PNAS RESEARCH ARTICLE | EVOLUTION

Microevolutionary change in wild stickleback: Using integrative time-series data to infer responses to selection

Kasha Strickland[a,b,1], Blake Matthews[c], Zophonías O. Jónsson[d], Bjarni K. Kristjánsson[b], Joseph S. Phillips[b,e], Árni Einarsson[d], and Katja Räsänen[f,g]

Affiliations are included on p. 9.

Edited by Nils Stenseth, Universitetet i Oslo, Oslo, Norway; received May 23, 2024; accepted August 7, 2024

A central goal in evolutionary biology is to understand how different evolutionary processes cause trait change in wild populations. However, quantifying evolutionary change in the wild requires linking trait change to shifts in allele frequencies at causal loci. Nevertheless, datasets that allow for such tests are extremely rare and existing theoretical approaches poorly account for the evolutionary dynamics that likely occur in ecological settings. Using a decade-long integrative phenome-to-genome time-series dataset on wild threespine stickleback (*Gasterosteus aculeatus*), we identified how different modes of selection (directional, episodic, and balancing) drive microevolutionary change in correlated traits over time. Most strikingly, we show that feeding traits changed by as much 25% across 10 generations which was driven by changes in the genetic architecture (i.e., in both genomic breeding values and allele frequencies at genetic loci for feeding traits). Importantly, allele frequencies at genetic loci related to feeding traits changed at a rate greater than expected under drift, suggesting that the observed change was a result of directional selection. Allele frequency dynamics of loci related to swimming traits appeared to be under fluctuating selection evident in periodic population crashes in this system. Our results show that microevolutionary change in a wild population is characterized by different modes of selection acting simultaneously on different traits, which likely has important consequences for the evolution of correlated traits. Our study provides one of the most thorough descriptions to date of how microevolutionary processes result in trait change in a natural population.

Mývatn | natural selection | whole genome resequencing | quantitative genetics | population genomics

Significance

Traits of wild animals can change over contemporary timescales, but concluding that evolution played a role requires demonstrating that trait change is linked to genetic change. This is because while selection acts on organisms' traits, evolution in the strict sense is a process resulting in changes to the genome. But natural selection operating in natural ecosystems rarely acts in a single direction, and many factors that cause selection vary through time. We study wild stickleback in a well-studied lake to characterize how the genetics of correlated traits respond to different types of selection (e.g., directional or fluctuating). Our study clearly demonstrates how evolutionary processes cause trait change in the wild on a contemporary timescale.

Identifying microevolutionary processes underlying phenotypic change in wild populations remains a fundamental challenge for evolutionary biologists. Existing theoretical models typically have limited power to predict the observed dynamics of trait change in nature (1–3). For example, short-term change predicted by estimates of selection and trait heritability are often not realized in wild populations, giving rise to the "paradox of stasis" (4). There are several possible explanations for this. First, although wild populations live in variable environments and likely experience shifts among agents and modes of selection, most microevolutionary studies focus on a single mode of selection (most commonly directional) (5–8). Second, while selection acts on the entire phenotype, components of the multivariate phenotype can differ in their evolutionary potential, expected mode of selection (9, 10), and degree of plasticity (11), all of which can interactively shape evolutionary responses. Third, covariances among traits can both accelerate and constrain evolutionary responses and can complicate the detection of responses to selection (12). Collectively, the inherent difficulties in detecting microevolutionary change in wild populations, together with the complexity of the biological processes governing evolutionary responses, make it challenging to study evolution in the wild. Here, we use theory from quantitative genetics and molecular genomics together with an integrative phenome-genome time-series to connect trait change with natural selection in a wild population of threespine stickleback (*Gasterosteus aculeatus*).

Robustly determining whether phenotypic change is caused by microevolutionary processes in natural populations requires determining whether observed phenotypic trends are caused by allele frequency changes at causal loci (13). This can be extended to describe how different modes of selection act simultaneously in a single population by linking observed phenotypic trends to analogous patterns in changes to allele frequencies. Quantitative genetics approaches enable the use of individual-level data to characterize microevolution as change to a population's mean breeding value (i.e., the expected trait value for an individual given their genes) (13, 14). Population genomics methods, on the other hand, can track allele frequency dynamics across generations to identify loci that are diverging beyond neutral expectations (15–19). Integration of these methods helps to

Author contributions: K.S., B.M., Z.O.J., and K.R. designed research; K.S., B.K.K., J.S.P., and Á.E. performed research; K.S., Z.O.J., and B.K.K. contributed new reagents/analytic tools; K.S. analyzed data; and K.S., B.M., and K.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: kasha.strickland@ed.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2410324121/-/DCSupplemental.

Published September 4, 2024.

alleviate key limitations involved with using either in isolation (20, 21). Specifically, by integrating quantitative genetic and population genomic approaches in longitudinal data, one can avoid the decoupling of genome-phenotype linkages that is often associated with inferences from allele frequency dynamics alone, while retaining the ability to identify change at the molecular level and allow for alternative genetic architectures by relaxing the assumptions of the infinitesimal model. We apply this approach to a 10-y time-series (ca. 10 stickleback generations) of whole-genome sequencing and phenotypic measures of functional traits (trophic and defense traits) from the threespine stickleback of Lake Mývatn, NE Iceland, allowing us to assess the extent to which temporal change in multiple trait types reflects different modes of selection (directional, episodic, and balancing).

Mývatn is a highly dynamic ecosystem in which multiple ecological agents of selection, including vertebrate and invertebrate abundances, are known to fluctuate through time (22–25), likely generating strong natural selection. In particular, stickleback density fluctuates periodically [Fig. 1*C*, (24)] which is expected to cause fluctuating density-dependent selection, and abundance of stickleback predators and prey fluctuate through time (Fig. 1 *D* and *E*), generating predator- or prey-mediated natural selection. The stickleback population is panmictic (26) with high levels of standing genetic variation (H_e = 0.26 ± 0.02, F_{IS} = −0.032 ± 0.08), despite relatively low effective population size (N_e = 1,752 ± 249) and regular population bottlenecks (24). Similar to other freshwater populations of stickleback around the Atlantic (27), genomic principal component analyses (PCA) suggest that Mývatn stickleback are polymorphic for several inversion haplotypes (Fig. 1*B*) which are rich in quantitative trait loci (QTL) and have been associated with both marine-freshwater and lake-stream divergence (28). Together, these data suggest that fluctuating selection may contribute to the maintenance of genetic and phenotypic variation in this population (29–31), although fluctuating selection could also deplete genetic variation (32, 33). Mývatn stickleback were sampled every 2 y from 2010 to 2020 as part of a long-term study of lake Mývatn (22, 24) (approximately 10 stickleback generations) which included an extreme population crash in the years 2014 to 2016 (24) (Fig. 1*C*) that likely reflects a strong episode of selection. We phenotyped 861 individuals and sequenced the genomes for 515 (*SI Appendix*, Table S1). After quality control of raw sequence data, genotyping, and filtering, we had just over 1.7 million biallelic single nucleotide polymorphisms (SNPs) located on autosomes that were used for all downstream analyses. We tested for evidence of directional, fluctuating, and balancing selection at both the phenotypic and genomic levels, aiming to determine whether different modes of selection act simultaneously in this highly dynamic ecosystem. While our tests for fluctuating selection follow a single episode of nonlinear selection that very likely scales to result in fluctuating selection given evidence for strong ecological

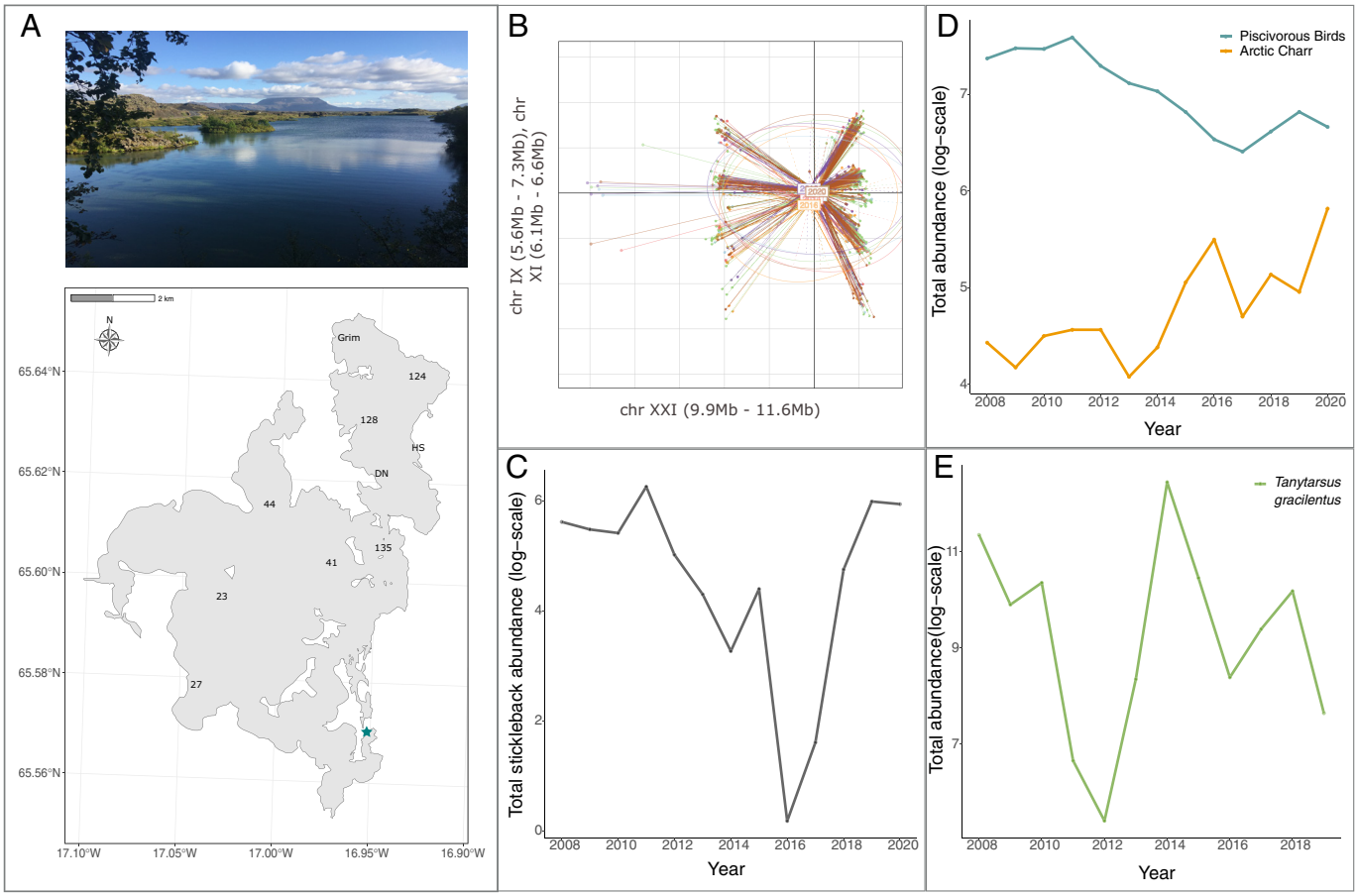

Fig. 1. Key aspects of Mývatn stickleback system. (*A*) photo and map of lake Mývatn with sampling locations named (location from which photo taken indicated on map as blue star) (*B*) summary figure of the first two axes from genome-wide clustering using genomic PCA analyses where axis labels describe genomic location of SNPs that segregate across each axis. Each of the genomic locations described on axis labels are a known inversion polymorphism in stickleback; (*C*) stickleback population dynamics across the duration of the study plotted as catch-per-unit-effort (CPUE) on the log-scale (see ref. 24 for further details); (*D*) population dynamics for key stickleback predators: sum total of piscivorous birds present at Mývatn [blue; Red-breasted merganser (*Mergus serrator*), red-throated diver (*Gavia stellata*), goosander (*Mergus merganser*), great northern diver (*Gavia immer*) and Slavonian grebe (*Podiceps auritus*)], and CPUE of Arctic charr (orange; *Salvelinus alpinus,* for further details, see ref. 25) both plotted on the log-scale; (*E*) population dynamics for *Tanytarsus gracilentus* (green) reflecting variation in chironomid midge abundance (key stickleback prey) plotted as CPUE on the log-scale.

fluctuations in this system (23, 24), due to our study spanning a single episode, we hereafter refer instead to an "episode of selection" to better reflect the nature of the selection we studied.

Results

Genetic Contribution to Observed Phenotypic Change. We focused on functionally important traits that fall into two general categories: the *defense traits* number of armor plates, thought to defend against bird predation, and length of pelvic and dorsal spines, thought to defend against gape-limited predators (34); the *trophic traits* gill-raker length, gill-raker gap width, and number, which in stickleback typically vary in relation to invertebrate prey communities, in particular, chironomid midges and cladocerans, and gut length which is correlated with the digestibility of diet (35). We also measured total *length* as a standard measure of body size and is also an approximate measure of age given that stickleback have indeterminate growth (36, 37). To identify whether traits changed through time, we ran a suite of mixed effects models to identify temporal trends in the phenotypic mean of the population after accounting for sexual dimorphism, length (and therefore age), and spatial divergence (*Materials and Methods*). We did not test whether length changed over time because while length is an important life history trait often under natural and sexual selection, we were not able to age individuals and were therefore not able to distinguish changes in length that were independent of age. The temporal trajectory of each trait was identified by comparing a model without year included to one which fit the year of capture as a linear term (as expected under directional selection) and one which further included a quadratic term (as expected under episodic selection) (38). Phenotypic covariances between all traits (after correcting for sexual dimorphism, age, and allometry) were estimated from the residual covariance matrix from a multivariate model. Note that because all analyses included length and sex, all results presented hereafter refer to traits relative to these measures.

Including a year term in the models improved model fit for all traits except pelvic spine (PS) length, and the means of all but one trait (PS length) changed over the 10-y time span (Table 1). The defense phenotype shifted toward fewer plates and longer dorsal spines over time (Fig. 2*B* and Table 1), suggesting adaptation as a response to predator-induced selection (34, 39). The trophic phenotype shifted toward fewer and longer gill-rakers with narrower gaps between them, and relatively longer guts (Fig. 2*B* and Table 1), suggesting adaptation to changes in diet. Most of the phenotypic trends were linear and directional, but there was nonlinearity in the effect of year on lengths of dorsal spines and the gut (Table 1). For traits that changed linearly, our analyses suggested that over the 10-y of the study, the number of plates decreased by 0.28 (mean = 5 plates) equating to a 6% decrease, the length of gill-rakers increased by 0.09 mm equating to a 8% increase, the number of gill-rakers decreased by 3.44 (mean = 14 rakers) equating to a 25% decrease, and gill raker gap width decreased by 0.03 mm (mean = 0.18 mm) equating to a 17% decrease across the duration of the study (all estimated for a female stickleback of average length, Table 1). Trait covariances generally reflected the directionality of overall phenotypic change (Fig. 2*A*): the lengths of the spines and of both gill-rakers were positively correlated with each other, as were the lengths of the dorsal spine and gut; gut length and gill-raker gap width were both negatively correlated with the number of gill-rakers; gill-raker gap width was positively correlated with the length of gill rakers.

Phenotypic change through time can result from phenotypic plasticity and/or changes to the genomic component of a trait (13). To identify whether observed trait change was caused by changes to the genetic architecture, which would suggest that trait change is caused by microevolutionary processes (13), we first estimated genomic estimated breeding values (GEBV) for each trait from

Table 1. Summary of results from linear models estimating phenotypic change through time. Summary data for raw data include overall mean ("Trait mean"), SD, variance among annual means ("Variance annual means"), and average SE of annual means ("Average SE annual means")

	Summary data				*WAIC*			*β*		
Trait (units)	Trait mean	SD	Variance annual means	Average SE annual means	No year	Year	Year + Year^2	Year	Year^2	r^2
GRL2 (mm)	1.09	0.26	0.02	0.0004	1066.1	**1047**	1046.1	**0.03 (0.02 to 0.05)**	-	0.66 (0.63 to 0.68)
GRL3 (mm)	1.12	0.26	0.02	0.0005	1054.5	**1024.3**	1024.6	**0.04 (0.03 to 0.05)**	-	0.68 (0.65 to 0.70)
GRN (n)	14.12	2.87	4.38	0.03	1580.4	**1474.4**	1472.5	**−0.12 (−0.14 to −0.09)**	-	0.25 (0.19 to 0.30)
GRW (mm)	0.18	0.05	0.001	0.00002	1309.7	**1252.2**	1252.3	**−0.06 (−0.08 to −0.05)**	-	0.52 (0.48 to 0.56)
Gut length (mm)	28.39	10.13	27.09	0.79	1122.9	1014.9	**998**	**−13.77 (−22.65 to −4.85)**	**0.003 (0.002 to 0.005)**	0.73 (0.70 to 0.74)
Plates (n)	4.97	0.92	0.02	0.01	1918.4	**1912**	1912.1	**−0.03 (−0.05 to -0.01)**	**-**	0.07 (0.04 to 0.10)
PS (mm)	5.12	1.06	0.24	0.007	**1336.1**	1337.6	1337.9	0.01 (−0.01 to 0.02)	-	0.59 (0.56 to 0.62)
SP1 (mm)	3.16	0.66	0.12	0.003	1264.2	1237.9	**1223.2**	**−13.22 (−22.16 to −4.18)**	**0.003 (0.001 to 0.005)**	0.65 (0.62 to 0.67)
SP2 (mm)	3.38	0.7	0.13	0.003	1304.8	1287.4	**1274.2**	**−12.51 (−21.57 to −3.52)**	**0.003 (0.001 to 0.005)**	0.63 (0.59 to 0.65)

WAIC for models with no effect of year ("No year"), a linear effect of time ("Year") or a linear and quadratic effect of time ("Year^2") are shown, with the optimal model (based on Δ*WAIC*) shown in bold. Regression coefficients (β) shown for the best fitting model (i.e., quadratic term shown only when model supports inclusion) with statistical support (i.e., posterior distribution does not overlap with zero) in bold. In cases where Δ*WAIC* between linear and quadratic model was within two (suggesting little difference to model fit) we assessed the statistical support for the quadratic term, and if not different from zero, we present results from the linear model. All β coefficients shown are posterior means with 95% CIs of posterior in parentheses. r^2 estimated for the best fitting model.

Traits included were GRL2 – length of 2nd gill-raker, GRL3 – length of 3rd gill-raker, GRN – gill-raker number, GRW – gill-raker gap width, gut length, Plates – number of armor plates, PS – pelvic spine length, SP1 – length of 1st dorsal spine, SP2 – length of 2nd dorsal spine. All traits were standardized to have mean of zero and SD of one before being fit in models.

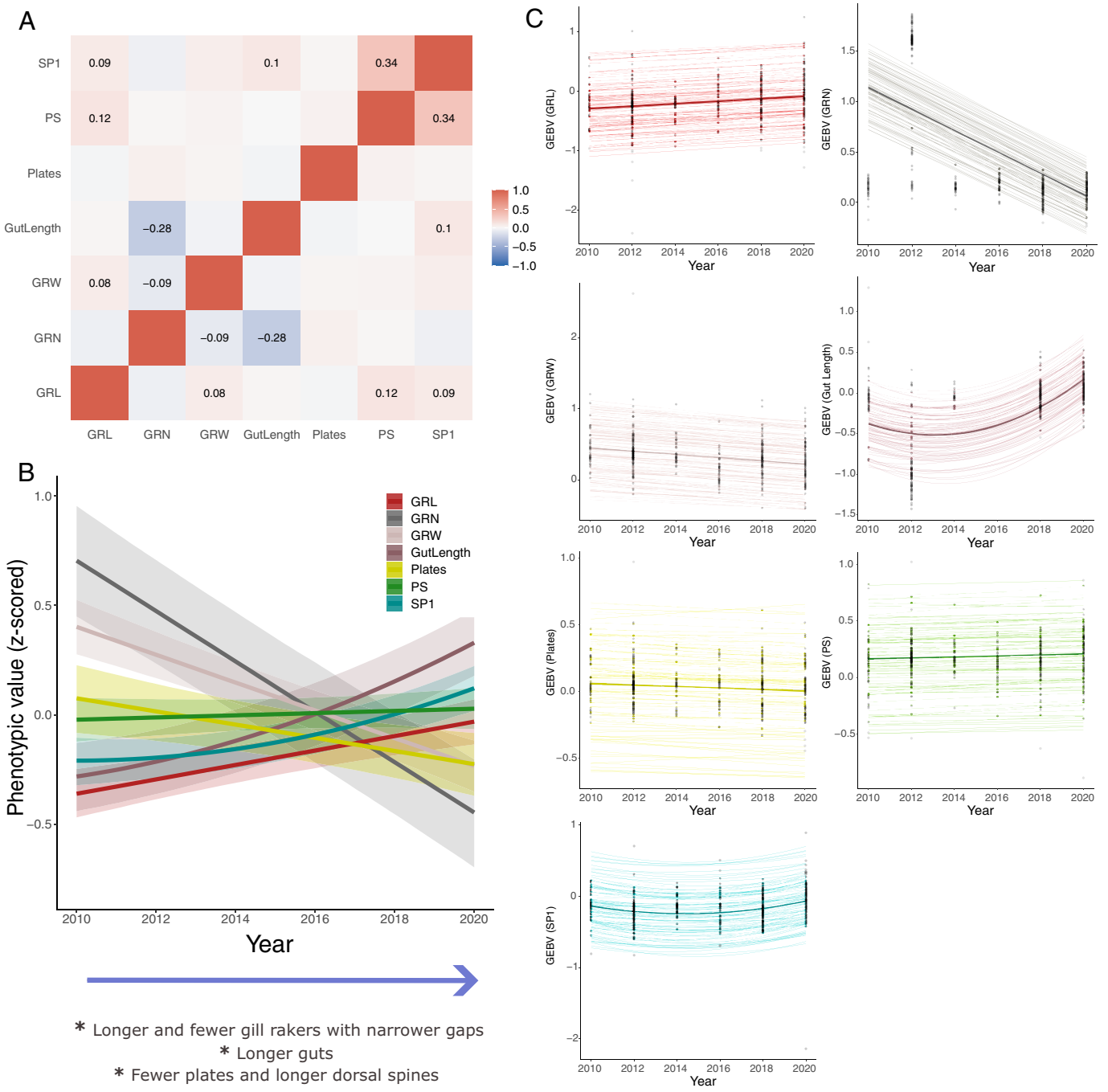

Fig. 2. Patterns of phenotypic variation. (*A*) Heatmap of phenotypic correlations accounting for length and sex. Values shown when 95% CIs of the posterior distribution of the correlation estimate was different from zero; (*B*) Phenotypic change through time (after accounting for length, sex, and space) where all traits are standardized to have zero mean and SD of one. Regression lines reflect the posterior means of the predicted temporal change in traits derived from univariate mixed effects models, and shaded ribbons show the 95% CIs of the posterior distribution of the predicted change in the trait. Regression lines were nonzero except for PS (Table 1); (*C*) Temporal trends in genomic estimated breeding values (GEBV) for all traits. Solid lines indicate the posterior mean of the predicted change in GEBVs for each trait with yearly average plotted as black points and the SD around that average plotted as error bars. Each thin line is estimated from a single draw from the posterior distribution of the model, which together generate 95% CI around the expected trend. Traits plotted are GRL – length of 2nd gill-raker, GRN – gill-raker number, GRW – gill-raker gap width, gut length, Plates – number of armor plates, PS – pelvic spine length, SP1 – length of 1st dorsal spine.

genome-wide SNPs. We used models implemented in the *hibayes* R package (40), correcting for total length (and therefore age), sex, year, and site of capture in each model (*Materials and Methods*). All traits had nonzero additive genetic variance (*SI Appendix*, Table S5) and were moderately to highly heritable (h^2 = 0.18 to 0.62, Table 2). Next, we ran linear regressions to model GEBVs as a function of year, which was fit as either a linear or quadratic term (*Materials and Methods*). GEBVs for the number of armor plates, and the lengths of dorsal and pelvic spines did not change through time (Fig. 2*C* and Table 2), suggesting that observed phenotypic trends in these traits may have occurred as a result of phenotypic plasticity. In contrast, GEBVs for number of gill-rakers, length of gill-rakers, gill-raker gap width, and gut length all changed between 2010 and 2020 (Fig. 2*C* and Table 2). Moreover, the shape and direction of the effect of year on GEBVs for these traits mirrored the observed phenotypic trends (Fig. 2), suggesting that phenotypic trends in these trophic traits were caused by microevolutionary processes causing changes to the genetic architecture.

Table 2. Summary of results from linear models on changes in GEBVs through time. h^2 shows narrow-sense heritability for all traits based on analyses across all years

	β		
Trait	Year	Year^2	h^2
GRL2 (mm)	**0.02 (0.01 to 0.03)**	-	0.62 (0.50 to 0.76)
GRL3 (mm)	**0.02 (0.01 to 0.03)**	-	0.49 (0.35 to 0.70)
GRN (n)	**−0.11 (−0.12 to −0.10)**	-	0.57 (0.48 to 0.66)
GRW (mm)	**−0.02 (−0.03 to −0.01)**	-	0.40 (0.30 to 0.49)
Gut length (mm)	**−58.34 (−78.58 to −43.14)**	**0.01 (0.01 to 0.02)**	0.60 (0.50 to 0.73)
Plates (n)	−0.005 (−0.01 to 0.004)	-	0.18 (0.14 to 0.24)
PS (mm)	0.004 (−0.004 to 0.01)	-	0.34 (0.24 to 0.45)
SP1 (mm)	−20.64 (−50.87 to −7.51)	0.005 (−0.0009 to 0.009)	0.37 (0.26 to 0.49)
SP2 (mm)	−26.71 (−60.36 to −10.96)	0.007 (−0.0003 to 0.01)	0.46 (0.32 to 0.69)

The regression coefficients (β) effects of year on GEBVs are shown, and the quadratic effect of year ("Year^2") was only fit where there was statistical support for the effect in phenotypic analyses (Table 1). All coefficients shown are posterior means with 95% CIs of posterior in parentheses. Regression coefficients where 95% CI of posterior does not overlap with zero shown in bold.
Traits included were GRL2 – length of 2nd gill-raker, GRL3 – length of 3rd gill-raker, GRN – gill-raker number, GRW – gill-raker gap width, gut length, Plates – number of armor plates, PS – pelvic spine length, SP1 – length of 1st dorsal spine, SP2 – length of 2nd dorsal spine. All traits were standardized to have mean of zero and SD of one before analysis.

Allele Frequency Dynamics Identify the Modality of Selection. While tracking temporal change in GEBVs can inform whether observed trait change has occurred as a result of changes to the underlying genetic architecture (13), it does not provide insight into changes on the genome nor directly infer whether that change has occurred as a result of selection (21). Molecular population genomics can therefore complement quantitative genetics by identifying genomic regions that are changing as a response to selection and, if QTL mapping has been achieved previously [as is the case for threespine stickleback (28, 41)], provide important clues as to which unmeasured phenotypic traits may be under selection. To identify genomic regions that were responding to selection, we compared allele frequency trajectories of all SNP loci to those expected under a Wright–Fisher model of neutrality (42, 43). Specifically, this model was used to predict allele frequency change caused by drift (or subsampling of the population) and was generated by randomly mating individuals at time t^0 to generate a population of size observed at time t^1 (repeated 10 times assuming a generation time of 1 y to replicate the full time-series). At each sampling point (n = 6), the simulated population was subsampled to match our observed sample sizes. This model was repeated 100 times to generate a distribution of expected allele frequency trajectories (*Materials and Methods*). A locus was then characterized as under: *directional selection* when its allele frequency trajectory changed linearly from 2010 to 2020 and deviated from that expected under the Wright–Fisher model; *episodic selection* if its allele frequency diverged from an expected trajectory during the crash years (2014 to 2016), but returned to an expected trajectory after this period, or; *balancing selection* if its allele frequency changed less than expected (Fig. 3*A*).

We identified 104 SNPs under directional selection (Fig. 3 and *SI Appendix*, Table S7). Of these, 81% (N = 83) fell on a known QTL in stickleback (41): 71% (N = 74) on a QTL for a feeding trait, 53% (N = 55) on a QTL for a defense trait, 25% (N = 26) on a QTL for a locomotion-related trait (e.g., vertebrae number, pterygiophore, and fin rays), 22% (N = 23) on a QTL for a respiration trait (e.g., operculum morphology) and <%1 (N = 1) on a QTL for pigmentation. Many regions of the stickleback genome have been linked to multiple phenotypes (41, 44, 45), and so we next identified whether there were trait categories that were over-represented in the regions of the genome identified as being under directional selection. We used Fisher's exact tests to identify whether the probability that a QTL identified as being under directional selection (i.e., on which a SNP under selection fell) was more likely to be associated with different functional trait categories than expected relative to the proportion of previously mapped QTL associated with that trait (41) (*Materials and Methods*). We found that QTL for feeding traits were significantly overrepresented in genomic regions under directional selection (odds ratio = 1.22, $P = 0.04$). The remaining trait categories were not significantly overrepresented in genomic regions under directional selection (*SI Appendix*, Table S8). The 104 SNPs under directional selection fell on 37 genes (*SI Appendix*, Table S9). Gene ontology (GO) analyses identified five enriched terms (*SI Appendix*, Table S10), all of which were associated with development of morphology, including one term associated with digestive tract development and one with development of head morphology.

We identified 818 SNPs to be under episodic selection (Fig. 3*C* and *SI Appendix*, Fig. S5 and Table S7). 701 of these SNPs (86.2%) fell on a known QTL in stickleback (41). 81% (N = 663) on a QTL for a feeding trait, 53% (N = 436) on a QTL for a defense trait, 30% (N = 242) on a QTL for a locomotion-related trait, 22% (N = 180) on a QTL for a respiration trait and 4% (N = 29) on a QTL for pigmentation. None of these trait types were significantly overrepresented when compared to the overall proportion of QTLs on the stickleback genome associated with each trait type, suggesting that observed associations with QTL could have occurred due to chance (*SI Appendix*, Table S8). The 818 SNPs under episodic selection fell on 280 genes (*SI Appendix*, Table S9), and GO analyses identified 14 enriched terms (*SI Appendix*, Table S10), seven of which were associated with anatomical structural development and two with processes involved in locomotory behavior. No SNPs were found to be under balancing selection.

Intriguingly, loci that were under either directional or episodic selection were located on QTL-rich regions of the genome. Feeding traits were highly overrepresented in genomic regions under directional selection, which aligns with analyses of GEBVs which suggested the genomic component of gill-raker number, length, and gap width changed linearly over time. Furthermore, one of the top GO terms associated with loci under directional selection was associated with digestive tract development, corroborating that dietary traits were under directional selection in this population during this 10-y period. The phenotypes implicated as under episodic selection may be linked to swimming physiology and behavior given that loci under episodic selection mapped to swimming-related traits, albeit at a rate expected by chance, and terms linked to locomotory behavior were overrepresented in GO analysis. This is a clear demonstration that while we were able to generate a large dataset consisting of individual-level measurements for many functionally relevant components of the stickleback phenotype, it remains unrealistic to be able to measure all components of wild organisms' phenotype. That is, it is likely that other, unmeasured aspects of the multivariate phenotype may have been responding to selection over the course of the study. This is where the integration with population genomics can be helpful as our GO analyses suggested that by far the most overrepresented gene functions under both directional and episodic selection were associated with

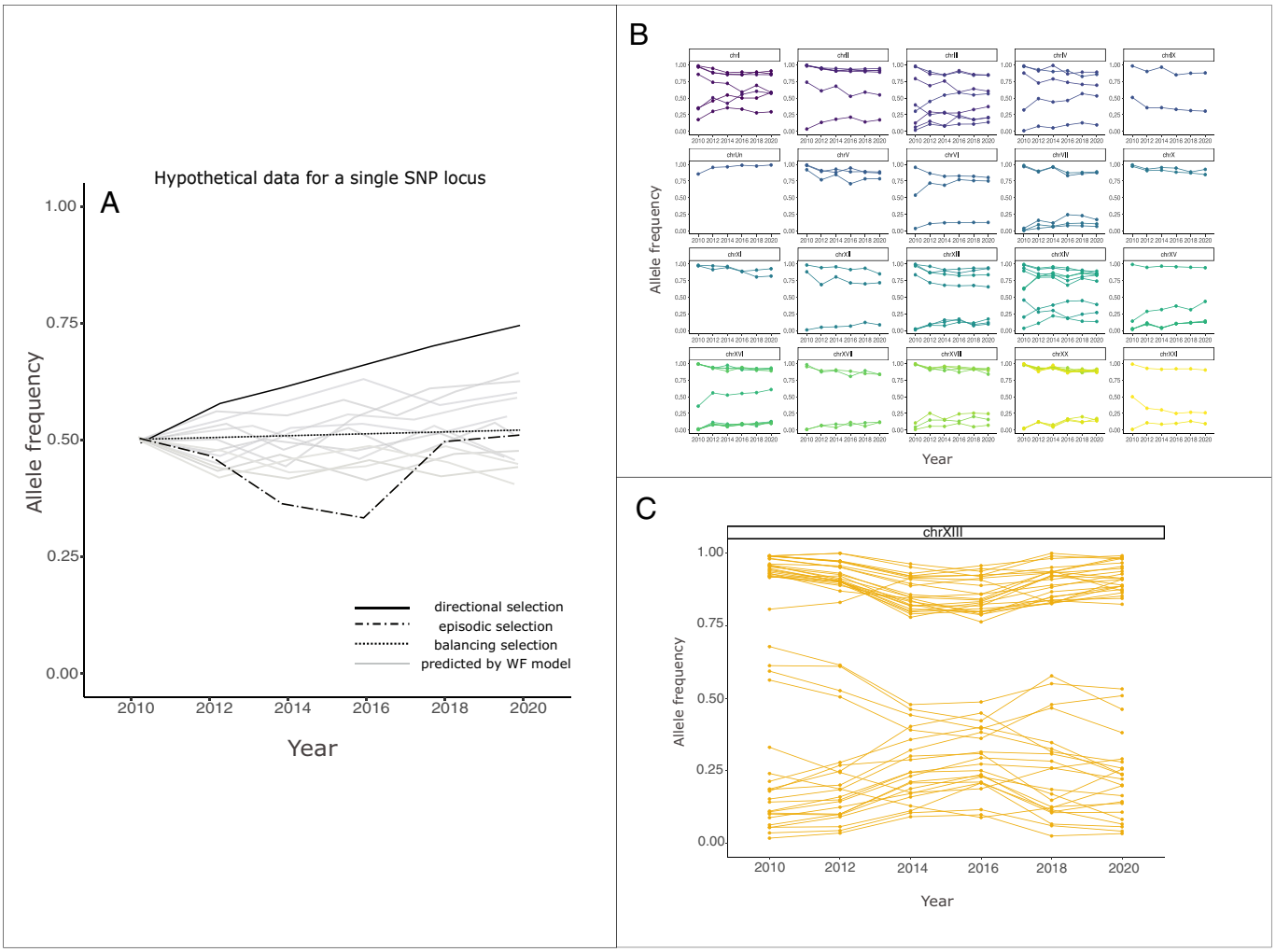

Fig. 3. Patterns of allele frequency change. (*A*) A schematic demonstrating how SNPs were identified as being under different modes of selection. For a single SNP, each solid gray line is a predicted allele frequency trajectory from a single run of a Wright-Fisher (WF) model, solid black line is a (hypothetical) observed allele frequency trajectory for a locus under directional selection, dotted line for a SNP under balancing selection, and dot-dashed line under episodic selection. Allele frequency trajectories for SNPs identified as (*B*) under directional selection (split across the 20 autosomes) and (*C*) under episodic selection in 2014 to 2016 (loci on chr XIII are shown for illustrative purposes; see *SI Appendix*, Fig. S5 for all loci).

structural development, indicating that other aspects of morphological phenotypes or their development are under selection. Taken together, these results clearly demonstrate how different modes of selection can act simultaneously on different components of the phenotype to shape trait evolution in a wild population.

Discussion

Integrating analytical approaches from quantitative genetics and molecular genomics with a decade of temporal sampling in a dynamic natural environment provided rich detail on patterns of microevolutionary change occurring in a wild population. The concurrent use of approaches proved especially powerful in revealing that dietary traits were under directional selection, as we were able to link changes in genomic breeding values for trophic traits with specific loci under directional selection that fell on QTL for trophic traits. Importantly, this result provides a unique empirical example which tests the assumptions used in each field: Although genomic estimated breeding values should sum the additive effects of causal loci affecting traits (46), and are derived directly from allelic variation (47), datasets that allow for an empirical examination of how predictions derived from breeding values correspond to allele frequency dynamics are extremely rare. Directional selection on trophic traits may have been driven by changes in available prey types over the studied time period (23, 48). However, ecological dynamics which generate the selective pressures acting on the phenotype are rarely linear (49), and fluctuating selection is thought to be a dominant mode of selection in natural populations (50). In concordance, we show that genes responding to an episode of selection during the stickleback population crash years were enriched for locomotory behavior, although this was not reflected in QTL-overlap analysis. Such selection on locomotion could have been caused by, for instance, negative density-dependent selection on dispersal (51), whereby in the low-density years, selection may have been favoring philopatry over dispersal due to reduced competition (52). However, linking ecological agents of selection to patterns of temporal change is challenging given that multiple ecological axes combine to generate the selective landscape organisms experience (53). This is further compounded by the role of phenotypic plasticity and habitat choice when organisms experience spatiotemporally variable environments (54), making genome- or phenotype-environment associations tricky to interpret.

Mývatn stickleback likely experience strong fluctuating density-dependent selection associated with the cyclic population dynamics of the population [approximately 6-y cycles (24)]. However, the episode of selection investigated here occurred over

a relatively short time-frame (approx. 10 generations) and it is hard to predict how these patterns scale to long-term change. Repeated episodes of selection may not always culminate into fluctuating selection per se because they may reflect long-term balancing selection, which would ultimately favor the maintenance of multiple alleles over time (55, 56). Investigating how short-term change shapes long-term patterns would require extensive sampling over multiple decades, and the rarity of empirical datasets such as the one generated in this study demonstrates the difficulty in achieving such a sampling design in natural populations. Indeed, while we were able to generate a large and powerful dataset that combined phenotype and genotypes at the individual level, this was achieved over relatively few sampling points, which clearly demonstrates the value of continued long-term studies of wild populations (57).

We have shown that correlated traits can differ in both the magnitude and shape of temporal change in response to selection, indicating that different modes of selection interact with different components of the phenotype within a single population. This could be important for understanding why changes that evolutionary models predict are often not observed (4). If we consider that selection acts on individual fitness, which is an emergent property of the effect of all the components of an organism's phenotype (58, 59), the simultaneity of different modes and directions of selection that act on correlated traits may explain a lack of an overall response. By working on a wild population of a model organism with exceptionally well-mapped trait architectures (28, 41), we were further able to gain insight into both measured and unmeasured traits under different modes of selection. By doing so, our study demonstrates how understanding the detailed evolutionary mechanisms affecting different parts of an individual's phenotype and genome can improve predictions about responses to selection in wild populations.

Materials and Methods

Study System and Sampling. Lake Mývatn is an environmentally heterogeneous ecosystem in North-East Iceland, and its ecological dynamics across multiple trophic levels, as well as patterns of spatial divergence, have been studied extensively (22, 24, 26, 60). For instance, chironomid midge, threespine stickleback, piscivorous birds, and Arctic charr population abundances vary strongly through time (23–25). The habitat of stickleback can be classified into five main types (60), and there is evidence for subtle phenotypic and genetic spatial divergence in Mývatn stickleback (26, 60, 61). Furthermore, there is spatial variation in water temperature (22), avian predators (26) as well as invertebrate prey communities (48). Stickleback have been surveyed since 1991 at eight lake sites across the two basins of the lake (North and South) as part of an ongoing long-term monitoring of population demographics (22, 24), with five shorelines sites added in 2009 (60) (Fig. 1). The sampling is done twice each year, in June and in August, by laying five unbaited minnow traps at predetermined locations (hereafter "sites") over two 12 h periods (night and day catch) (60). Stickleback from all traps are counted to estimate CPUE (24) and frozen for later analyses. Since 2009, a random subset of individuals (ca. N = 100 per site for each day and night catch) has been stored to allow phenotyping and/or genotyping.

Phenotyping. We randomly selected 20 individuals (of minimum total length 35 mm) from the June samples from 11 study sites (Fig. 1) every other year between 2010 and 2020 (N = 793, *SI Appendix*, Table S1). Where there were less than 20 individuals available (for instance in years where the stickleback population crashed), we used all the fish caught at that site in June of that year. Individuals were thawed, weighed on an electronic balance (wet mass, to the nearest mg) and their total length measured using a ruler (to the nearest mm). The right pectoral fin was cut and stored in 96% ethanol for DNA analyses. We measured traits that are known to be functionally relevant, typically under selection in stickleback and previously studied for spatial divergence (26): defense traits (armor plate number and length of spines) and trophic traits (gill raker morphology and gut length) (35, 62, 63).

On each individual, we measured the following 10 traits: total length, number of lateral armor plates (plate number, excluding the keel plates), length of the first dorsal spine (DS1), length of the second dorsal spine (DS2), length of the PS, length of the second gill raker on the first gill arch (GRL2), length of third gill raker on the first gill arch (GRL3), gap width between second and third gill rakers (GRW), number of long gill rakers on the first gill arch (GRN), and gut length. Note that we measured GRL2 and GRL3, rather than the length of the first gill raker (which is usually used in studies of stickleback trophic phenotype), because in some cases gill arches broke during dissection. After measurement of total length, each individual was dissected to remove the stomach and the gut, and any tapeworm (*Schistocephalus solidus*) parasites. Gut length was measured from the sphincter at the end of the esophagus to the nearest mm using a ruler. Unfortunately, the guts from fish caught in 2016 had been dissected prior to the commencement of the present study and they had not been measured at the time of dissection. As such, we did not have gut length measurements for fish caught in 2016. To aid morphological measurements, ethanol preserved fish were stained with alizarin red using standard protocols (26, 60). Fish were bleached using a 1:1 ratio of 3% H_2O_2 and 1% KOH and then stained in a solution of alizarin red and 1% KOH (64). After staining, digital images were taken of the left side of the fish with a digital camera (Canon EOS 600D), with mm paper for scale. From these images, plate number was counted and the length of DS1, DS2, and PS measured (in mm) to the nearest hundredth of a millimeter. After imaging, we dissected the first gill arch and, where necessary, restained it before mounting between two glass plates and photographing with a digital camera (Leica IC80 HD) mounted to a stereomicroscope (Leica M165-C), with mm paper for scale. We used the digital images of gill arches to measure GRL2, GRL3, and GRW (in mm) and counted GRN. All measurements from the digital images were taken using the segmented tool in ImageJ (65).

Whole Genome Resequencing and Bioinformatics. For genomic analyses, we randomly selected 10 of the 20 individuals that had been phenotyped from 10 of the 11 sites/y combination (N = 515, *SI Appendix*, Table S1). Where there were less than 10 individuals available, we used all fish from the June sampling. Genomic DNA was isolated and purified from the ethanol-stored fin clips using the Macherey-Nagel nucleomag tissue kit, following the manufacturer's protocol. Paired-end, PCR-free 150-bp insert libraries were then prepared for whole genome sequencing using the DNBSeq™ platform by BGI-Hong Kong to an average of 10× depth of coverage (26). All samples were mapped to v5 of the stickleback reference genome (66) and genotyped using the genome analysis tool kit (GATK) best practices pipeline (67). Only genotype calls with depth greater than six and less than 100 were retained, and autosomal SNPs with minor allele counts less than four were subsequently removed. The sex of individuals was confirmed using the proportion of reads with depth greater than eight mapped to the X vs. Y chromosome (68). For all analyses, we removed mitochondrial variants, indels, multiallelic variants, as well as variants identified on either of the sex chromosomes. SNPs with more than 50% genotype calls missing were removed, resulting in a total of 1,700,436 loci used for all downstream analyses.

Statistical Analyses.

Phenotypic trends. Our first aim was to characterize the covariance structure between pairs of all measured traits, as well as test for temporal change in the measured traits. To do this, we initially aimed to fit a single multivariate mixed model with all traits as a multivariate response as a function of sex, length, and year and including sampling site as a random effect. However, this full multivariate model had convergence issues, likely caused by the high level of complexity of the model for which current sample size was insufficient. Instead, we selected to run 1) a multivariate model to identify phenotypic covariances, and 2) a series of univariate models to test for temporal change in any of the measured traits. All models were fit with a Gaussian distribution, and traits were all standardized to have a mean of 0 and SD of 1 to ensure Gaussian errors were appropriate. All models were fit with sex and length as fixed effects to account for sexual dimorphism, allometry and age (because stickleback have indeterminate growth). All results presented therefore reflect trait measures relative to sex and length. While length is an important life history trait often under natural and sexual selection, given we were not able to age individuals, we were not able to distinguish between changes in length that were independent of individuals' age. Therefore, we selected to not analyze length as a quantitative trait. All models were fit using Stan via the brms package in *R* statistical environment (version 4.1.2) (69). All models were run with 6,000 iterations across four chains and a warm-up period

of 2,000 iterations, which was sufficient in all cases to achieve model convergence which was assessed by visually assessing mixing of chains and with $\widehat{R}$. Fixed effects were given normal priors with 0 mean and SD of 5. Random effects were given half-Cauchy priors with two degrees of freedom.

Pairwise phenotypic correlations between all traits were estimated using a multivariate mixed effects model. To estimate the full phenotypic covariance matrix, this model fitted all standardized traits as a multivariate response as a function of length and sex, and included year of capture as a random effect. Phenotypic correlations were then estimated from the residual covariance matrix.

To test whether any of the measured traits changed either linearly or nonlinearly across the study period, we compared three univariate mixed effects models for each trait. The first of these did not fit year as a fixed effect. The second fit year as a continuous linear effect to test whether the population mean of the trait changed linearly across time (as expected under directional selection). The third model was fit with an identical structure, except with an additional quadratic term for year in order to test whether the population mean of the trait changed quadratically following the population crash in the middle of the study period (as expected under episodic selection). In all models, intercepts were allowed to vary between sampling sites by fitting sampling site as a random effect. The three models were compared using *WAIC* to assess support either for a model with the linear term for year, or for a model with both the linear and quadratic term for year. In cases where Δ*WAIC* between the linear and quadratic models was ≤2, we then examined whether the quadratic term was different from zero. In cases where it was not, we present results from the linear model only. We present the 95% credible intervals of the posterior distribution for the effect of year on trait values and considered effect sizes to have statistical support when the credible intervals did not overlap zero. Note also that accounting for temporal autocorrelation in the residuals was not relevant here as individual fish were not repeatedly sampled through time. We modeled the mean trait change using linear and quadratic terms, and we acknowledge that the underlying evolutionary process might more closely resemble some autocorrelated stochastic process such as a random walk. However, it is difficult to statistically distinguish between these alternatives and the distinction also raises some conceptual issues about the deterministic vs. stochastic nature of the selection process that are beyond the scope of this paper.

Genetic contribution to phenotypic change through time. To test whether microevolutionary change was responsible for any observed trait change, we tested whether genomic estimated breeding values (GEBV) (13) for traits changed as a function of sampling year. To do this, GEBV for each trait were estimated for all individuals that were genotyped and phenotyped (N = 515, *SI Appendix*, Table S1). We used Bayesian regression models implemented in the *hibayes* R package (40, 70), which implements the "Bayes alphabet" models commonly used for genomic prediction (explained in detail in refs. 47 and 71–74). Breeding values are often estimated via analyses of pedigrees or genomic relatedness matrices under an infinitesimal model assuming a genomic architecture whereby all loci have an equally small effect on the trait (46). Although many traits are highly polygenic, estimating genomic breeding values directly from genomic variants using Bayes alphabet models allows for more appropriate mapping of genomic architectures, and should therefore improve our predictions of microevolutionary change (47). In addition, the prior for the distribution of marker effects can be flexibly implemented according to known, or predicted, genetic architectures of traits. We ran one model per trait with sex, length (and therefore also age), and year as fixed effects and sampling site as a multilevel random effect. Year was fit as a linear effect unless phenotypic analyses suggested there was a quadratic effect, in which case we fit year as both a linear and quadratic term. We ran all models using a *BayesBpi* model, which assumes that most markers have zero variance and a small proportion of markers (π) have nonzero variance, and the distribution of variances of these markers follows an inverse-chi-squared distribution. While it is possible to fit these models under a number of different assumed genetic architectures (70, 72–74), we selected the *BayesBpi* models because 1) this allows a flexible number of markers to be assigned as having nonzero variance, which is estimated within the model using variation in the empirical data, and 2) given genetic architectures are well mapped in stickleback, the prior distribution for marker variances used is a realistic assumed genetic architecture of our measured traits (41). Models were fit with 15,000 iterations, a burn-in set of 5,000 iterations, and a thinning interval of 10 (which was sufficient to achieve model convergence and an effective sample size of 1,000 for all traits) generating a posterior distribution for GEBVs for each trait. Convergence of models was assessed by visualizing the trace-plots of posterior distribution for all parameters estimated. SNP-based heritability (h^2) for each trait was as the proportion of total phenotypic variance (estimated as the sum of all variance components) attributed to additive genetic variance (V_A).

We then fit a linear model of GEBVs as a function of sampling year, and we repeated this for each draw of the posterior distribution to generate a posterior distribution of linear coefficient estimates for the relationship between GEBVs and sampling year. Year was fit as a linear effect effect unless phenotypic analyses suggested there was a quadratic effect, in which case we fit year as both a linear and quadratic term. Note that for the trait GRN the year 2012 appeared to be a rather extreme outlier as most individuals had more gill rakers than in all other years and there was greater phenotypic variance in this year than in other years. The linear effect of year was found to be the most appropriate model to fit temporal change in this trait over time (*Results*), but it may be that the quadratic term did not capture a true phenotypic trend that emerged in 2012 as it did not allow for the enough nonlinearity in the data. However, to ensure that this single year was not influencing our inferences regarding directional change we reran analyses of genomic breeding values without data from 2012. This further allowed us to ensure that anything that may have caused some stochastic change in this year was not influencing our conclusions. Removing 2012 did not change the overall trend we found when using the full dataset (*SI Appendix*, Fig. S4), and so we report results that include data from this year.

Allele frequency dynamics. To analyze changes in allele frequencies through time, we calculated allele frequencies per SNP per year (2010, 2012, 2014, 2016, 2018, 2020) for individuals in the North basin (24) (see *SI Appendix*, Table S1 for sample sizes). We investigated allele frequency dynamics using samples collected from the North basin only to avoid biases caused by variation in sample sizes between years in the South basin. Specifically, the majority of samples were collected in the North basin (*SI Appendix*, Table S1) and the sample sizes from the South basin were very variable. We do not believe this should affect downstream inferences because 1) stickleback density is much higher in the North and likely subsidizes the South basin via source-sink dynamics (24), and 2) although there is some evidence that allele frequencies and population densities vary between basins (24, 26), population genetic analyses suggest that the population is panmictic across the whole lake (26). To investigate allele frequency dynamics, we also subset the SNPs to retain those with a call rate of 50% within each year and a minor allele frequency in 2010 of at least 0.001 (*SI Appendix*, Fig. S2). This resulted in a total of 1,558,025 SNP loci used for this analysis.

We aimed to identify whether allele frequencies changed via directional selection, episodic selection, or balancing selection. The term episodic selection is used here rather than fluctuating selection because we tracked change over a single episode of selection (during a population crash), and detecting fluctuating selection would require tracking change across repeated episodes of selection. To isolate allele frequencies that changed due to selection as opposed to due to drift or sampling regimes, observed allele frequency trajectories of each SNP were compared to an allele frequency trajectory that would be expected under a Wright–Fisher model which predicted allele frequency change caused by drift (or other neutral processes). This model started with each SNP's observed allele frequency in 2010 and predicted the expected trajectory for each SNP accounting for changes in population size and variable sample sizes in the observed dataset. Specifically, this model started with a population of individuals of size $N_e\_t^0$ (observed effective population size, N_e, at time t^0, 2010, *SI Appendix*, Table S4) in which the allele frequency for a SNP at t^0 (2010) was the observed allele frequency in the empirical dataset. The population was then assumed to mate at random for two generations to generate a population of size Ne_ t^1 (observed effective population size *at* t^1 = 2012), assuming a generation time of 1 y. The population at t^1 was then sampled at N_t^1 (observed sample size at time t^1 = 2012) before an allele frequency for that time point was calculated. This process was then replicated for another four steps, resulting in a total of 10 generations and 6 sampling points (biannually between 2010 and 2020) to replicate the sampling process in the observed time-series dataset. The model was run for 100 iterations, generating a range of expected allele frequency trajectories, and the model was repeated for each SNP to generate an expected trajectory for each locus. Wright–Fisher models were run using custom R scripts which adapt the methods outlined in the *poolseq* R package (75).

The Wright–Fisher model we used accounted for differences in allele frequency because they started with the observed allele frequency for each locus at time t^0.

They also accounted for the variable populations sizes and sample sizes observed across the time-series by incorporating those observed parameters into the model (see above, *SI Appendix*, Tables S1 and S4). It should be noted, however, that it is likely that the strength of selection acting on loci of different starting frequencies is probably not even. That is, loci that start with a relatively low allele frequency and are identified as under selection are likely under stronger selection than one at intermediate frequency. This is because alleles that are at low frequency are much more likely to be lost due to drift over this time period that those that are at intermediate frequency. The Wright–Fisher model of drift is likely the most frequently used model of drift (42, 43, 76), especially in cases where a pedigree is not available [in which case alternative null models are available, e.g., gene-dropping (19)]. Nevertheless, the assumptions of the Wright–Fisher model (i.e., discrete generations, constant population size, no selection or mutation, and random mating) may be unlikely to occur in wild populations (77). In our analyses, we have dealt with most of these assumptions explicitly: our model incorporates variable population sizes observed through the time-series (see above), and our goal was to generate a null model in which allele frequency changes were not caused by selection (including sexual selection which would cause nonrandom mating). Furthermore, it is unlikely that mutation rates in this population are high enough to be exacting a significant influence on allele frequencies during our time-series (78). We acknowledge that the remaining assumption of discrete generations may not hold in this population, as individuals are thought to start breeding in their first year with some surviving and breeding in their second (or more rarely third) year. The consequence of this is that 1) the generation time is likely longer than we have modeled in our null model (i.e., somewhere between 1 and 2 y), and 2) observed allele frequencies calculated at each time point are calculated from a population that includes some individuals from previous cohorts. Although we recognize that such structure may cause issues in drawing inference from Wright–Fisher models in certain scenarios, in our study, given the presumed age-structure and generation time in the population, we believe that our implementation of the Wright–Fisher model represents quite a conservative approach as we are likely overestimating the rate of drift in the null model. This is because in our null model we include a faster generation time and population turnover (resulting in higher rates of neutral genetic change) than are probably occurring in the population. As such, while we may miss some signatures of selection in our approach (i.e., there may be false negatives in our results), we believe that the signatures that we do isolate are more likely to be accurate.

A given SNP was characterized as under *directional selection* if its allele frequency trajectory was linear and different from that expected under the Wright–Fisher model (i.e., does not fall within the expected AF distribution) in all sampling time points after t^0. Mývatn stickleback likely experience strong fluctuating density-dependent selection associated with the cyclic population dynamics of the population [approximately 6-y cycles (24)]. Our time series overlapped with one of these cycles, with a strong population crash in 2014 to 2016. As such, SNPs were characterized as responding to an *episode of selection* if allele frequency was different to the model during the years (2014, 2016) that the stickleback population crashed, but not at any other time point. That is, their allele frequency diverged when the population crashed but returned to a trajectory expected under the Wright–Fisher model afterward. Both of these approaches to determining whether a SNPs allele frequency trajectory deviates from a Wright–Fisher model represent a highly conservative method; for a SNP to be determined as being under either mode of selection, the probability (P) that its allele frequency trajectory occurred as a result of neutral processes must be as close to 0 as statistically possible (i.e., it does not fall anywhere in the range of expected trajectories from a Wright–Fisher model). As such, this approach avoids the need for extensive correcting for multiple testing as all SNPs identified as under selection would eventually have a P-value of = 1/100. SNPs were considered to be *under balancing selection* if their allele frequency changed less than expected under a Wright–Fisher model. To do this, we calculated the absolute difference in allele frequency between subsequent time points (ΔAF) and compared this to the distribution of allele frequency change in the Wright–Fisher model to generate a P-value for each SNP as the proportion of times that the expected ΔAF was less than the observed. All p-values were corrected for multiple testing by converting them to q-values using an FDR rate of 5%. SNPs were then considered as putatively under balancing selection if they were significantly different from expectations in each time point. This method for identifying balancing selection was designed to identify short-term balancing selection, and while it may be a relatively conservative approach, this definition has been used in previous research aiming to identify short-term balancing selection (79). Note that our definition of episodic selection here may scale over time to reflect a process of long-term balancing selection whereby relatively short-term episodes of selection repeated over time act as long-term balancing selection that maintains phenotypic and allelic variation in a population (55, 56).

For each set of SNPs identified to be putatively under a given type of selection, we identified the QTL these SNPs fell (exactly) on using a previously compiled list of all mapped QTL in threespine stickleback (41). To do this, genome locations for loci used in our analyses that were mapped to v5 of the reference genome were converted to genome positions on the Glazer genome assembly using LiftOver to facilitate overlap analyses with QTL locations referenced on other versions of the reference assembly (26). Then, using trait categories described in ref. 41 (i.e., feeding, defense, swimming/locomotion, pigmentation, and respiration), we used Fishers exact tests to compare the proportion of QTL putatively under selection (i.e., on which a SNP under selection fell) that were associated with the different trait categories to the proportion of all QTL previously mapped in stickleback that fall into that trait category.

Finally, we identified the protein-coding genes on which these SNPs were located (i.e., within the transcribed regions, hereafter "candidate genes") and ran GO analyses to explore whether any molecular functions were overrepresented in sets of genes associated with selection in our data. To do this, we compared candidate genes with the reference set of 20 805 genes across the stickleback genome ("gene universe"). GO information was obtained from the stickleback reference genome on ENSEMBL using the R package BIOMART (80), and functional enrichment was investigated using the package TOPGO 2.42 (81) and the Fisher's exact test (at $P < 0.01$). To reduce false positives, we pruned the GO hierarchy by requiring that each GO term had at least 10 annotated genes in our reference list ("nodeSize = 10").

Data, Materials, and Software Availability. Whole genome sequencing data generated and analysed in this project are available in the European nucleotide archive (ENA) accession numbers PRJEB79151 (82) and PRJEB58765 (83). Phenotypic data generated in this project and meta data associated with the genomic data are available at https://github.com/kashastrickland/stickleback-PhenoGenome (84).

ACKNOWLEDGMENTS. We would like to thank all students and volunteers that have helped with field work, processing samples, and phenotyping fish, especially Coralie Pallet, Pauline Gautier, Lucie Roques and Ragna Guðrún Snorradottir and Antoine Millet, who as a PhD student laid the foundations for field sampling in the early years of our study but sadly passed away in 2023. Sampling was conducted under the auspices of the Mývatn Research Station, which has government approval for collecting fish specimens from the lake. Data for Arctic charr abundances used in Fig. 1 came from previous publications, but we would like to thank Guðni Guðbergsson for collecting these data. Finally, we would like to thank Timotheé Bonnet and an anonymous reviewer for constructive comments on a previous version of this manuscript.

Author affiliations: [a]Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom; [b]Department of Aquaculture and Fish Biology, Háskólinn á Hólum, Hólum í Hjaltadal, Sauðárkrókur 551, Iceland; [c]Department of Fish Ecology and Evolution, Swiss Federal Institute of Aquatic Science and Technology, EAWAG, Kastanienbaum CH-6047, Switzerland; [d]Institute of Life and Environmental Sciences, School of Engineering and Natural Sciences, University of Iceland, Reykjavík 102, Iceland; [e]Department of Biology, Creighton University, Omaha, NE 68178; [f]Department of Aquatic Ecology, Swiss Federal Institute of Aquatic Science and Technology, EAWAG, Duebendorf 8600, Switzerland; and [g]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä 40014, Finland

1. T. Bonnet *et al.*, Genetic variance in fitness indicates rapid contemporary adaptive evolution in wild animals. *Science* **376**, 1012–1016 (2022).
2. J. Merilä, B. C. Sheldon, L. E. B. Kruuk, Explaining stasis: Microevolutionary studies in natural populations. *Genetica* **112**, 199–222 (2001).
3. R. J. A. Buggs, The challenge of demonstrating contemporary natural selection on polygenic quantitative traits in the wild. *Mol. Ecol.* **31**, 6383–6386 (2022).
4. B. Pujol *et al.*, The missing response to selection in the wild. *Trends Ecol. Evol.* **33**, 337–346 (2018).

5. M. Pfenninger, Q. Foucault, Population genomic time series data of a natural population suggests adaptive tracking of fluctuating environmental changes. *Integr. Comp. Biol.* **62**, 1812–1826 (2022).
6. M. Pfenninger, Q. Foucault, A.-M. Waldvogel, B. Feldmeyer, Selective effects of a short transient environmental fluctuation on a natural population. *Mol. Ecol.* **32**, 335–349 (2023).
7. S. M. Rudman *et al.*, Direct observation of adaptive tracking on ecological time scales in Drosophila. *Science* **375**, eabj7484 (2023).
8. P. de Villemereuil *et al.*, Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31969–31978 (2020).
9. A. M. Siepielski, J. D. DiBattista, J. A. Evans, S. M. Carlson, Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. *Proc. Biol. Sci.* **278**, 1572–1580 (2011).
10. J. G. Kingsolver, D. W. Pfennig, Patterns and power of phenotypic selection in nature. *BioScience* **57**, 561–572 (2007).
11. A. Felmy, D. N. Reznick, J. Travis, T. Potter, T. Coulson, Life histories as mosaics: Plastic and genetic components differ among traits that underpin life-history strategies. *Evolution* **76**, 585–604 (2022).
12. S. F. Chenoweth, H. D. Rundle, M. W. Blows, The contribution of selection and genetic constraints to phenotypic divergence. *Am. Nat.* **175**, 186–196 (2010).
13. B. Walsh, M. Lynch, *Evolution and Selection of Quantitative Traits* (Oxford University Press, 2018).
14. T. Bonnet *et al.*, The role of selection and evolution in changing parturition date in a red deer population. *PLoS Biol.* **17**, e3000493 (2019).
15. V. Buffalo, G. Coop, Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20672–20680 (2020).
16. A. A. Snead, F. Alda, Time-series sequences for evolutionary inferences. *Integr. Comp. Biol.* **62**, 1771–1783 (2022).
17. S. U. Franssen, R. Kofler, C. Schlötterer, Uncovering the genetic signature of quantitative trait evolution with replicated time series data. *Heredity* **118**, 42–51 (2017).
18. H. A. Orr, Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.* **10**, 531–539 (2009).
19. N. Chen *et al.*, Allele frequency dynamics in a pedigreed natural population. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2158–2164 (2019).
20. J. R. Stinchcombe, H. E. Hoekstra, Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity* **100**, 158–170 (2008).
21. N. Barghi, J. Hermisson, C. Schlötterer, Polygenic adaptation: A unifying framework to understand positive selection. *Nat. Rev. Genet.* **21**, 769–781 (2020).
22. Á. Einarsson *et al.*, The ecology of Lake Myvatn and the River Laxá: Variation in space and time. *Aquat. Ecol.* **38**, 317–348 (2004).
23. A. R. Ives, Á. Einarsson, V. A. A. Jansen, A. Gardarsson, High-amplitude fluctuations and alternative dynamical states of midges in Lake Myvatn. *Nature* **452**, 84–87 (2008).
24. J. Phillips *et al.*, Demographic basis of spatially structured fluctuations in a threespine stickleback metapopulation. *Am. Nat.* **201**, E41–E55 (2023).
25. J. S. Phillips, G. Guðbergsson, A. R. Ives, Opposing trends in survival and recruitment slow the recovery of a historically overexploited fishery. *Can. J. Fish. Aquat. Sci.* **79**, 1138–1144 (2022).
26. K. Strickland *et al.*, Genome-phenotype-environment associations identify signatures of selection in a panmictic population of threespine stickleback. *Mol. Ecol.* **32**, 1708–1725 (2023).
27. S. Liu, A.-L. Ferchaud, P. Grønkjær, R. Nygaard, M. M. Hansen, Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. *Mol. Ecol.* **27**, 4725–4743 (2018).
28. M. Roesti, B. Kueng, D. Moser, D. Berner, The genomics of ecological vicariance in threespine stickleback fish. *Nat. Commun.* **6**, 8767 (2015).
29. S. Ellner, N. G. Hairston, Role of overlapping generations in maintaining genetic variation in a fluctuating environment. *Am. Nat.* **143**, 403–417 (1994).
30. M. Rees, S. P. Ellner, Why so variable: Can genetic variance in flowering thresholds be maintained by fluctuating selection? *Am. Nat.* **194**, E13–E29 (2019).
31. A. Sasaki, S. Ellner, Quantitative genetic variance maintained by fluctuating selection with overlapping generations: Variance components and covariances. *Evolution* **51**, 682–696 (1997).
32. L. R. Hallsson, M. Björklund, Selection in a fluctuating environment leads to decreased genetic variation and facilitates the evolution of phenotypic plasticity. *J. Evol. Biol.* **25**, 1275–1290 (2012).
33. O. L. Johnson, R. Tobler, J. M. Schmidt, C. D. Huber, Fluctuating selection and the determinants of genetic variation. *Trends Genet.* **39**, 491–504 (2023).
34. T. E. Reimchen, P. Nosil, Temporal variation in divergent selection on spine number in threespine stickleback. *Evolution* **56**, 2472–2483 (2002).
35. K. Reid, M. A. Bell, K. R. Veeramah, Threespine stickleback: A model system for evolutionary genomics. *Annu. Rev. Genom. Hum. Genet.* **22**, 357–383 (2021).
36. P. Yershov, A. Sukhotin, Age and growth of marine three-spined stickleback in the White Sea 50 years after a population collapse. *Polar Biol.* **38**, 1813–1823 (2015).
37. R. J. Snyder, Migration and life histories of the threespine stickleback: Evidence for adaptive variation in growth rate between populations. *Environ. Biol. Fishes* **31**, 381–388 (1991).
38. M. B. Morrissey, I. B. J. Goudie, Analytical results for directional and quadratic selection gradients for log-linear models of fitness functions. *Evolution* **76**, 1378–1390 (2022).
39. T. E. Reimchen, Injuries on stickleback from attacks by a toothed predator (Oncorhynchus) and implications for the evolution of lateral plates. *Evolution* **46**, 1224–1230 (1992).
40. L. Yin, H. Zhang, X. Liu, hibayes: Individual-Level, Summary-Level and Single-Step Bayesian Regression Model (R package version 3.0.3, 2024). https://CRAN.R-project.org/package=hibayes. Accessed 25 June 2024.
41. 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).
42. T. D. Tran, J. Hofrichter, J. Jost, An introduction to the mathematical structure of the Wright-Fisher model of population genetics. *Theory Biosci.* **132**, 73–82 (2013).
43. P. Tataru, M. Simonsen, T. Bataillon, A. Hobolth, Statistical inference in the Wright-Fisher model using allele frequency data. *Syst. Biol.* **66**, e30–e46 (2017).
44. S. L. Archambeault, L. R. Bärtschi, A. D. Merminod, C. L. Peichel, Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback Eda locus. *Evol. Lett.* **4**, 282–301 (2020).
45. 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).
46. W. G. Hill, Understanding and using quantitative genetic variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 73–85 (2010).
47. G. Moser *et al.*, Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genet.* **11**, e1004969 (2015).
48. M. Bartrons *et al.*, Spatial patterns reveal strong abiotic and biotic drivers of zooplankton community composition in Lake Mývatn, Iceland. *Ecosphere* **6**, art105 (2015).
49. O. N. Bjørnstad, B. T. Grenfell, Noisy clockwork: Time series analysis of population fluctuations in animals. *Science* **293**, 638–643 (2001).
50. G. Bell, Fluctuating selection: The perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 87–97, (2010).
51. J. Wright, G. H. Bolstad, Y. G. Araya-Ajoy, N. J. Dingemanse, Life-history evolution under fluctuating density-dependent selection and the adaptive alignment of pace-of-life syndromes. *Biol. Rev.* **94**, 230–247 (2019).
52. C. B. Baines, J. M. J. Travis, S. J. McCauley, G. Bocedi, Negative density-dependent dispersal emerges from the joint evolution of density- and body condition-dependent dispersal strategies. *Evolution* **74**, 2238–2249 (2020).
53. D. C. Laughlin, J. Messier, Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends Ecol. Evol.* **30**, 487–496 (2015).
54. P. Edelaar, R. Jovani, I. Gomez-Mestre, Should I change or should I go? Phenotypic plasticity and matching habitat choice in the adaptation to environmental heterogeneity. *Am. Nat.* **190**, 506–520 (2017).
55. K. M. Siewert, B. F. Voight, Detecting long-term balancing selection using allele frequency correlation. *Mol. Biol. Evol.* **34**, 2996–3005 (2017).
56. D. Koenig *et al.*, Long-term balancing selection drives evolution of immunity genes in Capsella. *Elife* **8**, e43606 (2019).
57. B. C. Sheldon, L. E. B. Kruuk, S. C. Alberts, The expanding value of long-term studies of individuals in the wild. *Nat. Ecol. Evol.* **6**, 1799–1801 (2022).
58. I. Fragata, A. Blanckaert, M. A. Dias Louro, D. A. Liberles, C. Bank, Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.* **34**, 69–82 (2019).
59. I. Salazar-Ciudad, M. Marín-Riera, Adaptive dynamics under development-based genotype–phenotype maps. *Nature* **497**, 361–364 (2013).
60. A. Millet, B. K. Kristjánsson, Á. Einarsson, K. Räsänen, Spatial phenotypic and genetic structure of threespine stickleback (Gasterosteus aculeatus) in a heterogeneous natural system, Lake Mývatn, Iceland. *Ecol. Evol.* **3**, 3219–3232 (2013).
61. G. Á. Ólafsdóttir, S. S. Snorrasson, M. G. Ritchie, Postglacial intra–lacustrine divergence of Icelandic threespine stickleback morphs in three neovolcanic lakes. *J. Evol. Biol.* **20**, 1870–1881 (2007).
62. A. P. Hendry, D. I. Bolnick, D. Berner, C. L. Peichel, Along the speciation continuum in sticklebacks. *J. Fish Biol.* **75**, 2000–2036 (2009).
63. A. Härer, D. I. Bolnick, D. J. Rennison, The genomic signature of ecological divergence along the benthic-limnetic axis in allopatric and sympatric threespine stickleback. *Mol. Ecol.* **30**, 451–463 (2021).
64. M. A. Bell, Differentiation of adjacent stream populations of threespine sticklebacks. *Evolution* **36**, 189–199 (1982).
65. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
66. S. Nath, D. E. Shaw, M. A. White, Improved contiguity of the threespine stickleback genome using long-read sequencing. *G3 (Bethesda)* **11**, jkab007 (2021).
67. G. A. Van der Auwera *et al.*, From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 10–11 (2013).
68. C. L. Peichel *et al.*, Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* **21**, 177 (2020).
69. P.-C. Bürkner, brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
70. L. Yin, H. Zhang, X. Li, S. Zhao, X. Liu, hibayes: An R Package to Fit Individual-Level, Summary-Level and Single-Step Bayesian Regression Models for Genomic Prediction and Genome-Wide Association Studies. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.02.12.480230 (Accessed 19 July 2024).
71. D. C. Hunter *et al.*, Using genomic prediction to detect microevolutionary change of a quantitative trait. *Proc. R. Soc. B Biol. Sci.* **289**, 20220330 (2022).
72. B. Ashraf *et al.*, Genomic prediction in the wild: A case study in Soay sheep. *Mol. Ecol.* **31**, 6541–6555 (2022).
73. J. C. H. Aspheim, K. Aase, G. H. Bolstad, H. Jensen, S. Muff, Bayesian marker-based principal component ridge regression–A flexible multipurpose framework for quantitative genetics in wild study systems. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.06.01.596874 (Accessed 19 July 2024).
74. J. Guhlin *et al.*, Species-wide genomics of kākāpō provides tools to accelerate recovery. *Nat. Ecol. Evol.* **7**, 1693–1705 (2023).
75. T. Taus, A. Futschik, C. Schlötterer, Quantifying selection with Pool-Seq time series data. *Mol. Biol. Evol.* **34**, 3023–3034 (2017).
76. C. Paris, B. Servin, S. Boitard, Inference of selection from genetic time series using various parametric approximations to the Wright–Fisher model. *G3 (Bethesda)* **9**, 4073–4086 (2019).
77. B. C. Haller, P. W. Messer, SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
78. A.-L. Ferchaud, M. M. Hansen, The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: Three-spine sticklebacks in divergent environments. *Mol. Ecol.* **25**, 238–259 (2016).
79. M. A. Stoffel, S. E. Johnston, J. G. Pilkington, J. M. Pemberton, Purifying and balancing selection on embryonic semi-lethal haplotypes in a wild mammal. *Evol. Lett.* (2023), 10.1093/evlett/qrad053.
80. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
81. A. Alexa, J. Rahnenfuhrer, TopGo: Enrichment Analysis for Gene Ontology (Version 2.42.0. 2020, R Package, 2020). https://bioconductor.org/packages/topGO. Accessed 8 February 2023.
82. K. Strickland *et al.*, *Gasterosteus aculeatus*. Microevolutionary change in wild stickleback. European Nucleotide Archive. https://www.ebi.ac.uk/ena/browser/view/PRJEB79151. Deposited 17 August 2024.
83. K. Strickland *et al.*, WGS reads used in "Genome-phenotype-environment associations identify signatures of selection in a panmictic population of threespine stickleback." European Nucleotide Archive. https://www.ebi.ac.uk/ena/browser/view/PRJEB58765. Deposited 4 January 2023.
84. K. Strickland *et al.*, Phenotypic and meta data for "Microevolutionary change in wild stickleback: using integrative time-series data to infer responses to selection." Github. https://github.com/kashastrickland/sticklebackPhenoGenome. Accessed 21 August 2024.