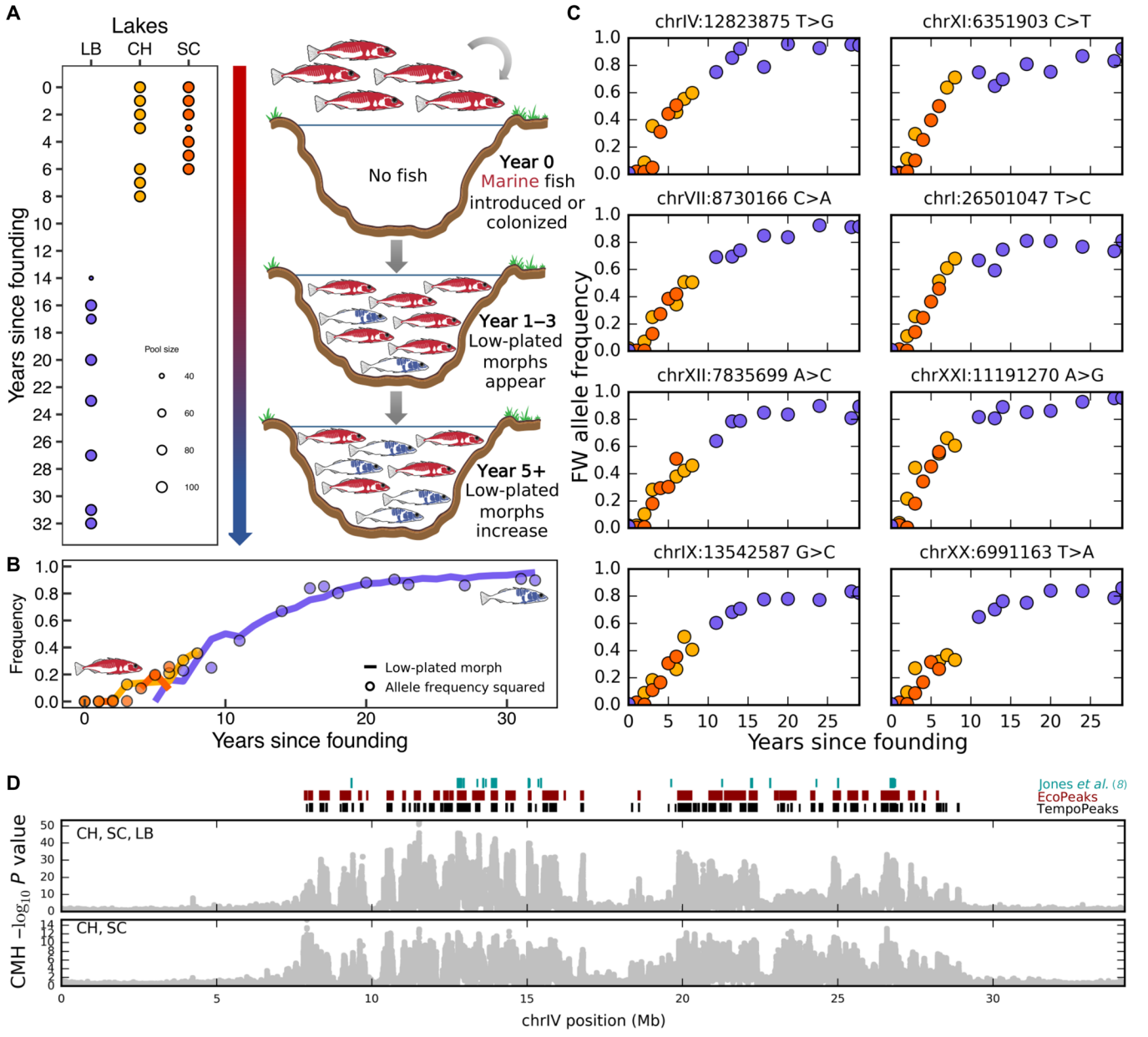

Fig. 2. Contemporary evolution occurring in freshwater transplants in Cook Inlet, Alaska. (**A**) The timing (years since founding) and approximate size of subsequent sequencing sample pools from lake populations [Loberg Lake (LB), Cheney Lake (CH), and Scout Lake (SC)] founded recently by anadromous stickleback (left) and the scenario for divergence of anadromous populations after colonizing the lakes (right). Red and blue fish represent the complete armor-plated and armor-reduced phenotypes, respectively. (**B**) Frequency of armor-reduced morphological phenotype across our CH, SC, and LB time series overlaid with the frequency squared for the freshwater (FW) *Eda* allele. LB data are based on a combination of individual genotypes and pool-seq frequencies, while CH and SC are based only on pool-seq frequencies. (**C**) Allele frequency trajectories for eight SNPs found within TempoPeaks on distinct chromosomes with the highest Cochran-Mantel-Haenszel (CMH) scores (except for chrIV:12823875, the *Eda*-plate regulatory region SNP). (**D**) Genomewide distribution of window-based CMH scores across chrIV for different combinations of transplant lakes discussed in the main text. Black, dark red, and teal bars above figure represent specific CH + SC + LB TempoPeaks, northeast Pacific EcoPeaks, and significant loci from Jones *et al.* (*8*) identified using CSS [5% FDR (*8*)], respectively.

genomic feature enabling extremely fast evolution of freshwater phenotypes in stickleback from the northeast Pacific basin.

The *Eda* SNP associated with armor plate variability (chrIV: 12,823,875 T>G (*49*)) is within the second most significant specific TempoPeak on chrIV. In both CH and SC, the G allele increases rapidly from an initial frequency of <1% to over 50% within 8 years, while approaching fixation in LB by 15 years. Notably, the square of G-allele frequencies (i.e., the expected number of GG homozygotes) tracks closely with frequencies of the low–armor plate phenotype, consistent with almost complete recessiveness (h = 0.0) for the G allele for this phenotype (Fig. 2B). Nonetheless, to fit the allele frequency trajectory of this SNP, and, in particular, the extremely rapid

increase in CH and SC, it was necessary to impose a dominance coefficient (h) of 1.0 along with a very large selection coefficient (s) of 0.55, as in a recent paper focusing on this locus (*18*).

Like *Eda*, most TempoPeaks display similarly sharp left-shifted sigmoidal allele frequency trajectories, indicating very strong and dominant-positive selection (Fig. 2C and section S20). When modeling each peak SNP as independent, we find an extremely high mean s of 0.30 (5th, 95th percentile 0.08 to 0.53) and h of 0.98 (5th, 95th percentile 0.95 to 1.0) for the 344 specific TempoPeaks found in CH + SC + LB. The estimated s values for chrIV, where there are 69 TempoPeaks, are particularly high (mean s = 0.38), consistent with the accelerated evolution of this whole chromosome observed via a chromosome-wide F_{ST} analysis comparing the founding generation of CH, SC, and LB to all subsequent years (section S15).

Features associated with EcoPeak evolution

The remarkable speed at which northeast Pacific stickleback adapt to new freshwater environments suggests that analysis of EcoPeaks may provide unique insights into optimal genomic properties for evolution. Using *Gasterosteus nipponicus*, *Gasterosteus wheatlandi*, and *Pungitius pungitius* for calibration, we estimated molecular divergence time between a pair of freshwater (Little Campbell upstream) and marine (Little Campbell downstream) stickleback in windows tiled across the genome (section S11). We find that EcoPeaks as a whole are significantly older than the rest of the genome [1600 thousand years (ka) versus 700 ka, $P < 5 \times 10^{-324}$]. Although peaks shared globally trend older than those found just within the northeast Pacific (1800 ka versus 1600 ka, $P = 0.18$), the imputed ages overlap considerably (Fig. 3A). We estimate that the majority (161 of 209) are over a million years old and have cycled between freshwater and marine environments many times during this long history, likely persisting at high frequency in freshwater habitats south of the zone of glaciation during the Ice Ages and at more northerly latitudes during previous interglacials and the Holocene.

Contrary to our expectations that recombination would disassemble regions over time, we found that older EcoPeaks are larger than younger ones (Fig. 3B). This signature is strongest at the most significant markers within each EcoPeak, which are typically older than more distal sequences (Fig. 3C). This suggests that individual regions may grow over time, with alleles originally based on an initial

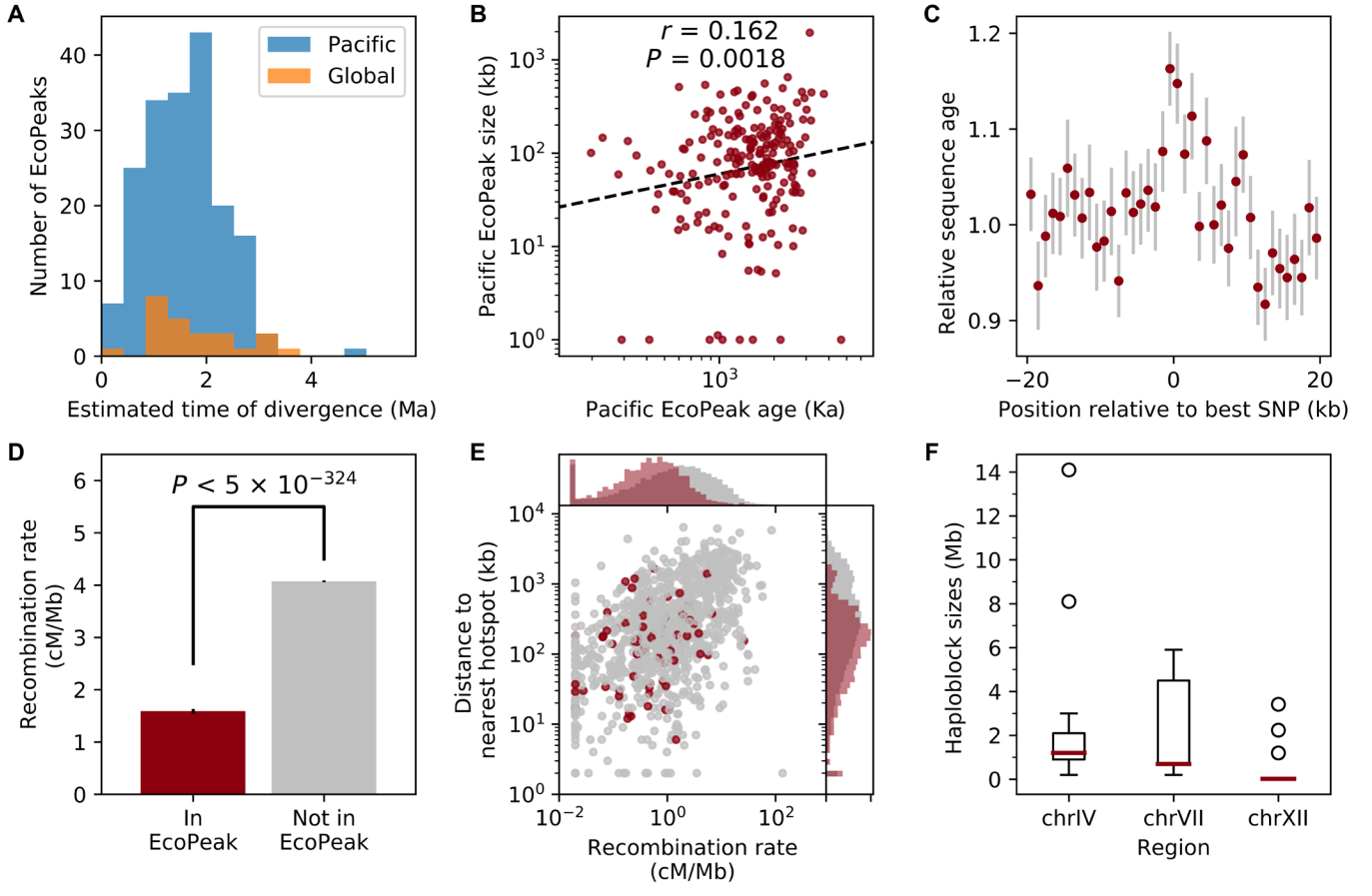

Fig. 3. EcoPeak associations with age, region size, and recombination rate. (**A**) Distribution of estimated molecular age for those EcoPeaks either shared worldwide (orange) or within the northeast Pacific (blue). Ma, million years. (**B**) EcoPeaks with older estimated molecular ages tend to be larger. (**C**) Estimated ages decline with distance on either side of EcoPeaks. Each dot represents mean age in 1-kb windows flanking the EcoPeak centers (gray bars, 1 SE). (**D**) Recombination rates tend to be lower within EcoPeaks compared to the rest of the genome, ±1 SE. (**E**) Recombination rates and distances to nearest 20× recombination hotspots, plotted for randomly subsampled 1-kb windows tiled across the genome, with marginal histograms of all windows. Locations overlapping EcoPeaks (red) are shifted to both smaller hotspot distances and lower recombination rates compared to other genomic regions (gray). (**F**) Observed haploblock size in marine fish carrying freshwater EcoPeaks on the indicated chromosomes across three marine populations. For all, specific northeast Pacific EcoPeaks are used.

beneficial mutation accumulating additional linked favorable mutations, snowballing over time to form a finely tuned haplotype with multiple adaptive changes. This is consistent with work in other species identifying examples of evolution through multiple linked mutations that together modify function of a gene (*50*–*52*) and implies that progressive allelic improvement may be common.

We also observed that EcoPeaks frequently overlap major quantitative trait loci (QTLs) in stickleback [73 of 209 overlaps observed versus 32 of 209 expected, $P < 1 \times 10^{-15}$; Fig. 1G (*53*)], suggesting that these variants underlie many mapped phenotypic traits. Just as the QTLs cluster in "supergene" complexes (*54*), so too do EcoPeaks (median observed interpeak distance 192 kb versus 795 kb expected, $P = 4.88 \times 10^{-10}$). One particularly large complex (chrIV: 8 to 17 Mb) contains 22 EcoPeaks and the major QTLs controlling many aspects of both defensive armor and trophic morphology (e.g., the length of dorsal and pelvic spines, the number of armor plates through *Eda*, gill rakers, and teeth). Thus, clustering may have important functional effects by allowing multiple traits and underlying EcoPeaks to be selected and inherited as a single unit, especially when in tight linkage. A fine-scale recombination map of RS stickleback (generated with LDhelmet (*55*)) shows that EcoPeaks are highly enriched in regions of low average recombination, forming tightly linked haploblocks (Fig. 3D, compare Fig. 1, C and D; section S14). EcoPeaks are also enriched near local recombination hotspots within their neighborhood (Fig. 3E), potentially facilitating reassembly of larger haplotype blocks upon freshwater colonization (also see section S19).

To further examine the frequency and size of haploblocks in individual fish, we surveyed 1643 stickleback from three Alaskan marine populations by SNP array genotyping (sections S5 and S12). While most marine fish heterozygous for freshwater alleles carry a relatively small haploblock, some carry multi-megabase haploblocks containing multiple EcoPeaks (Fig. 3F). Thus, a proper treatment of rapid stickleback evolution needs to account for the complex linkage of EcoPeaks rather than treating them independently.

Modeling the genomic landscape of contemporary evolution

To estimate a more realistic distribution of fitness effects (DFE) that incorporates the genome's recombination landscape, we developed a deep neural network (DNN) approach that uses forward simulations (section S21). Our simulations, which are conceptually similar to those of Galloway *et al.* (*56*), attempted to replicate the dynamics of the "transporter model" (*29*), with one large (N_e = 10,000) anadromous population connected independently by gene flow to 10 smaller (N_e = 1000) established freshwater populations. After 1000 generations, we founded three new freshwater populations from the anadromous population, thus generating simulated allele frequency trajectories that reflect our annual LB, CH, and SC samples (Fig. 4A).

Focusing our DNN analysis on a subset of 19 specific TempoPeak SNPs separated by ≥0.4 cM (~100 kb) along chrIV, we closely replicated observed allele trajectories of positively selected freshwater alleles across all SNPs simultaneously using a beta distribution–shaped DFE, for which the mean *s* across the 19 TempoPeaks was 0.063 and the standard deviation was 0.030, with reciprocal fitness costs implemented in the marine population (Fig. 4C). The estimated *s* from our DNN was thus substantially smaller than the mean of 0.48 when each SNP was considered independently. In addition, 18 of 19 SNPs were predicted to be fully dominant and none fully recessive under the best model.

We validated our best-fit DNN model by simulating the 19 selected TempoPeaks SNPs with the estimated DFE along with ~400k neutral SNPs distributed randomly along chrIV. Despite the neutral SNPs not being used in training the DNN, we were able to mimic the overall topology of the CMH scores across the entire genome, suggesting that our model was capturing the overall genomic architecture of freshwater adaptation (Fig. 4D). Our best-fit DNN model also appeared to recapitulate much of the haplotype structure of the array data from individuals from RS, LB1999, and LB2013 (Fig. 4B). Notably, the transition to freshwater alleles appears to be somewhat slower on the right half of chrIV, where there are fewer EcoPeaks, TempoPeaks, and QTLs, and this difference was observable in both the empirical and simulated data.

Overall, our model suggests that extremely rapid and replicable allele frequency increases on chrIV in LB, CH, and SC are mostly driven by multiple linked (primarily) dominant alleles, each with relatively smaller *s* values that act in concert, with recombination hotspots between them (section S19) allowing rapid reassembly of optimum freshwater haplotypes, consistent with the transporter hypothesis. The lower individual *s* values may allow these dominant alleles to persist in the marine environment at low frequency after being disassembled by recombination, especially if some act in epistasis.

Biological features with predictive power

Given the genomewide dynamism of the earliest stages of freshwater adaptation, we attempted to identify genomic features that predict the speed of evolution at TempoPeaks and understand why some peaks are consistently selected more rapidly than others (section S22). We used CMH scores as a proxy of evolutionary speed for each TempoPeak in CH + SC + LB and regressed these against a variety of sequence features.

The best predictor for the speed of evolution is the degree of ecotypic differentiation between marine and long-established freshwater populations (Pacific EcoPeak *P* value), with variants more commonly differentiated in the northeast Pacific being selected more quickly (Fig. 5A and fig. S81). Fisher's geometric model indicates that alleles with large effects are usually disfavored; however, the "prefiltering" of ancient SGV that counters this tendency (*12*) largely benefits alleles that are broadly positively selected, possibly explaining this result.

We also found that larger TempoPeaks are typically selected more rapidly. Similarly, greater TempoPeak density predicts more rapid divergence, suggesting that our simulation accurately reflects how nearby loci mutually reinforce their collective selection. Overlap with major QTLs also has a strong association with rapid evolution, while other variables such as increased sequence divergence, decreased recombination rate, increased gene overlap, increased sequence conservation, increased Ka/Ks, and decreased ancestral marine frequency have smaller contributions to predictive power for speed of selection (Fig. 5A).

We also tested whether underlying sequence characteristics could predict not only the speed of selection in CH + SC + LB but also the location of the selected regions themselves (section S23). Recombination rate, QTL overlap, allelic age, and an integrated genomic context score (section S23) that incorporate the previous features are all useful predictors (Fig. 5B). By combining these fundamental features into a logistic model trained on the survey of extant populations, the most confident predictions of selected regions in the rest of the genome achieve 85% precision. This model performs 67% as well as predictions based only on empirical repeatability in extant populations in the northeast Pacific (Fig. 5B). Thus, our understanding of underlying principles reflects an incomplete yet substantial proportion of evolutionary repeatability.

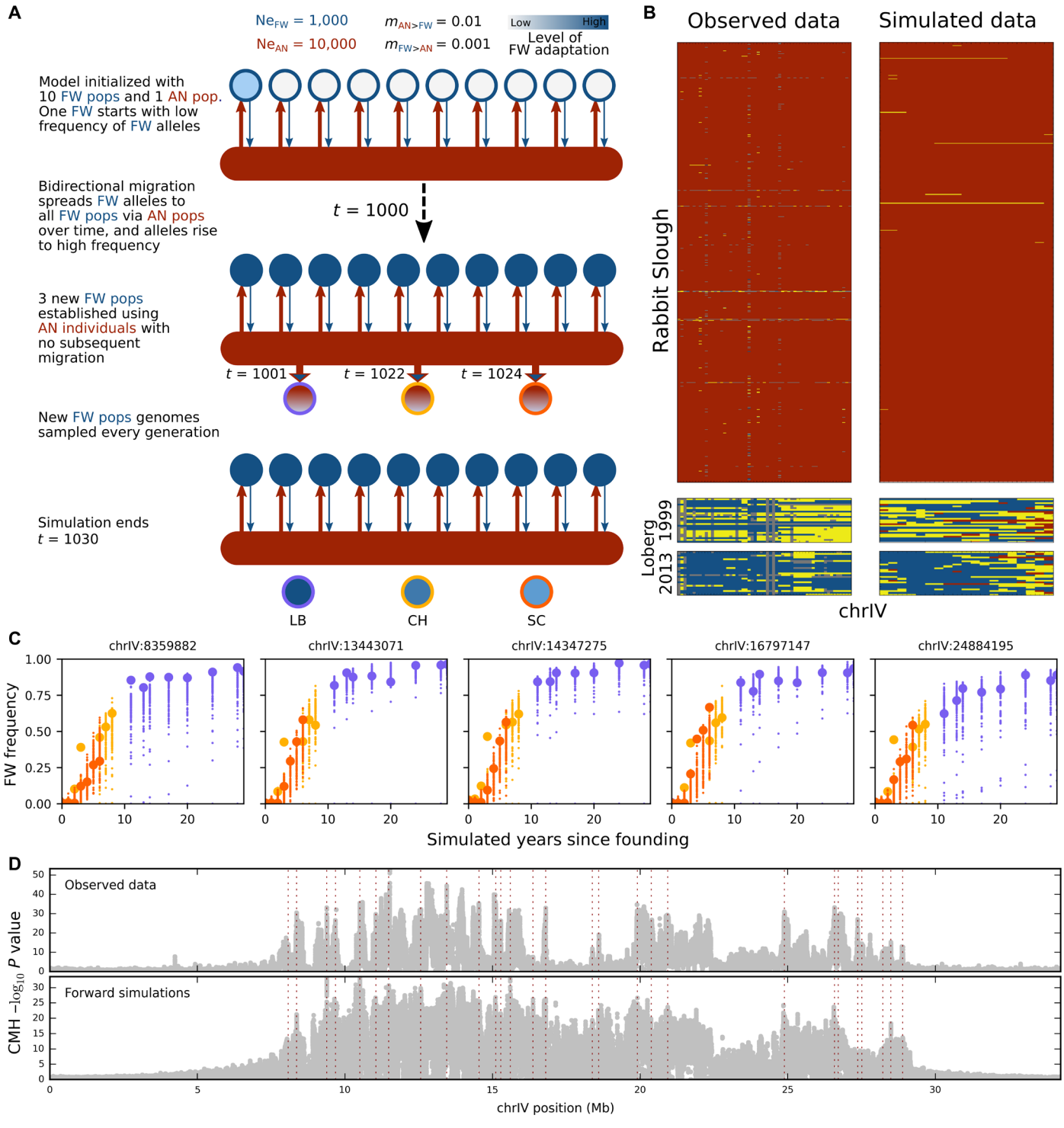

Fig. 4. DNN simulation–based modeling of rapid and repeated stickleback evolution. (**A**) Schematic showing evolutionary model of forward simulations under the transporter hypothesis. Red horizontal bars, anadromous (AN) ancestor; blue circles, descendant freshwater isolates; red to blue shaded circles, three adapting freshwater populations (i.e., LB, CH, and SC) founded recently by anadromous stickleback; and arrows, gene flow or founding events. (**B**) Genotypes across chrIV for freshwater-associated SNPs in RS ($n = 750$), LB in 1999 ($n = 25$), and LB in 2013 for (left) observed and (right) simulated data under best-fit DNN model. anadromous homozygous, red; heterozygous, yellow; and freshwater homozygous genotypes, blue; respectively. (**C**). Allele frequency trajectories for LB, CH, and SC in 100 simulations under the best-fit DNN model for five randomly selected SNPs. Larger points, observed data. (**D**) Distribution of average CMH scores in windows of 2500 bp across chrIV for (top) observed and (bottom) simulated data under best-fit DNN model. Red dotted lines, locations of SNPs under selection and used to fit DNN.

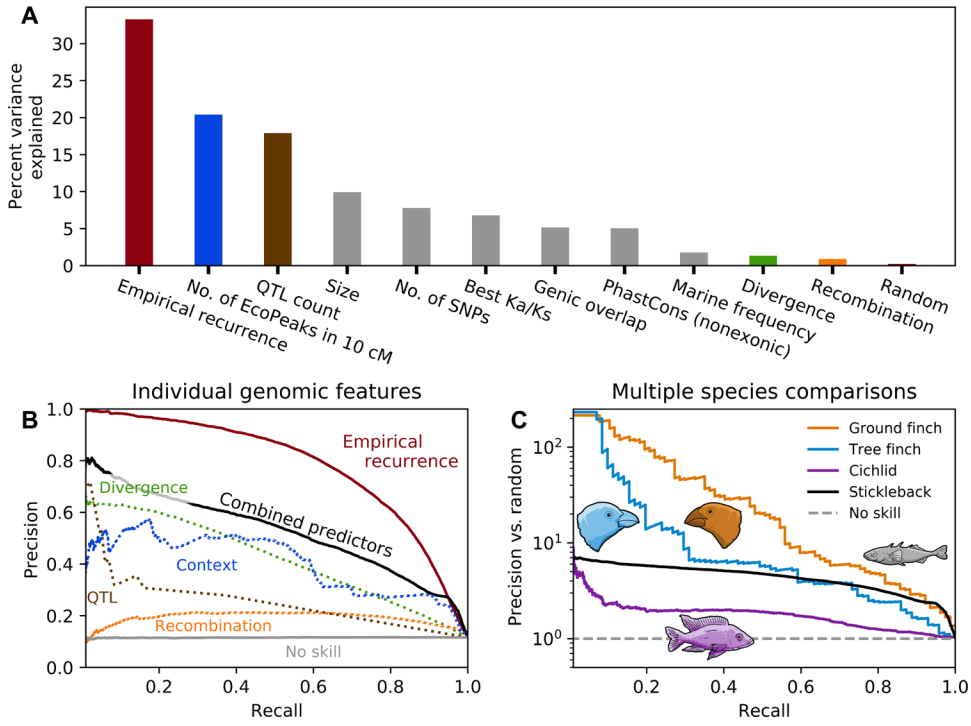

Fig. 5. Properties underlying speed and locus of selection in stickleback, cichlids, and Darwin's finches. (**A**) Variance in the speed of TempoPeak selection explained by different underlying genomic features, including colored bars: empirical recurrence of marine-freshwater differentiation (peak Pacific ecotypic *P* value), number of additional Pacific EcoPeaks within 10 cM, number of major QTLs overlapped, sequence divergence, and recombination rate; gray bars: genomic size of EcoPeak, total number of variable nucleotides, elevated Ka/Ks in coding regions, overlap with genic sequences, overlap with conserved noncoding sequence (PhastCons nonexonic), and carrier frequency of freshwater alleles in marine populations. (**B**) Precision-recall curve for predicting the locations of selected loci in CH + SC + LB lakes by either individual genomic features (dotted lines), a composite model trained with these basic predictors, or the empirical expectation of recurrence based on many extant populations. Precision is the fraction of predictions that are accurate, while recall is the fraction of true positives that are correctly predicted. "No skill" refers to the performance expected by random chance. (**C**) Performance above chance of the composite model applied to stickleback, cichlids, and two representative pairs of species of Darwin's finches (ground finches: *Geospiza magnirostris* versus *Geospiza propinqua*; tree finches: *Camarhynchus pauper* versus *Camarhynchus psittacula*).

Parallels in distant species

To test the generality of these predictive factors, we applied the stickleback-trained model to a dataset of 12 pairs of species of Darwin's finches (section S23) (*57*). Darwin's finches have undergone adaptive radiation in the Galápagos Islands over the last several hundred thousand years, are ~435 million years divergent from stickleback, and face very different selective pressures. As in stickleback, however, the "islands of divergence" of all 12 analyzed pairs of species of Darwin's finches (sensu Han *et al.*) are enriched for ancient alleles overlapping mapped QTLs with low recombination rates. The top 100 windows predicted by the stickleback model recover a median of 28-fold more previously identified islands of divergence than expected by chance ($P < 1 \times 10^{-10}$; Fig. 5C), including the *Alx1* and *Hmga2* loci implicated in beak morphology in multiple species pairs (even without QTL input). The model also recovers a substantial proportion of differentiated loci in a recent case of cichlid speciation (*58*). Thus, a handful of basic genomic properties allow strong quantitative predictions of the location of key evolutionary loci, even across widely separated branches of life.

DISCUSSION

The importance of SGV for evolution is becoming increasingly apparent, especially in species with large genome sizes (*59*), including humans (*60*). At first glance, the dependence of threespine stickleback on SGV for freshwater adaptation may appear to be a peculiarity in terms of repeatability and speed and their particular natural history. However, by more comprehensively understanding the dynamics of this highly optimized process, we have extracted general features of genome architecture and evolution that successfully translate to species on distant branches of the tree of life, thus demonstrating the tremendous power of the stickleback system to identify unifying principles that underlie evolutionary change.

MATERIALS AND METHODS

Sample collection and DNA preparation

Fish for all downstream genomic analyses were trapped following Institutional Animal Care and Use Committee (IACUC) guidelines using unbaited minnow traps set near shore, immediately euthanized with MS-222 (tricaine methanesulfonate), and then preserved in 70 or 95% ethanol (section S1). DNA extraction was performed using either phenol:chloroform isolation (*61*) and quantified using a NanoDrop spectrophotometer or using the DNeasy 96 Blood & Tissue Kit following the standard "animal tissue" protocol and quantified using the Qubit High-sensitivity DNA Assay (section S2). Equimolar pooling of samples from RS, CH, SC, and LB was performed using an Opentrons OT-2 robot (section S3).

Genome sequencing and genotyping

Samples from long-established populations underwent WGS sequencing on a HiSeq 2000 using 2 × 76-bp paired-end sequencing libraries. Contemporary pools were sequenced on either a NovaSeq 6000 or Illumina HiSeq 2500 using 2 × 150-bp paired-end sequencing libraries. Contemporary WGS was performed by Beijing Genomics Institute using their proprietary DNBseq technology using 2 × 100-bp paired-end libraries (section S4). A custom Illumina 384 GoldenGate array was designed for SNP genotyping (section S5).

Bioinformatic processing

We constructed a slightly modified reference genome based on the recent Hi-C–guided improvement of the stickleback genome (*62*), to which we refer as *gasAcu1-4*, that includes a new chrP and a new mitochondrial genome (section S6). Reads were mapped to *gasAcu1-4* using bwa mem (*63*) and Picard was used to add read groups and mark duplicate reads. Indel realignment and base quality recalibration were performed using Genome Analysis Toolkit (GATK) (*64*). HaplotypeCaller and GenotypeGVCFs were used for variant calling for WGS data. Allele frequencies in pool-seq populations were calculated using the maximum-likelihood method of Lynch *et al.* (*65*) and PoPoolation2 (v1201) (section S7) (*66*).

Analysis

EcoPeaks were identified using two approaches, the first following the same genetic distance–based approach as Jones *et al.* (*8*) based on 2500-bp windows sliding every 500 bp, and the second analyzing the distribution of allele counts between marine and freshwater populations at every base position in the genome with two alleles present at >10% frequency in the combined analysis metapopulation. For both the SNP-based and 2500-bp window–based analyses, nearby significant values were grouped into the EcoPeaks that behaved as a single unit using a greedy algorithm. Peaks were filtered at either a 1% FDR for the specific calls or at 5% for the sensitive calls. The single base and window peaks were then intersected for the final specific calls or unioned for the final sensitive calls (section S9). Allelic divergence and age were computed from five upstream (freshwater) and five downstream (marine) fish from Little Campbell River. Nonoverlapping 1-kb windows were tiled across the genome, and variants homozygous for different alleles were counted and used to compute marine-freshwater sequence divergence d (section S11). A rho-based recombination map was constructed on the basis of 20 RS genomes using LDhelmet (*55*) following a similar methodology to that of Shanfelter *et al.* (*67*), though with some minor modifications, and converted to genetic distance based on the pedigree-based linkage map generated by Glazer *et al.* (*68*) (section S14). F_{ST} was calculated for each pool-seq population against its youngest counterpart using the ratio of averages method implemented by Bhatia *et al.* (*69*) (section S18). We used a modified CMH test (*48*) to identify SNPs that had shown a significant change in allele frequency in our contemporary time-series data (section S17). We followed the same general EcoPeak identification methodology to define TempoPeaks from our CMH P values. Sensitive TempoPeaks were based on a 5% Bonferroni-corrected P value threshold merging SNP and window-defined peaks. Specific TempoPeaks were based on a 1% P value threshold only considering window-defined peaks (section S18). We applied the deterministic method described by Taus *et al.* (*70*) to estimate s for the SNP with the largest CMH score in each significant TempoPeak. We estimated s and $p0$ (initial allele frequency), both assuming that the dominance coefficient h is 0.5, as well as simultaneously estimating s, $p0$, and h (section S20). We additionally developed a DNN approach to estimate the DFE of multiple linked TempPeaks on chrIV. This analysis includes three main stages: (i) simulating linked loci with positive selection parameterized with various randomly drawn DFEs within the context of the demographic and evolutionary model of the transporter hypothesis, (ii) using the DNN framework to estimate the best DFE given the observed allele frequency trajectory data from our pool-seq experiments, and (iii) comparing the transporter hypothesis under the best-fit DFE to various features of the observed genomic data (section S21). Genomic features of interest were then used to predict speed of TempoPeak selection in the contemporary evolution populations using a linear regression model (section S22). We also applied genomic features of interest to predict the genomic loci of selection in the contemporary evolution populations via multivariate logistic regression. Last, we applied this model with the same terms and weights to datasets from Darwin's finches (*57*) and Lake Victoria cichlids (*58*) using the previously published analyses as our truth sets (section S23).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/25/eabg5285/DC1

REFERENCES AND NOTES

1. C. Darwin, *Fertilisation of Orchids* (Murray, 1862).
2. A. R. Wallace, Creation by Law. *Q. J. Sci.* **4**, 470–488 (1867).
3. N. I. Vavilov, The law of homologous series in variation. *J. Genet.* **12**, 47–89 (1922).
4. S. J. Gould, *Wonderful Life* (Norton, 1989).
5. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).
6. D. Blank, L. Wolf, M. Ackermann, O. K. Silander, The predictability of molecular evolution during functional innovation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3044–3049 (2014).
7. Z. D. Blount, R. E. Lenski, J. B. Losos, Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).
8. F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya; Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. D. Palma, K. Lindblad-Toh, D. M. Kingsley, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
9. Y.-T. Lai, C. K. L. Yeung, K. E. Omland, E.-L. Pang, Y. Hao, B.-Y. Liao, H.-F. Cao, B.-W. Zhang, C.-F. Yeh, C.-M. Hung, H.-Y. Hung, M.-Y. Yang, W. Liang, Y.-C. Hsu, C.-T. Yao, L. Dong, K. Lin, S.-H. Li, Standing genetic variation as the predominant source for adaptation of a songbird. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2152–2157 (2019).
10. Y.-H. E. Loh, E. Bezault, F. M. Muenzel, R. B. Roberts, R. Swofford, M. Barluenga, C. E. Kidd, A. E. Howe, F. Di Palma, K. Lindblad-Toh, J. Hey, O. Seehausen, W. Salzburger, T. D. Kocher, J. T. Streelman, Origins of shared genetic variation in African cichlids. *Mol. Biol. Evol.* **30**, 906–917 (2013).
11. P. F. Colosimo, K. E. Hosemann, S. Balabhadra, G. Villarreal Jr., M. Dickson, J. Grimwood, J. Schmutz, R. Myers, D. Schluter, D. M. Kingsley, Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
12. R. D. H. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
13. J. A. Endler, *Natural Selection in the Wild* (Princeton Univ. Press, 1986).
14. J. G. Kingsolver, H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, P. Beerli, The strength of phenotypic selection in natural populations. *Am. Nat.* **157**, 245–261 (2001).
15. D. A. Reznick, H. Bryga, J. A. Endler, Experimentally induced life-history evolution in a natural population. *Nature* **346**, 357–359 (1990).
16. P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711 (2002).
17. M. A. Bell, W. E. Aguirre, N. J. Buck, Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution* **58**, 814–824 (2004).

18. D. Schluter, K. B. Marchinko, M. E. Arnegard, H. Zhang, S. D. Brady, F. C. Jones, M. A. Bell, D. M. Kingsley, Fitness maps to a large-effect locus in introduced stickleback populations. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e1914889118 (2021).
19. A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, D. A. Petrov, Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLOS Genet.* **10**, e1004775 (2014).
20. S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, G. Sherlock, Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).
21. P. Nosil, R. Villoutreix, C. F. de Carvalho, T. E. Farkas, V. Soria-Carrasco, J. L. Feder, B. J. Crespi, Z. Gompert, Natural selection and the predictability of evolution in *Timema* stick insects. *Science* **359**, 765–770 (2018).
22. J. T. Anderson, C.-R. Lee, C. A. Rushworth, R. I. Colautti, T. Mitchell-Olds, Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol. Ecol.* **22**, 699–708 (2013).
23. N. O. Therkildsen, A. P. Wilder, D. O. Conover, S. B. Munch, H. Baumann, S. R. Palumbi, Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science* **365**, 487–490 (2019).
24. R. D. H. Barrett, S. Laurent, R. Mallarino, S. P. Pfeifer, C. C. Y. Xu, M. Foll, K. Wakamatsu, J. S. Duke-Cohan, J. D. Jensen, H. E. Hoekstra, Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019).
25. T. J. Thurman, R. D. H. Barrett, The genetic consequences of selection in natural populations. *Mol. Ecol.* **25**, 1429–1448 (2016).
26. M. A. Bell, W. E. Aguirre, Contemporary evolution, allelic recycling, & adaptive radiation of the threespine stickleback. *Evol. Ecol. Res.* **15**, 377–411 (2013).
27. M. A. Bell, S.A. Foster, Ed., *The Evolutionary Biology of the Threespine Stickleback* (Oxford Univ. Press, 1994).
28. B. Fang, J. Merilä, M. Matschiner, P. Momigliano, Estimating uncertainty in divergence times among three-spined stickleback clades using the multispecies coalescent. *Mol. Phylogenet. Evol.* **142**, 106646 (2020).
29. D. Schluter, G. L. Conte, Genetics and ecological speciation. *Proc. Natl. Acad. Sci. U.S.A.* **106** (Suppl. 1), 9955–9962 (2009).
30. P. A. Hohenlohe, S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, W. A. Cresko, Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLOS Genet.* **6**, e1000862 (2010).
31. F. C. Jones, Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady, A. M. Southwick, D. M. Absher, R. M. Myers, T. E. Reimchen, B. E. Deagle, D. Schluter, D. M. Kingsley, A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.* **22**, 83–90 (2012).
32. B. E. Deagle, F. C. Jones, D. M. Absher, D. M. Kingsley, T. E. Reimchen, Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Mol. Ecol.* **22**, 1917–1932 (2013).
33. M. Roesti, S. Gavrilets, A. P. Hendry, W. Salzburger, D. Berner, The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* **23**, 3944–3956 (2014).
34. T. C. Nelson, W. A. Cresko, Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol. Lett.* **2**, 9–21 (2018).
35. B. Fang, P. Kemppainen, P. Momigliano, X. Feng, J. Merilä, On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat. Ecol. Evol.* **4**, 1105–1115 (2020).
36. E. A. Lescak, S. L. Bassham, J. Catchen, O. Gelmond, M. L. Sherbick, F. A. von Hippel, W. A. Cresko, Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E7204–E7212 (2015).
37. N. V. Terekhanova, M. D. Logacheva, A. A. Penin, T. V. Neretina, A. E. Barmintseva, G. A. Bazykin, A. S. Kondrashov, N. S. Mugue, Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLOS Genet.* **10**, e1004696 (2014).
38. S. Bassham, J. Catchen, E. Lescak, F. A. von Hippel, W. A. Cresko, Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. *Genetics* **209**, 921–939 (2018).
39. W. E. Aguirre, M. A. Bell, Twenty years of body shape evolution in a threespine stickleback population adapting to a lake environment. *Biol. J. Linn. Soc. Lond.* **105**, 817–831 (2012).
40. A. Garcia-Elfring, A. Paccard, T. J. Thurman, B. A. Wasserman, E. P. Palkovacs, A. P. Hendry, R. D. H. Barrett, Using seasonal genomic changes to understand historical adaptation to new environments: Parallel selection on stickleback in highly-variable estuaries. *Mol. Ecol.* **30**, 2054–2064 (2021).
41. D. A. Marques, F. C. Jones, F. Di Palma, D. M. Kingsley, T. E. Reimchen, Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nat. Ecol. Evol.* **2**, 1128–1138 (2018).
42. R. D. H. Barrett, S. M. Rogers, D. Schluter, Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255–257 (2008).
43. T. G. Laurentino, D. Moser, M. Roesti, M. Ammann, A. Frey, F. Ronco, B. Kueng, D. Berner, Genomic release-recapture experiment in the wild reveals within-generation polygenic selection in stickleback fish. *Nat. Commun.* **11**, 1928 (2020).
44. N. Yu, F.-C. Chen, S. Ota, L. B. Jorde, P. Pamilo, L. Patthy, M. Ramsay, T. Jenkins, S.-K. Shyue, W.-H. Li, Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269–274 (2002).
45. D. Schluter, E. A. Clifford, M. Nemethy, J. S. McKinnon, Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**, 809–822 (2004).
46. J. A. Baker, D. C. Heins, J. E. Baum, Trajectory and rate of change in female life-history traits following colonization of a freshwater, lacustrine environment by oceanic threespine stickleback. *Evol. Ecol. Res.* **20**, 247–263 (2019).
47. M. A. Bell, D. C. Heins, M. A. Wund, F. A. von Hippel, R. Massengill, K. Dunker, G. A. Bristow, W. E. Aguirre, Reintroduction of threespine stickleback into Cheney and Scout Lakes, Alaska. *Evol. Ecol. Res.* **17**, 157–178 (2016).
48. K. Spitzer, M. Pelizzola, A. Futschik, Modifying the Chi-square and the CMH test for population genetic inference: Adapting to overdispersion. *Ann. Apl. Stats.* **14**, 202–220 (2020).
49. N. M. O'Brown, B. R. Summers, F. C. Jones, S. D. Brady, D. M. Kingsley, A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *eLife* **4**, e05290 (2015).
50. A. P. McGregor, V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, D. L. Stern, Morphological evolution through multiple *cis*-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).
51. S. Koshikawa, M. W. Giorgianni, K. Vaccaro, V. A. Kassner, J. H. Yoder, T. Werner, S. B. Carroll, Gain of *cis*-regulatory activities underlies novel domains of wingless gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7524–7529 (2015).
52. L. F. Stam, C. C. Laurie, Molecular dissection of a major gene effect on a quantitative trait: The level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* **144**, 1559–1564 (1996).
53. C. L. Peichel, D. A. Marques, The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20150486 (2017).
54. C. T. Miller, A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman, M. D. Shapiro, B. L. Cole, C. L. Peichel, D. Schluter, D. M. Kingsley, Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. *Genetics* **197**, 405–420 (2014).
55. A. H. Chan, P. A. Jenkins, Y. S. Song, Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1003090 (2012).
56. J. Galloway, W. A. Cresko, P. Ralph, A few stickleback suffice for the transport of alleles to new lakes. *G3* **10**, 505–514 (2020).
57. F. Han, S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, M. T. Webster, Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* **27**, 1004–1015 (2017).
58. J. I. Meier, D. A. Marques, C. E. Wagner, L. Excoffier, O. Seehausen, Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**, 1489–1506 (2018).
59. W. Mei, M. G. Stetter, D. J. Gates, M. C. Stitzer, J. Ross-Ibarra, Adaptation in plant genomes: Bigger is different. *Am. J. Bot.* **105**, 16–19 (2018).
60. R. D. Hernandez, J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, G. McVean; 1000 Genomes Project, G. Sella, M. Przeworski, Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
61. C. L. Peichel, K. S. Nereng, K. A. Ohgi, B. L. E. Cole, P. F. Colosimo, C. A. Buerkle, D. Schluter, D. M. Kingsley, The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901–905 (2001).
62. C. L. Peichel, S. T. Sullivan, I. Liachko, M. A. White, Improvement of the threespine stickleback genome using a Hi-C-based proximity-guided assembly. *J. Hered.* **108**, 693–700 (2017).
63. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
64. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
65. M. Lynch, D. Bost, S. Wilson, T. Maruki, S. Harrison, Population-genetic inference from pooled-sequencing data. *Genome Biol. Evol.* **6**, 1210–1218 (2014).
66. R. Kofler, R. V. Pandey, C. Schlötterer, PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435–3436 (2011).

67. A. F. Shanfelter, S. Archambeault, M. A. White, Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biol. Evol.* **11**, 1573–1585 (2019).
68. A. M. Glazer, E. E. Killingbeck, T. Mitros, D. S. Rokhsar, C. T. Miller, Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3* **5**, 1463–1472 (2015).
69. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting FST: The impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).
70. T. Taus, A. Futschik, C. Schlötterer, Quantifying selection with pool-seq time series data. *Mol. Biol. Evol.* **34**, 3023–3034 (2017).
71. W. Aguirre, P. Doherty, M. Bell, Genetics of lateral plate and Gillraker phenotypes in a rapidly evolving population of threespine stickleback. *Behaviour* **141**, 1465–1483 (2004).
72. A. Y. K. Albert, S. Sawaya, T. H. Vines, A. K. Knecht, C. T. Miller, B. R. Summers, S. Balabhadra, D. M. Kingsley, D. Schluter, The genetics of adaptive shape shift in stickleback: Pleiotropy and effect size. *Evolution* **62**, 76–85 (2008).
73. S. Arif, W. E. Aguirre, M. A. Bell, Evolutionary diversification of opercle shape in Cook Inlet threespine stickleback. *Biol. J. Linn. Soc. Lond.* **97**, 832–844 (2009).
74. W. E. Aguirre, K. E. Ellis, M. Kusenda, M. A. Bell, Phenotypic variation and sexual dimorphism in anadromous threespine stickleback: Implications for postglacial adaptive radiation. *Biol. J. Linn. Soc. Lond.* **95**, 465–478 (2008).
75. M. D. Shapiro, M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng, B. Jónsson, D. Schluter, D. M. Kingsley, Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
76. Y. F. Chan, M. E. Marks, F. C. Jones, G. Villarreal Jr., M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, D. M. Kingsley, Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).
77. A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, W. J. Kent, The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
78. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution's cauldron: Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11484–11489 (2003).
79. W. J. Kent, BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
80. B. L. Aken, S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. Fernandez Banet, K. Billis, C. García Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F. J. Martin, D. N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y. A. Tang, J.-H. Vogel, S. White, A. Zadissa, P. Flicek, S. M. J. Searle, The Ensembl gene annotation system. *Database*. 2016, baw093 (2016).
81. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC. Res. Notes* **9**, 88 (2016).
82. D. W. Hagen, Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). *J. Fish. Res. Board Can.* **24**, 1637–1692 (1967).
83. K. Yoshida, T. Makino, K. Yamaguchi, S. Shigenobu, M. Hasebe, M. Kawata, M. Kume, S. Mori, C. L. Peichel, A. Toyoda, A. Fujiyama, J. Kitano, Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLOS Genet.* **10**, e1004223 (2014).
84. M. Higuchi, A. Goto, Genetic evidence supporting the existence of two distinct species in the genus *Gasterosteus* around Japan. *Environ. Biol. Fishes* **47**, 1–16 (1996).
85. S. Varadharajan, P. Rastas, A. Löytynoja, M. Matschiner, F. C. F. Calboli, B. Guo, A. J. Nederbragt, K. S. Jakobsen, J. Merilä, A high-quality assembly of the nine-spined stickleback (*Pungitius pungitius*) genome. *Genome Biol. Evol.* **11**, 3291–3308 (2019).
86. M. Kirkpatrick, N. Barton, Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
87. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
88. B. S. Gaut, A. D. Long, The lowdown on linkage disequilibrium. *Plant Cell* **15**, 1502–1506 (2003).
89. R. C. Lewontin, The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67 (1964).
90. K. Osoegawa, K. C. Mallempati, S. Gangavarapu, A. Oki, K. Gendzekhadze, S. R. Marino, N. K. Brown, M. P. Bettinotti, E. T. Weimer, G. Montero-Martín, L. E. Creary, T. A. Vayntrub, C.-J. Chang, M. Askar, S. J. Mack, M. A. Fernández-Viña, HLA alleles and haplotypes observed in 263 US families. *Hum. Immunol.* **80**, 644–660 (2019).
91. P. A. Hohenlohe, S. Bassham, M. Currey, W. A. Cresko, Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 395–408 (2012).
92. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
93. S. Purcell, C. Chang, *PLINK*; www.cog-genomics.org/plink/1.9/.
94. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).
95. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
96. O. Delaneau, B. Howie, A. J. Cox, J.-F. Zagury, J. Marchini, Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013).
97. M. A. White, J. Kitano, C. L. Peichel, Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome. *Mol. Biol. Evol.* **32**, 1981–1995 (2015).
98. G. Lunter, M. Goodson, Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* **21**, 936–939 (2011).
99. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
100. M. A. Bell, J. D. Stewart, P. J. Park, The world's oldest fossil threespine stickleback fish. *Copeia* **2009**, 256–265 (2009).
101. B. Guo, F. J. J. Chain, E. Bornberg-Bauer, E. H. Leder, J. Merilä, Genomic divergence between nine- and three-spined sticklebacks. *BMC Genomics* **14**, 756 (2013).
102. R. Kawahara, M. Miya, K. Mabuchi, T. J. Near, M. Nishida, Stickleback phylogenies resolved: Evidence from mitochondrial genomes and 11 nuclear genes. *Mol. Phylogenet. Evol.* **50**, 401–404 (2009).
103. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
104. C. S. Smukowski Heil, C. Ellison, M. Dubin, M. Noor, Recombining without hotspots: A comprehensive evolutionary portrait of recombination in two closely related species of *Drosophila*. *Genome Biol. Evol.* **7**, 2829–2842 (2015).
105. R. R. Hudson, M. Slatkin, W. P. Maddison, Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
106. Á. Jónás, T. Taus, C. Kosiol, C. Schlötterer, A. Futschik, Estimating the effective population size from temporal allele frequency changes in experimental evolution. *Genetics* **204**, 723–735 (2016).
107. N. O. Rode, Y. Holtz, K. Loridon, S. Santoni, J. Ronfort, L. Gay, How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Mol. Ecol. Resour.* **18**, 194–203 (2018).
108. C. Vlachos, C. Burny, M. Pelizzola, R. Borges, A. Futschik, R. Kofler, C. Schlötterer, Benchmarking software tools for detecting and quantifying selection in evolve and resequencing studies. *Genome Biol.* **20**, 169 (2019).
109. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
110. J. H. Gillespie, *Population Genetics: A Concise Guide* (Johns Hopkins University Press, 1998).
111. Z. He, X. Dai, M. Beaumont, F. Yu, Detecting and quantifying natural selection at two linked loci from time series data of allele frequencies with forward-in-time simulations. *Genetics* **216**, 521–541 (2020).
112. M. Kimura, A model of a genetic system which leads to closer linkage by natural selection. *Evolution* **10**, 278–287 (1956).
113. J. Terhorst, C. Schlötterer, Y. S. Song, Multi-locus analysis of genomic time series data from experimental evolution. *PLOS Genet.* **11**, e1005069 (2015).
114. C. J. Battey, P. L. Ralph, A. D. Kern, Predicting geographic location from genetic variation with deep neural networks. *eLife* **9**, e54507 (2020).
115. A. D. Kern, D. R. Schrider, diploS/HIC: An updated approach to classifying selective sweeps. *G3* **8**, 1959–1970 (2018).
116. D. R. Schrider, A. D. Kern, Supervised machine learning for population genetics: A new paradigm. *Trends Genet.* **34**, 301–312 (2018).
117. B. C. Haller, P. W. Messer, SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
118. C. D. Huber, A. Durvasula, A. M. Hancock, K. E. Lohmueller, Gene expression drives the evolution of dominance. *Nat. Commun.* **9**, 2750 (2018).
119. A. F. Agrawal, M. C. Whitlock, Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* **187**, 553–566 (2011).
120. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (O'Reilly Media Inc., 2019).
121. T. A. O'Brien, K. Kashinath, N. R. Cavanaugh, W. D. Collins, J. P. O'Brien, A fast and objective multidimensional kernel density estimation method: fastKDE. *Comput. Stat. Data Anal.* **101**, 148–160 (2016).
122. T. A. O'Brien, W. D. Collins, S. A. Rauscher, T. D. Ringler, Reducing the computational cost of the ECF using a nuFFT: A fast and objective probability density estimation method. *Comput. Stat. Data Anal.* **79**, 222–234 (2014).

123. J. A. Chaves, E. A. Cooper, A. P. Hendry, J. Podos, L. F. De León, J. A. M. Raeymaekers, W. O. MacMillan, J. A. C. Uy, Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol. Ecol.* **25**, 5282–5295 (2016).
124. S. Singhal, E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean, M. Przeworski, Stable recombination hotspots in birds. *Science* **350**, 928–932 (2015).
125. S. Lamichhaney, F. Han, J. Berglund, C. Wang, M. S. Almén, M. T. Webster, B. R. Grant, P. R. Grant, L. Andersson, A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352**, 470–474 (2016).
126. S. Lamichhaney, J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, L. Andersson, Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).

Acknowledgments: We thank the many individuals who contributed to this study. Fish samples used for geographic surveys: S. Arnott, B. Blackman, F. Chan, P. Colosimo, A. Dalziel, B. Deagle, D. P. Højgaard, J. A. Jacobsen, B. Jonsson, R. King, D. Kueltz, A. Maccoll, J. McKinnon, C. Miller, S. Mori, K. O'Brien, C. Peichel, M. Ravinet, M. Rhodes-Reese and NOAA, T. Reimchen, J. Richmond, D. Schluter, M. Shapiro, B. Summers, and T. Vines. Undergraduate laboratory assistants: J. Ancona, H. Babalola, P. Chohan, J. F. Gaige, Z. Khan, and J. Mallozzi. Sampling access: M. Tauriainen and the staff of T & J Gravel Products permitted us to collect stickleback on their property from Scout and Loberg lakes, respectively. Field assistants, Alaska: S. Abrams, D. Arciari, D. B. Bell, M. R. Bell, S. R. Bell, B. Berland, M. Bobb, G. A. Bristow, K. T. Ellis, V. Ely, J. Fitzgerald, A. K. Gangavelli, M. A. Hahn, A. C. Havens, A. Hernandez, L. Hitt, J. Johnson, E. Kalabacas, A. Karve, H. Knoper, F. Kreier, M. Kurz, R. Lucas, A. McGarry, M. McGee, B. K. Lohman, R. Paitz, A. Plaunova, J. L. Rollins, H. Schultz, M. Sekiya, D. L. Soltz, L. Stein, A. C. Thompson, M. P. Travis Heide Viitaniemi, J. I. Wucherpfennig, and K. T. Xie. We thank A. Hinrichs, H. Clawson, K. Smith, and D. Karolchik for contribution to the UCSC Stickleback Genome Browser annotations for *gasAcu1*, which were lifted to *gasAcu1-4* for this study. IACUC approvals: K.R.V.: 1446584, Stony Brook University. M.A.B.: 237429, Stony Brook University. D.M.K: 13834, Stanford University. F.C.J.: 35/9185.82-5 EB01/09 A, Baden-Württemberg Regierungspräsidium, Germany. Friedrich Miescher Laboratory of Max Planck Society, Tübingen, Germany. D.C.H.: 0304R-UT-C, 0304R2, 0304R3, and 0304R4, Tulane University. **Funding:** This work was supported by NSF BSR8905758, BSR9046191, DEB0211391, DEB0322818, DEB0509070, and DEB0919184 to M.A.B.; NIH R01GM124330 to K.R.V.; DFG SPP1819 and ERC 617219 to F.C.J.; by Newcomb Institute grants to D.C.H.; NSF GRFP to G.A.R.K.; NSF GRFP 1656518 and NIH 5T32GM007790 and Stanford CEHG Fellowship to H.I.C.; and NIH 3P50HG002568 and 3P50HG002568-09S1 to D.M.K. D.M.K. is an investigator of HHMI.

Author contributions: M.A.B: experimental design, Alaskan sampling and population founding, tissue sampling, and morphological data collection. P.J.P.: Alaskan sampling and population founding. F.A.v.H.: Alaskan sampling and logistical support. W.E.A.: Alaskan sampling and morphological data collection. D.C.H.: Alaskan sampling. G.A.R.K.: analysis of geographic populations and genomic properties predictive of evolution. D.N.V., M.M., and T.S.B.: DNA extraction, quantitation, selection, and pooling of Alaskan samples. F.C.J.: design of geographic sampling and SNP array and SNP array analysis. S.D.B.: sample curation, DNA preparation, and SNP array calling. D.M.A. and R.M.M.: SNP array genotyping. M.K.: SNP array analysis. H.I.C.: transfer and visualization of genome annotations. F.D.P.: geographic population sequencing. D.M.K.: experimental design and conceptual guidance. K.R.V. and K.R.: analysis and modeling of contemporary populations. G.A.R.K. and K.R.V. wrote the manuscript with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All WGS Illumina data of extant populations have been deposited in the Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) at accession PRJNA247503. All WGS Illumina data for contemporary pool-seq experiments have been deposited in the Sequence Read Archive at accession PRJNA671824. All WGS BGI data for RS 2009 genomes used to construct the recombination map have been deposited in the Sequence Read Archive at accession PRJNA671690. SNP genotyping array data and ancestral reference sequence have been deposited on Data Dryad (https://doi.org/10.5061/dryad.pvmcvdnjm). EcoPeak, TempoPeak, and RS recombination rate data can be visualized and downloaded [via the Table Browser (*71*) of the UCSC Genome Browser (http://genome.ucsc.edu/) (*72*)] by copying the following track hub (*73*) URL into the "My Hubs" tab at https://genome.ucsc.edu/cgi-bin/hgHubConnect: https://sbwdev.stanford.edu/kingsleyAssemblyHub/hub.txt. The assembly hub must be opened through the UCSC Genome Browser, not directly. We provide the *gasAcu1-4* reference genome as well as liftOver chains for converting to and from the original Broad S1 stickleback reference genome (*gasAcu1*) via Data Dryad at https://doi.org/10.5061/dryad.547d7wm6t.

Submitted 11 January 2021
Accepted 5 May 2021
Published 18 June 2021
10.1126/sciadv.abg5285

Citation: G. A. Roberts Kingman, D. N. Vyas, F. C. Jones, S. D. Brady, H. I. Chen, K. Reid, M. Milhaven, T. S. Bertino, W. E. Aguirre, D. C. Heins, F. A. von Hippel, P. J. Park, M. Kirch, D. M. Absher, R. M. Myers, F. Di Palma, M. A. Bell, D. M. Kingsley, K. R. Veeramah, Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv.* **7**, eabg5285 (2021).