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

Kasha Strickland*[1,2], Blake Matthews[3], Zophonías O. Jónsson[4], Bjarni K Kristjánsson[2], Joseph S Phillips[2,5], Árni Einarsson[4], Katja Räsänen[6,7]

1. Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, UK
2. Department of Aquaculture and Fish Biology, Hólar University, Iceland
3. Department of Fish Ecology and Evolution, Swiss Federal Institute of Aquatic Science and Technology, EAWAG, Kastanienbaum, Switzerland
4. Faculty of Life and Environmental Sciences, University of Iceland, Reykjavík, Iceland
5. Department of Biology, Creighton University, Omaha, Nebraska, USA
6. Department of Aquatic Ecology, Swiss Federal Institute of Aquatic Science and Technology, EAWAG, Duebendorf, Switzerland
7. Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

* corresponding author: Email: kasha.strickland@ed.ac.uk

Abstract

Identifying microevolutionary change in the wild requires linking trait change to shifts in allele frequencies, but existing approaches poorly account for different modes of selection that act simultaneously on correlated traits. Using an integrative phenome-genome time-series dataset collected on wild threespine stickleback (*Gasterosteus aculeatus*), we identified how different modes of selection (directional, balancing, and episodic) drive trait change over time. Specifically, we show that dietary traits in our population changed linearly by an average of 16% across 10 generations, which was linked to changes in both genomic breeding values and allele frequencies at quantitative trait loci for dietary traits. Importantly, allele frequencies at quantitative trait loci were changing at a rate greater than expected under neutrality, suggesting that dietary traits were under directional selection. We further show that swimming-related traits may be under episodic selection caused by an extreme population crash. Our study provides a unique empirical demonstration of microevolution in a wild population in which multiple modes of selection act simultaneously on different traits, which likely has important downstream consequences for the evolution of correlated traits.

Main

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 realised 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, whilst 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 if 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 characterise 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 [15–18]. Integration of these methods helps to alleviate key limitations involved with using either in isolation [19,20]. 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, whilst 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 our approach to a 10-year time-series (ca. 10 stickleback generations) of whole-genome sequencing and phenotypic measures of functional traits (trophic and defence 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).

Lake Mývatn is a highly dynamic ecosystem in which multiple ecological agents of selection, including vertebrate and invertebrate abundances, are known to change through time [21–24] likely generating strong natural selection. In particular, stickleback density fluctuates periodically (Figure 1c [23]) which would cause fluctuating density-dependent selection, and abundance of stickleback predators and prey change through time (Figure 1d and e), generating predator- or prey-mediated natural selection. The stickleback population is panmictic [25] with high levels of standing genetic variation ($H_e = 0.26 \pm 0.02$, $F_{IS} = -0.032 \pm 0.08$), despite relatively low effective population

size (N_e = 1752 ± 249) and regular population bottlenecks [23]. Similar to other freshwater populations of stickleback around the Atlantic [26], genomic PCA analyses suggest that Mývatn stickleback are polymorphic for several inversion haplotypes (Figure 1b) which are rich in QTLs and are associated with both marine-freshwater and lake-stream divergence [27]. Together, these data suggest that fluctuating selection may contribute to the maintenance of genetic and phenotypic variation in this population. We sampled Mývatn stickleback every 2 years from 2010 to 2020 (approximately 10 stickleback generations) which included an extreme population crash in the years 2014 - 2016 [23] (Figure 1c) that likely caused a strong episode of selection. We phenotyped 793 individuals, and sequenced the genomes for 515 of them. After quality control of raw sequence data, genotyping and filtering, we had just over 1.7 million biallelic SNPs located on autosomes that were used for all downstream analyses. We tested for evidence of directional, episodic and balancing selection at both the phenotypic and genomic level, aiming to determine if and how different modes of selection act simultaneously in this highly dynamic ecosystem.

Results and discussion

Genetic contribution to observed phenotypic change

We focussed on functionally important traits that fall into two general categories: the *defence traits* number of armour plates, thought to defend against fish predation, and length of pelvic and dorsal spines, thought to defend against gape-limited predators [28]; 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 [29]. We also measured total *length* as a standard measure of body size and, given that stickleback have indeterminate growth, it is also an approximate measure of age [30,31]. 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 (see materials and methods). We did not test whether length changed over time because whilst 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 between changes in length that were independent of individuals' 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) [32]. 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 trait measures relative to these parameters.

Including a year term in the models improved model fit for all traits except pelvic spine length, and the means of all but one trait (pelvic spine length) changed over the 10-year time span (Table 1). The defence phenotype shifted toward fewer plates and longer dorsal spines over time (Table 1, Figure 2b), suggesting adaptation as a response to predator induced selection [28,33]. The trophic phenotype, on the other hand, shifted toward fewer and longer gill-rakers with narrower gaps between them, and relatively longer guts (Table 1, Figure 2b), 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 those traits changing linearly, our analyses suggested that over the 10 years of the study, the number of plates decreased by 0.28 which reflected a 6% decrease across the duration of the study (mean = 5 plates), the length of gill-rakers increased by 0.09mm equating to a 8% decrease across the duration of the study (mean = 1.11mm), the number of gill-rakers decreased by 3.44 equating to a 25% increase across the duration of the study (mean = 14 rakers), and the gap width between gill-rakers decreased by 0.03mm equating to a 17% decrease across the duration of the study (mean = 0.18mm) (all estimated for a female stickleback of average length, Table 1). Trait covariances generally reflected the directionality of overall phenotypic change (Figure 2a): the lengths of the spines, of both gill-rakers and of the gut were positively correlated with each other, but negatively correlated with the number of gill-rakers. Gill-raker gap width was negatively correlated with gut length.

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 the genomic breeding values (EGBV) for each trait from genome-wide SNPs using mixture models implemented in bayesR [34–36]. We ran bayesR to estimate EGBVs by fitting trait values as residuals from a regression of traits as a function of total length (and therefore age), sex and site of capture. All traits studied here followed oligogenic expectations,

with few SNPs of large effect and many of small effect (Table S4). The traits were all associated with non-zero additive genetic variance (Table S4), indicating a high degree of evolutionary potential in all traits, and were moderately to highly heritable (h^2 = 0.32 - 0.81, Table 2). Next, we ran linear regressions to model EGBVs as a function of year which was fit as either a linear or quadratic term (see methods). EGBVs for the number of armour plates, and the lengths of dorsal spines and gill-rakers did not change through time (Table 2, Figure 2c), suggesting that the observed phenotypic trends in these traits occurred as a result of phenotypic plasticity. In contrast, EGBVs for number of gill-rakers, gill-raker gap width and gut length all changed between 2010 and 2020 (Table 2, Figure 2c). Moreover, the shape and direction of the effect of year on EGBVs for these traits mirrored the observed phenotypic trends (see Figure 2), suggesting that phenotypic trends in these trophic traits were caused by microevolutionary processes causing changes to the genetic architecture.

Allele frequency dynamics identify the modality of selection

Whilst tracking temporal change in EGBVs can inform whether observed trait change has occurred as a result of changes to the underlying genetic architecture, it does not provide insight into changes at the genomic sequence level nor directly infer whether that change has occurred as a result of selection [20]. 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 with threespine stickleback [27,37]), 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 [38,39]. 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 one year 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 (see materials and methods). Loci were then characterised as under: *directional selection* when its allele frequency trajectory changed linearly from 2010 - 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 - 2016), but returned to an expected trajectory after this period, or; *balancing selection* if its allele frequency changed less than expected (Figure 3a).

We identified 104 SNPs under directional selection (Figure 3, Table S6). Of these, 81% (N = 83) were on a known QTL in stickleback [37]: 71% (N = 74) fell on a QTL for a feeding trait, 53% (N = 55) for a defence trait, 25% (N = 26) for a locomotion related trait (e.g., fin rays, vertebrae number and pterygiophore), 22% (N = 23) for a respiration trait (e.g. operculum morphology) and <%1 (N = 1) for pigmentation. Many regions of the stickleback genome have been linked to multiple phenotypes [37,40,41], and so we next identified whether there were trait categories that were overrepresented in the regions identified as being under directional selection. To do this, we used Fisher's exact tests to identify whether the probability that a QTL identified as under 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 QTLs associated with that trait [37]. We found that QTLs 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 (Table S7). The 104 SNPs under directional selection fell on 37 genes (Table S8). Gene ontology (GO) analyses identified 11 enriched terms (Table S9), including four terms associated with fatty acid metabolism, three with development of left-right asymmetry, one with neuron differentiation and one with response to oestrogen stimulus.

We identified 818 SNPs under episodic selection (Figure 3c, Figure S4, Table S6). 701 of these SNPs (86.2%) fell on a known QTL in stickleback [37]. 81% (N = 663) for a feeding trait, 53% (N = 436) for a defence trait, 30% (N = 242) for a locomotion related trait, 22% (N = 180) for a respiration trait and 4% (N = 29) for pigmentation. None of these trait types were significantly overrepresented when compared to the overall probability of a QTL on the stickleback genome being associated with each trait type, suggesting that observed associations with QTLs could have occurred due to chance (Table S7). The 818 SNPs under episodic selection fell on 280 genes (Table S8), and GO analyses identified 14 enriched terms (Table S9), seven of which were associated with anatomical structural development and two with processes involved in locomotory behaviour. 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 EGBVs: the genomic component of both gill-raker number and gap width changed linearly over time. Furthermore, the top GO term associated with loci under directional selection was associated with fatty acid metabolism, corroborating that dietary traits were under directional selection in this population. The phenotypes implicated as under episodic selection may be linked to swimming physiology and behaviour given that loci under episodic selection mapped to swimming related traits, albeit at a rate expected given the distribution of swimming QTLs on the genome, and terms linked to locomotory behaviour were over represented in GO analysis. This is a clear demonstration that whilst we were able to generate a large dataset consisting of individual-level measurements for many important 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 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 for both directional and episodic selection were associated with structural development, indicating that unmeasured morphological phenotypes are under selection. Taken together, these results clearly demonstrate how multiple modes of selection act simultaneously on different components of the phenotype to shape trait evolution in a wild population.

Conclusions

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 QTLs for trophic traits. Importantly, this result provides a unique empirical example which tests the assumptions used in each field: although genomic breeding values should sum the additive effects of causal loci affecting traits [42], and are derived directly from allelic variation [36], 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 [22,43]. However, ecological dynamics which generate the selective pressures acting on the phenotype are rarely linear [44], and fluctuating selection is thought to be a dominant mode of selection in natural populations [45]. In concordance, we show that genes under episodic selection during the years the population crashed were enriched for locomotory behaviour, 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 [46], whereby in the low density years selection may have been favouring philopatry over dispersal due to reduced competition [47]. 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 [48]. This is further compounded by the role of phenotypic plasticity and habitat choice when organisms experience spatio-temporally variable environments [49], making genome- or phenotype-environment associations tricky to interpret.

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, although Mývatn stickleback likely experience strong fluctuating density-dependent selection associated with the cyclic population dynamics of the population (approximately 6-year cycles [23]). Such repeated episodes of selection may not always culminate into fluctuating selection *per se* as they may instead reflect long-term balancing selection, which would ultimately favour the maintenance of multiple alleles over time [50,51]. 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. Indeed, whilst 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 [52].

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 [53,54], 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 [27,37], 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.

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 [21,23,25,55]. For instance, chironomid midge, threespine stickleback, piscivorous birds and Arctic charr population abundances are all known to vary through time [22–24]. Stickleback habitats consist of five main types [55], and there is evidence for subtle phenotypic and genetic spatial divergence in Mývatn stickleback [25,55,56]. Furthermore, there is spatial variation in water temperature [21], avian predators [25] and invertebrate prey abundances and community structure [43]. 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 [21,23], with five shorelines sites added in 2009 [55] (Figure 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 hr periods (night and day catch) [55]. Stickleback from all traps are counted to estimate catch-per-unit-effort (CPUE) [23] 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 35mm) from the June samples from 10 study sites (Figure 1) every other year between 2010 and 2020 (N = 793, 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 [25]: defence traits (armour plate number and length of spines) and trophic traits (gill raker morphology and gut length) [29,57,58].

On each individual we measured the following 10 traits: total length, number of lateral armour plates (plate number, excluding the keel), length of the first dorsal spine (DS1), length of the

second dorsal spine (DS2), length of the pelvic spine (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 oesophagus 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 alizarine red using standard protocols[25,55]. 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 [59]. 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 millimetre. After imaging, we dissected the first gill arch and, where necessary, re-stained it before mounting between two glass plates and photographing with a digital camera (Nikon Coolpix 4500) mounted to a stereomicroscope (Leica MZ12), 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 [60].

Whole genome resequencing and bioinformatics

For genomic analyses, we randomly selected 10 of the 20 individuals that had been phenotyped from each site/year combination (N = 515, 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 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 DNBSeqTM platform by BGI-Hongkong to an average of 10X depth of coverage [25]. All samples were mapped to v5 of the stickleback reference genome [61] and genotyped using the GATK best practices pipeline [62]. 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 [63]. 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 1700436 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 fixed effects 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 less complex 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 standard deviation of 1 to ensure gaussian errors were appropriate. All models were fit with sex as a fixed effect to account for sexual dimorphism and length as a fixed effect to account for allometry and age (because stickleback have indeterminate growth). All results presented therefore reflect trait measures relative to sex and length. Whilst 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 analyse length as a quantitative trait. All models were fit using Stan via the brms package in *R* statistical environment (version 4.1.2) [64]. All models were run with 6000 iterations across four chains and a warm up period of 2000 iterations, which was sufficient in all cases to achieve model convergence which was assessed by visually assessing mixing of chains and with $\hat{R}$. Fixed effects were given normal priors with 0 mean and standard deviation of 5. Random effects were given half-cauchy priors with 2 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 standardised traits as a multivariate response as a function of length and sex but without accounting for spatial or temporal effects. Phenotypic correlations were then estimated from the residual covariance matrix.

To test whether any of the measured traits were changing 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 was changing across time. The third model was fit with an identical structure, except with an additional quadratic term for year in order to test whether population mean of the trait was changing nonlinearly. 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 model support for the addition of either the linear, or linear and quadratic term for year. In cases where $\Delta WAIC$ 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 modelled 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 if microevolutionary change was responsible for any observed trait change, we tested whether genomic breeding values [13] for traits changed as a function of sampling year. To do this, estimated genomic breeding values (EGBV) for each trait were estimated for all individuals that were genotyped and phenotyped (N = 515, Table S1) using bayesR following methodology detailed in [34–36]. Breeding values are normally 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 [42]. Although many traits are highly polygenic, analyses such as those implemented in bayesR allow for more accurate mapping of genomic architectures, and should therefore improve our predictions of microevolutionary change [36]. Briefly, we first corrected trait values for sex, sampling site and length (and therefore also age) using linear models in R. Then, we used the residuals from these models as the trait values in a mixture model in bayesR to estimate EGBVs per individual per trait. The bayesR models were fit with 50000 iterations, a burn-in set of 20000 iterations and a thinning interval of 10, generating a posterior distribution for EGBVs for each trait. SNP-based h^2 for each trait was estimated directly from estimates of additive genetic variance (V_A) and residual variance (V_e) from bayesR, and was calculated as the proportion of total phenotypic variance attributed to V_A.

We then fit a linear model of EGBVs 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 EGBVs and sampling year. 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.

Allele frequency dynamics

To analyse changes in allele frequencies through time, we calculated allele frequency per SNP per year (2010, 2012, 2014, 2016, 2018, 2020) for individuals in the North basin (see Table S1 for sample sizes). We investigated allele frequency dynamics using samples collected from the North basin in order 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 (Table S1) and the sample sizes from the South basin were very variable. We do not believe this should effect downstream inferences because (1) stickleback density is much higher in the North and likely subsidises the South basin [23], and (2) although there is some evidence that allele frequencies and population densities vary between basins [23,25], population genetic analyses suggest that the population is panmictic across the whole lake [25]. To investigate allele frequency dynamics, we also subset the SNPs to retain those with a call rate of 50% within each year and a MAF in 2010 of at least 0.001 (Figure S2). This resulted in a total of 1558025 SNP loci used for this analysis.

We aimed to identify whether loci were under one of three modes of selection: those where allele frequencies changed via directional selection, episodic selection or balancing selection. We identified these by comparing the observed allele frequency change for each locus to the expected change in allele frequency caused by drift (or other neutral processes). To isolate SNPs where allele frequency was changing due to selection as opposed to due to drift or sampling regimes, observed allele frequency trajectories of each SNP was compared to an allele frequency trajectory that would be expected under a Wright-Fisher model. 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, Table S3) 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 year. 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. The process was then replicated for another 4 steps, resulting in a total of 10 generations and 6 sampling points (bi-annually 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 [65].

A given SNP was characterised as under *directional selection* if its allele frequency was 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-year cycles [23]). Out time series was known to overlap with one of these cycles, with a large population crash in 2014 – 2016. As such, SNPs were characterised as under *episodic 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 diverges when the population crashes but returns to a trajectory expected under Wright-Fisher afterwards. Both of these approaches to determining if a SNPs allele frequency trajectory deviates from

a Wright-Fisher model represent a highly conservative method; for a SNP to be determined as under either mode of selection, the probability that its allele frequency trajectory occurred as a result of neutral processes must be 0 (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 have $P = 0$. 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 when loci changed less than expected by neutral processes, and whilst it may be a relatively conservative approach, this definition has been used in previous research aiming to identify short-term balancing selection [66]. 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 to maintain phenotypic and allelic variation in a population [50,51]. The Wright-Fisher model we used to compare observed allele frequency trajectories accounted for differences in allele frequency by starting the model with the observed allele frequency for each locus at time t^0, and also accounted for the variable sample sizes and populations sizes observed across the time-series by incorporating those observed parameters into the model (Table S1 and S3). 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.

For each set of SNPs identified to be putatively under a given type of selection, we identified the QTLs these SNPs fell on using lists from [37]. 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 [25] . Using the trait categories described in [37] (i.e., feeding, defence, swimming/locomotion, pigmentation and respiration), we used Fishers exact tests to compare the proportion of QTLs 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 QTLs previously mapped in stickleback that fall into that trait category.

We finally identified the protein-coding genes on which these SNPs were located (i.e., within the transcribed regions) and ran gene ontology (GO) analyses to explore whether any molecular functions were overrepresented in sets of genes associated with selection. 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 [67], and functional enrichment was investigated using the package TOPGO 2.42 [68] 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").

Acknowledgements: 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 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 Figure 1 came from previous publications [24], but we would like to thank Guðni Guðbergsson for collecting these data.

Funding: This work was supported by an Icelandic Research Council (RANNIS) Grant of Excellence (grant number: 195571-052).

Author contributions:

Conceptualization: KS, BM, KR

Data Curation: KS, ZOJ, BKK, JP, AE

Formal Analysis: KS

Funding acquisition: BM, ZOJ, BKK, AE, KR

Investigation: KS, BM, KR

Methodology: KS, BM, ZOJ

Project Administration: KS, BKK, KR

Resources: KS, ZOJ, BKK

Software: KS, ZOJ

Supervision: ZOJ, BKK, KR

Validation: KS, BM, JP, KR

Visualisation: KS

Writing – original draft: KS, BM, KR

Writing – review & editing: all authors

Competing interests: Authors declare that they have no competing interests.

References

1. Bonnet, T. *et al.* Genetic variance in fitness indicates rapid contemporary adaptive evolution in wild animals. *Science* **376**, 1012–1016 (2022).
2. Merilä, J., Sheldon, B. C. & Kruuk, L. E. B. Explaining stasis: microevolutionary studies in natural populations. *Genetica* **112**, 199–222 (2001).
3. Buggs, R. J. A. The challenge of demonstrating contemporary natural selection on polygenic quantitative traits in the wild. *Mol. Ecol.* **31**, 6383–6386 (2022).
4. Pujol, B. *et al.* The Missing Response to Selection in the Wild. *Trends Ecol. Evol.* **33**, 337–346 (2018).
5. Pfenninger, M. & Foucault, Q. Population Genomic Time Series Data of a Natural Population Suggests Adaptive Tracking of Fluctuating Environmental Changes. *Integr. Comp. Biol.* **62**, 1812–1826 (2022).
6. Pfenninger, M., Foucault, Q., Waldvogel, A.-M. & Feldmeyer, B. Selective effects of a short transient environmental fluctuation on a natural population. *Mol. Ecol.* **32**, 335–349 (2023).
7. Rudman, S. M. *et al.* Direct observation of adaptive tracking on ecological time scales in Drosophila. *Science* **375**, eabj7484 (2023).
8. de Villemereuil, P. *et al.* Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. *Proc. Natl. Acad. Sci.* **117**, 31969–31978 (2020).
9. Siepielski, A. M., DiBattista, J. D., Evans, J. A. & Carlson, S. M. Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. *Proc. Biol. Sci.* **278**, 1572–1580 (2011).
10. Kingsolver, J. G. & Pfennig, D. W. Patterns and Power of Phenotypic Selection in Nature. *BioScience* **57**, 561–572 (2007).

11. Felmy, A., Reznick, D. N., Travis, J., Potter, T. & Coulson, T. Life histories as mosaics: Plastic and genetic components differ among traits that underpin life-history strategies. *Evolution* **76**, 585–604 (2022).
12. Chenoweth, S. F., Rundle, H. D. & Blows, M. W. The Contribution of Selection and Genetic Constraints to Phenotypic Divergence. *Am. Nat.* **175**, 186–196 (2010).
13. Walsh, B. & Lynch, M. *Evolution and Selection of Quantitative Traits*. (Oxford University Press, 2018).
14. Bonnet, T. *et al.* The role of selection and evolution in changing parturition date in a red deer population. *PLOS Biol.* **17**, e3000493 (2019).
15. Buffalo, V. & Coop, G. Estimating the genome-wide contribution of selection to temporal allele frequency change. *Proc. Natl. Acad. Sci.* **117**, 20672–20680 (2020).
16. Snead, A. A. & Alda, F. Time-Series Sequences for Evolutionary Inferences. *Integr. Comp. Biol.* **62**, 1771–1783 (2022).
17. Franssen, S. U., Kofler, R. & Schlötterer, C. Uncovering the genetic signature of quantitative trait evolution with replicated time series data. *Heredity* **118**, 42–51 (2017).
18. Orr, H. A. Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.* **10**, 531–539 (2009).
19. Stinchcombe, J. R. & Hoekstra, H. E. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158–170 (2008).
20. Barghi, N., Hermisson, J. & Schlötterer, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nat. Rev. Genet.* **21**, 769–781 (2020).
21. Einarsson, Á. *et al.* The ecology of Lake Myvatn and the River Laxá: Variation in space and time. *Aquat. Ecol.* **38**, 317–348 (2004).

22. Ives, A. R., Einarsson, Á., Jansen, V. A. A. & Gardarsson, A. High-amplitude fluctuations and alternative dynamical states of midges in Lake Myvatn. *Nature* **452**, 84–87 (2008).
23. Phillips, J. *et al.* Demographic basis of spatially structured fluctuations in a threespine stickleback metapopulation. *Am. Nat.* (2023).
24. Phillips, J. S., Guðbergsson, G. & Ives, A. R. Opposing trends in survival and recruitment slow the recovery of a historically overexploited fishery. *Can. J. Fish. Aquat. Sci.* **79**, 1138–1144 (2022).
25. Strickland, K. *et al.* Genome-phenotype-environment associations identify signatures of selection in a panmictic population of threespine stickleback. *Mol. Ecol.* **32**, 1708–1725 (2023).
26. Liu, S., Ferchaud, A.-L., Grønkjær, P., Nygaard, R. & Hansen, M. M. Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. *Mol. Ecol.* **27**, 4725–4743 (2018).
27. Roesti, M., Kueng, B., Moser, D. & Berner, D. The genomics of ecological vicariance in threespine stickleback fish. *Nat. Commun.* **6**, 8767 (2015).
28. Reimchen, T. E. & Nosil, P. Temporal variation in divergent selection on spine number in threespine stickleback. *Evolution* **56**, 2472–2483 (2002).
29. Reid, K., Bell, M. A. & Veeramah, K. R. Threespine Stickleback: A Model System For Evolutionary Genomics. *Annu. Rev. Genomics Hum. Genet.* **22**, 357–383 (2021).
30. Yershov, P. & Sukhotin, A. Age and growth of marine three-spined stickleback in the White Sea 50 years after a population collapse. *Polar Biol.* **38**, 1813–1823 (2015).
31. Snyder, R. J. Migration and life histories of the threespine stickleback: evidence for adaptive variation in growth rate between populations. *Environ. Biol. Fishes* **31**, 381–388 (1991).

32. Morrissey, M. B. & Goudie, I. B. J. Analytical results for directional and quadratic selection gradients for log-linear models of fitness functions. *Evolution* **76**, 1378–1390 (2022).
33. Reimchen, T. E. Injuries on Stickleback from Attacks by a Toothed Predator (Oncorhynchus) and Implications for the Evolution of Lateral Plates. *Evolution* **46**, 1224–1230 (1992).
34. Hunter, D. C. *et al.* Using genomic prediction to detect microevolutionary change of a quantitative trait. *Proc. R. Soc. B Biol. Sci.* **289**, 20220330 (2022).
35. Ashraf, B. *et al.* Genomic prediction in the wild: A case study in Soay sheep. *Mol. Ecol.* **31**, 6541–6555 (2022).
36. Moser, G. *et al.* Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *PLOS Genet.* **11**, e1004969 (2015).
37. Peichel, C. L. & Marques, D. A. The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20150486 (2017).
38. Tran, T. D., Hofrichter, J. & Jost, J. An introduction to the mathematical structure of the Wright–Fisher model of population genetics. *Theory Biosci.* **132**, 73–82 (2013).
39. Tataru, P., Simonsen, M., Bataillon, T. & Hobolth, A. Statistical Inference in the Wright–Fisher Model Using Allele Frequency Data. *Syst. Biol.* **66**, e30–e46 (2017).
40. Archambeault, S. L., Bärtschi, L. R., Merminod, A. D. & Peichel, C. L. Adaptation via pleiotropy and linkage: association mapping reveals a complex genetic architecture within the stickleback Eda locus. *Evol. Lett.* **4**, 282–301 (2020).
41. O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D. & Kingsley, D. M. A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *eLife* **4**, e05290 (2015).

42. Hill, W. G. Understanding and using quantitative genetic variation. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 73–85 (2010).
43. Bartrons, M. *et al.* Spatial patterns reveal strong abiotic and biotic drivers of zooplankton community composition in Lake Mývatn, Iceland. *Ecosphere* **6**, art105 (2015).
44. Bjørnstad, O. N. & Grenfell, B. T. Noisy Clockwork: Time Series Analysis of Population Fluctuations in Animals. *Science* **293**, 638–643 (2001).
45. Bell, G. Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 87–97 (2010).
46. Wright, J., Bolstad, G. H., Araya-Ajoy, Y. G. & Dingemanse, N. J. Life-history evolution under fluctuating density-dependent selection and the adaptive alignment of pace-of-life syndromes. *Biol. Rev.* **94**, 230–247 (2019).
47. Baines, C. B., Travis, J. M. J., McCauley, S. J. & Bocedi, G. Negative density-dependent dispersal emerges from the joint evolution of density- and body condition-dependent dispersal strategies. *Evolution* **74**, 2238–2249 (2020).
48. Laughlin, D. C. & Messier, J. Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends Ecol. Evol.* **30**, 487–496 (2015).
49. Edelaar, P., Jovani, R. & Gomez-Mestre, I. 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).
50. Siewert, K. M. & Voight, B. F. Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. *Mol. Biol. Evol.* **34**, 2996–3005 (2017).
51. Koenig, D. *et al.* Long-term balancing selection drives evolution of immunity genes in Capsella. *eLife* **8**, e43606 (2019).

631 52. Sheldon, B. C., Kruuk, L. E. B. & Alberts, S. C. The expanding value of long-term studies of
632 individuals in the wild. *Nat. Ecol. Evol.* **6**, 1799–1801 (2022).
633 53. Fragata, I., Blanckaert, A., Dias Louro, M. A., Liberles, D. A. & Bank, C. Evolution in the
634 light of fitness landscape theory. *Trends Ecol. Evol.* **34**, 69–82 (2019).
635 54. Salazar-Ciudad, I. & Marín-Riera, M. Adaptive dynamics under development-based
636 genotype–phenotype maps. *Nature* **497**, 361–364 (2013).
637 55. Millet, A., Kristjánsson, B. K., Einarsson, Á. & Räsänen, K. Spatial phenotypic and genetic
638 structure of threespine stickleback (Gasterosteus aculeatus) in a heterogeneous natural
639 system, Lake Mývatn, Iceland. *Ecol. Evol.* **3**, 3219–3232 (2013).
640 56. Ólafsdóttir, G. Á., Snorrasson, S. S. & Ritchie, M. G. Postglacial intra-lacustrine divergence
641 of Icelandic threespine stickleback morphs in three neovolcanic lakes. *J. Evol. Biol.* **20**,
642 1870–1881 (2007).
643 57. Hendry, A. P., Bolnick, D. I., Berner, D. & Peichel, C. L. Along the speciation continuum in
644 sticklebacks. *J. Fish Biol.* **75**, 2000–2036 (2009).
645 58. Härer, A., Bolnick, D. I. & Rennison, D. J. The genomic signature of ecological divergence
646 along the benthic-limnetic axis in allopatric and sympatric threespine stickleback. *Mol. Ecol.*
647 **30**, 451–463 (2021).
648 59. Bell, M. A. Differentiation of adjacent stream populations of threespine sticklebacks.
649 *Evolution* 189–199 (1982).
650 60. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image
651 analysis. *Nat. Methods* **9**, 671–675 (2012).
652 61. Nath, S., Shaw, D. E. & White, M. A. Improved contiguity of the threespine stickleback
653 genome using long-read sequencing. *G3 GenesGenomesGenetics* **11**, (2021).

62. Van der Auwera, G. A. *et al.* From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 10–11 (2013).
63. Peichel, C. L. *et al.* Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. *Genome Biol.* **21**, 177 (2020).
64. Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
65. Taus, T., Futschik, A. & Schlötterer, C. Quantifying Selection with Pool-Seq Time Series Data. *Mol. Biol. Evol.* **34**, 3023–3034 (2017).
66. Stoffel, M. A., Johnston, S. E., Pilkington, J. G. & Pemberton, J. M. Purifying and balancing selection on embryonic semi-lethal haplotypes in a wild mammal. *Evol. Lett.* qrad053 (2023) doi:10.1093/evlett/qrad053.
67. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
68. Alexa, A. & Rahnenfuhrer, J. TopGo: Enrichment Analysis for Gene Ontology; R Package Version 2.42. 0. 2020. (2020).

Data and materials availability: Upon acceptance for publication, we will submit all raw reads from sequencing to ENA and phenotypic data to dryad.

Tables & Figures

Figure 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 are a known inversion polymorphism in stickleback; (c) stickleback population dynamics across the duration of the study plotted as total abundance on the log-scale (see [23] for further details); (d) population dynamics for key stickleback predators, sum total of all piscivorous birds (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 Arctic charr (orange; *Salvelinus alpinus*), across the duration of the study plotted as total abundances on the log-scale; (e) population dynamics for *Tanytarsus gracilentus* (green), a key stickleback prey item, across the duration of the study plotted as total abundance on the log-scale.

Figure 2. Patterns of phenotypic variation. (a) Heatmap of phenotypic correlations estimated from a multivariate model accounting for length and sex. Values shown when posterior distribution of a 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 standard deviation 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 non-zero except for PS (Table 1); (c) Temporal trends in estimated genomic breeding values (EGBV) for all traits. Solid lines indicate the posterior mean of the predicted change in EGBVs for each trait with yearly average plotted as black points and the standard deviation 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% confidence intervals around the expected trend. Traits plotted are Plates - number of armour plates, SP1 – length of 1st dorsal spine, PS - pelvic spine length, GRN – gill-raker number, GRW – gill-raker gap width, GRL2 – length of 2nd gill-raker, gut length.

Figure 3. Patterns of allele frequency change. (a) Schematic demonstrating how SNPs were identified as under different modes of selection. For a single SNP, each solid grey 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) (c) under episodic selection (loci on chr XIII are shown for illustrative purposes, see Figure S4 for all loci).

Table 1. Table summarising results from linear models estimating phenotypic change through time. Summary data for raw data included, including overall mean ("Trait mean"), standard deviation ("SD"), variance among annual means ("Variance annual means") and average standard error of annual means ("Average SE annual means"). *WAIC* for models with no effect of year, ("No year") a linear effect of time ("Year") vs a linear and quadratic effect of time ("Year^2") is shown, with lowest value (indicating the best fit model) shown in bold. Regression coefficients ($\boldsymbol{\beta}$) shown for the best fitting model (i.e., quadratic term shown only when model supports inclusion) and in bold shows values with statistical support (i.e., posterior distribution does not overlap with zero). In cases where $\Delta WAIC$ between linear and quadratic model was within 2 (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 $\boldsymbol{\beta}$ coefficients shown are posterior means with 95% CIs of posterior in parentheses. r^2 estimated for the best fitting model.

Trait (units)	**Summary data**				***WAIC***			***β***		***r²***
	Trait mean	SD	Variance annual means	Average SE annual means	No year	Year	Year + Year^2	Year	Year^2	
GRL2 (mm)	1.09	0.26	0.02	0.0004	1066.1	**1047**	1046.1	**0.03 (0.02 - 0.05)**	-	0.66 (0.63 - 0.68)
GRL3 (mm)	1.12	0.26	0.02	0.0005	1054.5	**1024.3**	1024.6	**0.04 (0.03 - 0.05)**	-	0.68 (0.65 - 0.70)
GRN (n)	14.12	2.87	4.38	0.03	1580.4	1474.4	**1472.5**	**-0.12 (-0.14 - -0.09)**	-	0.25 (0.19 - 0.30)
GRW (mm)	0.18	0.05	0.001	0.00002	1309.7	**1252.2**	1252.3	**-0.06 (-0.08 - -0.05)**	-	0.52 (0.48 - 0.56)
Gut length (mm)	28.39	10.13	27.09	0.79	1122.9	1014.9	**998**	**-13.77 (-22.65 - -4.85)**	**0.00 (0.00 - 0.01)**	0.73 (0.70 - 0.74)
Plates (n)	4.97	0.92	0.02	0.01	1918.4	**1912**	1912.1	**-0.03 (-0.05 - -0.01)**	-	0.07 (0.04 - 0.10)
PS (mm)	5.12	1.06	0.24	0.007	**1336.1**	1337.6	1337.9	0.01 (-0.01 - 0.02)	-	0.59 (0.56 - 0.62)
SP1 (mm)	3.16	0.66	0.12	0.003	1264.2	1237.9	**1223.2**	**-13.22 (-22.16 - -4.18)**	**0.003 (0.001 - 0.01)**	0.65 (0.62 - 0.67)
SP2 (mm)	3.38	0.7	0.13	0.003	1304.8	1287.4	**1274.2**	**-12.51 (-21.57 - -3.52)**	**0.003 (0.001 - 0.01)**	0.63 (0.59 - 0.65)

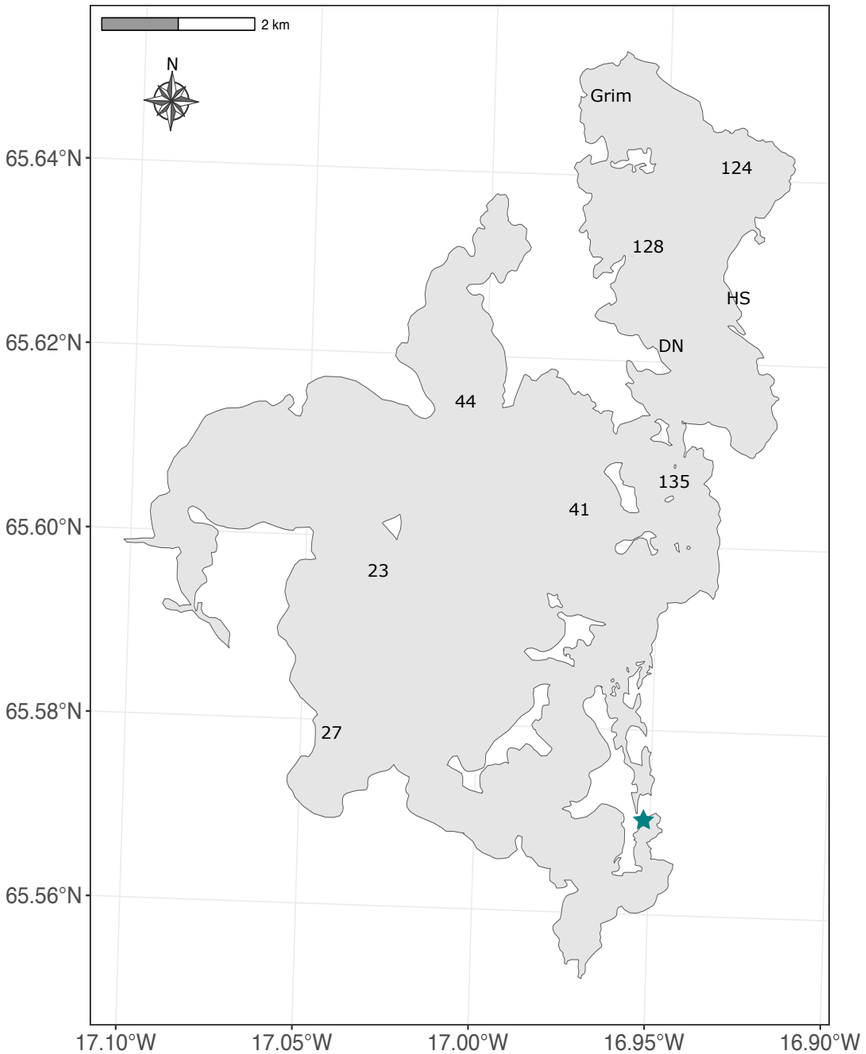
Traits included were Plates - number of armour plates, SP1 – length of 1st dorsal spine, SP2 – length of 2nd dorsal spine, PS - pelvic spine length, GRN – gill-raker number, GRW – gill-raker gap width, GRL2 – length of 2nd gill-raker, GRL3 – length of 3rd gill-raker, gut length. All traits were standardised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

Table 2. Table summarising results from linear models estimating changes in estimated genomic breeding values (EGBVs) through time. h^2 shows narrow-sense heritability for all traits based on analyses across all years. The regression coefficients (***β***) effects of year on EGBVs are shown, and the quadratic effect of year ("Year^2") was only fit where there was statistical support for the effect in phenotypic analyses (see Table 1). All coefficients shown are posterior means with 95% CIs of posterior in parentheses.

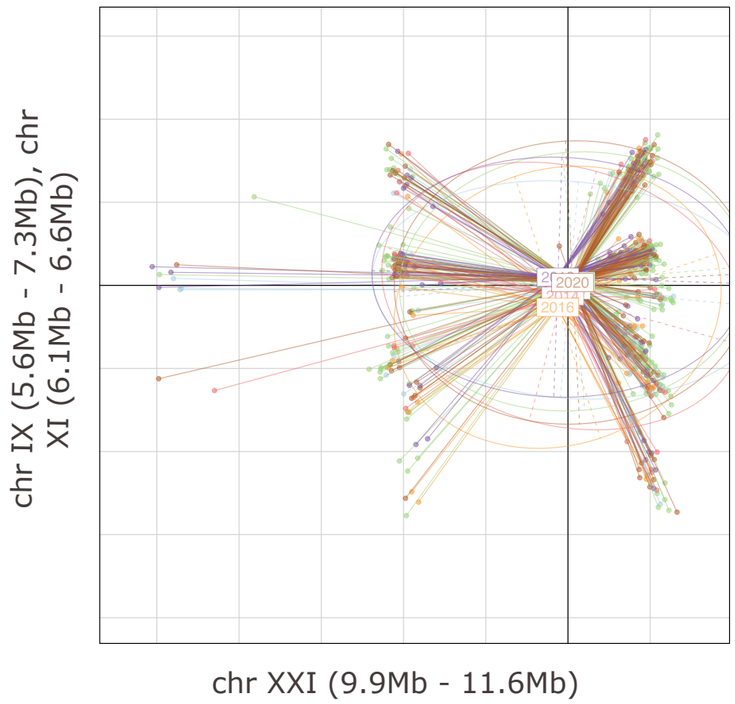
Trait (units)	***β***		***h²***
	Year	Year^2	
GRL2 (mm)	0.03 (-0.00003 - 0.04)	-	0.76
GRL3 (mm)	0.03 (-0.0001 - 0.05)	-	0.67
GRN (n)	**-0.12 (-0.13 - -0.10)**	-	0.81
GRW (mm)	**-0.03 (-0.05 - -0.01)**	-	0.8
Gut length (mm)	**-68.95 (-88.17 - -46.55)**	**0.02 (0.01 - 0.02)**	0.32
Plates (n)	-0.01 (-0.03 - 0.0002)	-	0.45
PS (mm)	0.003 (-0.002 - 0.01)	-	0.58
SP1 (mm)	-37.77 (-63.68 - 0.14)	0.01 (-0.00003 - 0.02)	0.63
SP2 (mm)	-44.48 (-62.88 - 0.02)	0.01 (-0.00001 - 0.02)	0.74

Traits included were Plates - number of armour plates, SP1 – length of 1st dorsal spine, SP2 – length of 2nd dorsal spine, PS - pelvic spine length, GRN – gill-raker number, GRW – gill-raker gap width, GRL2 – length of 2nd gill-raker, GRL3 – length of 3rd gill-raker, gut length. All traits were standardised to have mean of zero and standard deviation of one before analysis.

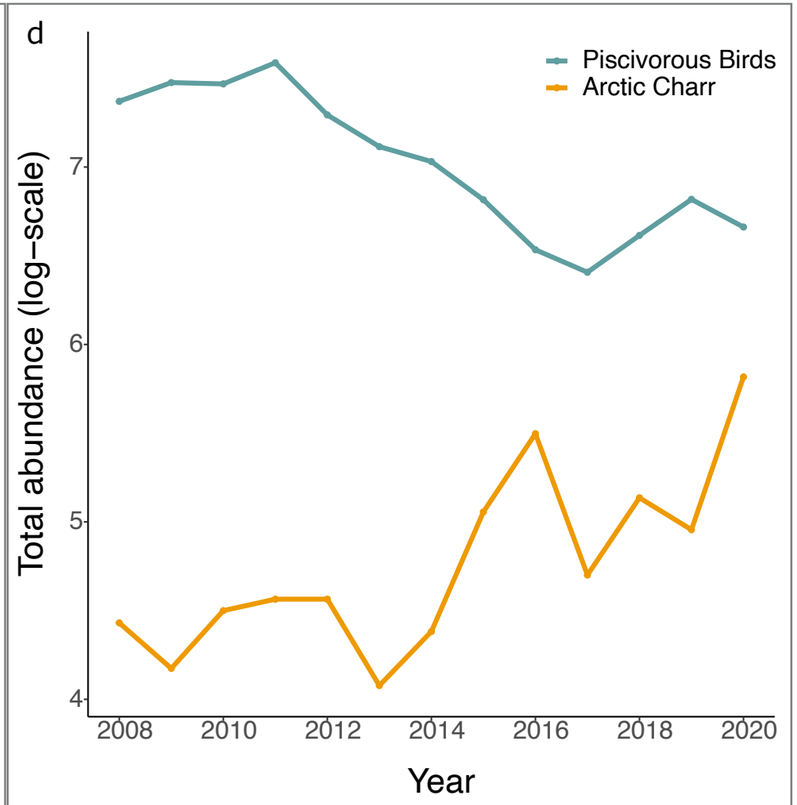
a


2 km
N
65.64°N
65.62°N
65.60°N
65.58°N
65.56°N
17.10°W
17.05°W
17.00°W
16.95°W
16.90°W
Grim
124
128
HS
DN
44
135
41
23
27

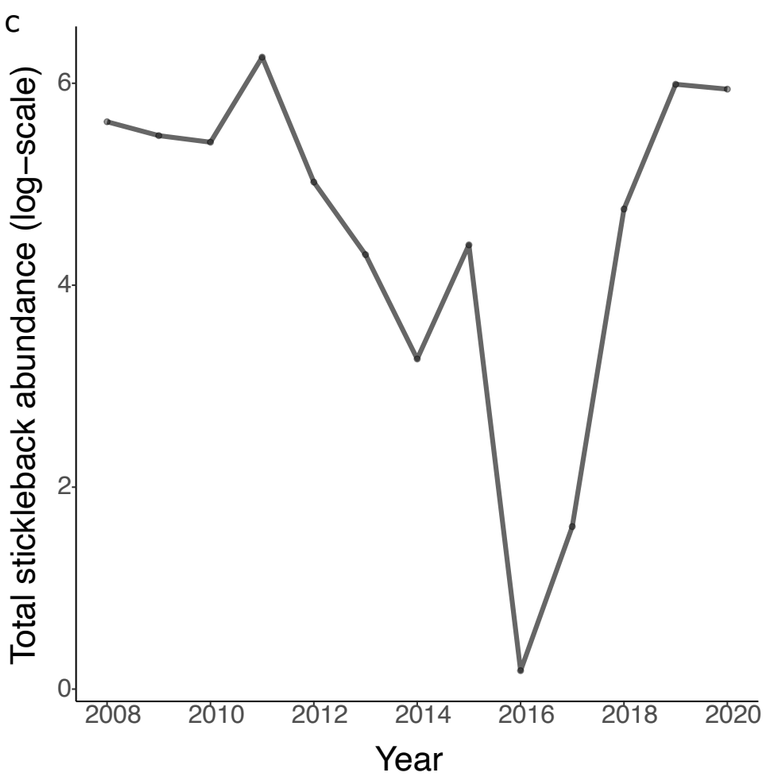
b


chr IX (5.6Mb - 7.3Mb), chr XI (6.1Mb - 6.6Mb)
chr XXI (9.9Mb - 11.6Mb)
2020
2016

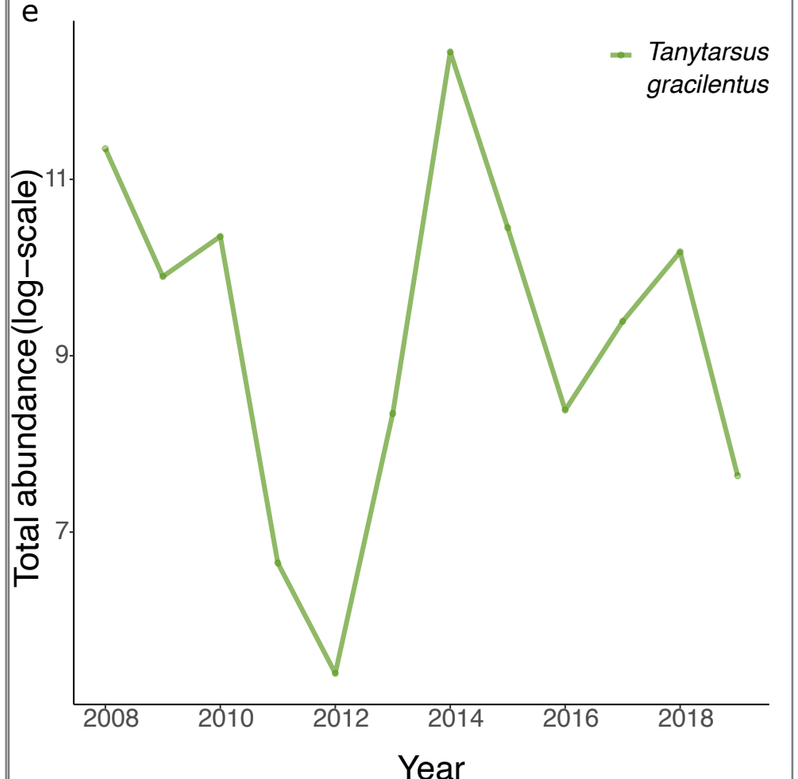
d


Piscivorous Birds
Arctic Charr
Total abundance (log−scale)
7
6
5
4
2008
2010
2012
2014
2016
2018
2020
Year

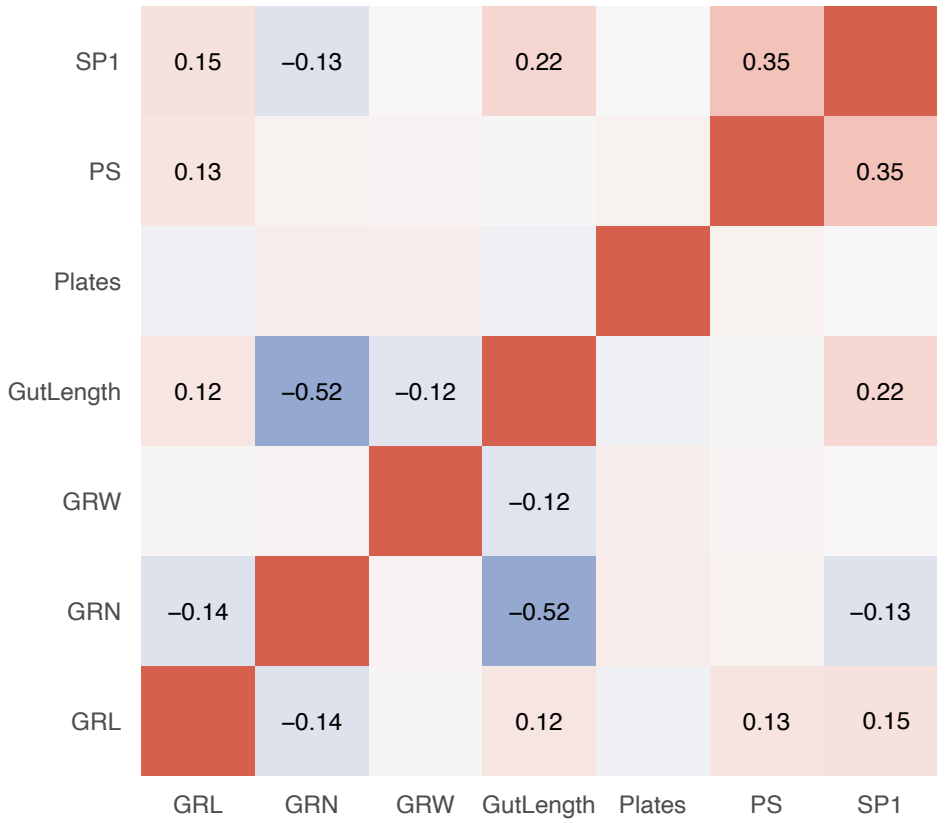
c


Total stickleback abundance (log−scale)
6
4
2
0
2008
2010
2012
2014
2016
2018
2020
Year

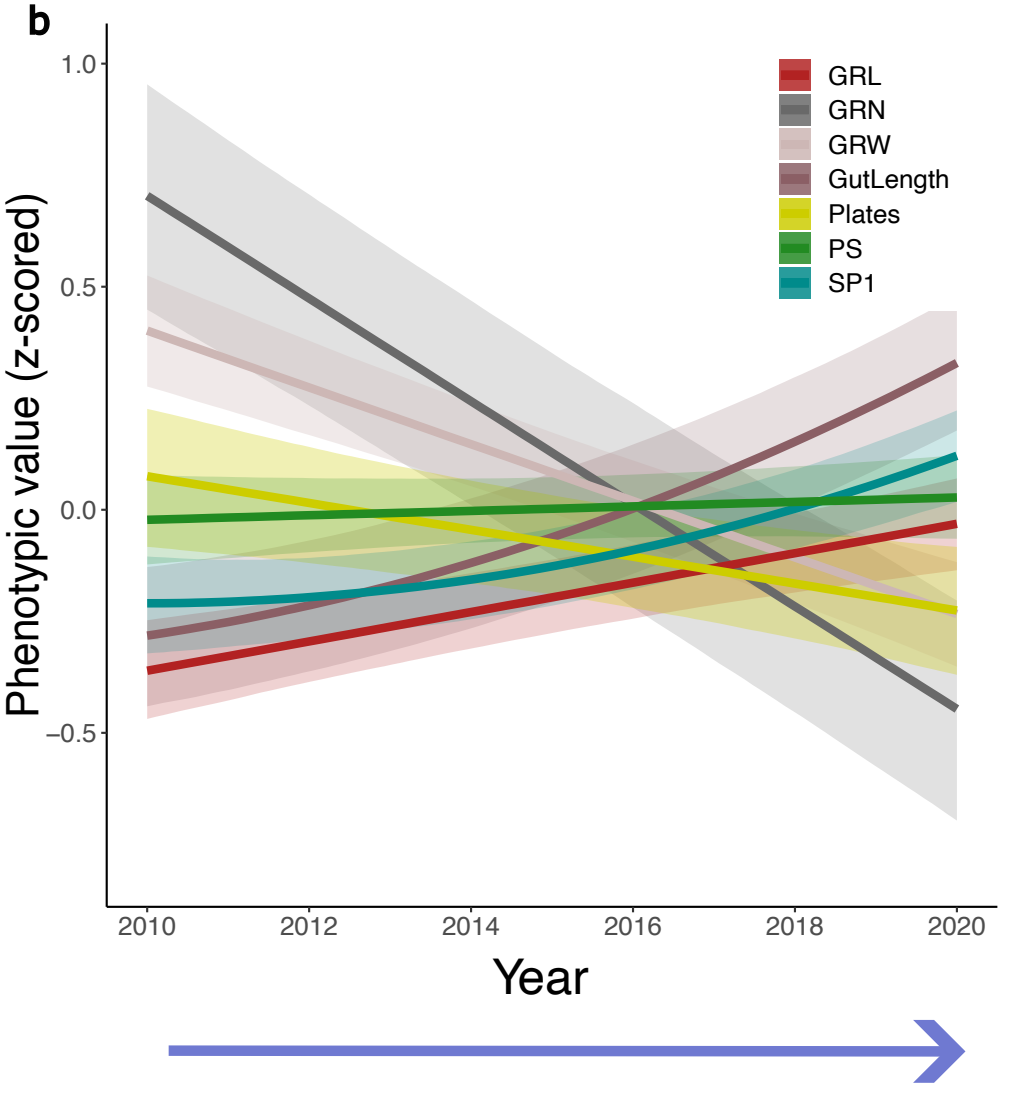
e


Tanytarsus gracilentus
Total abundance (log−scale)
11
9
7
2008
2010
2012
2014
2016
2018
Year

a

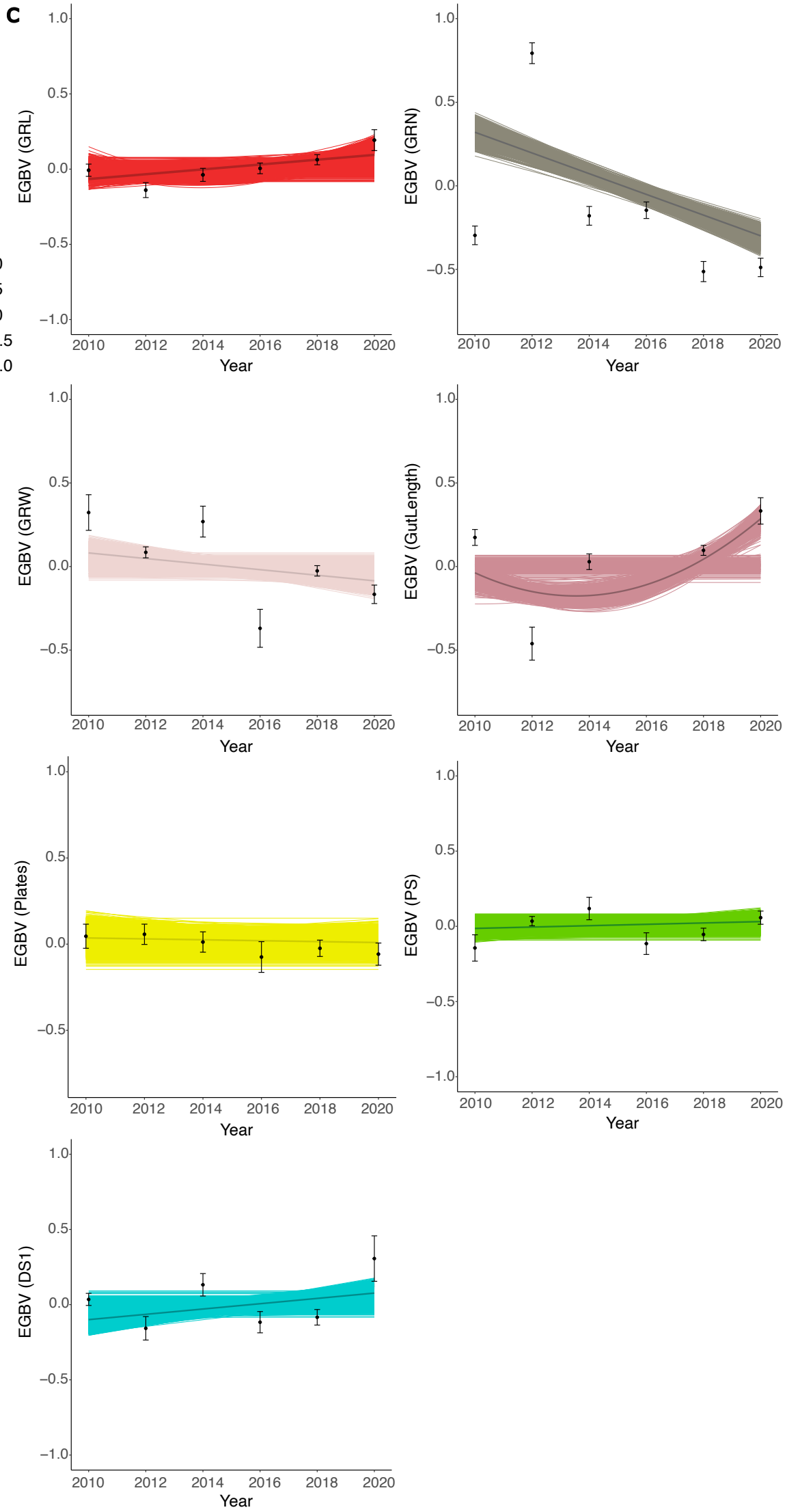


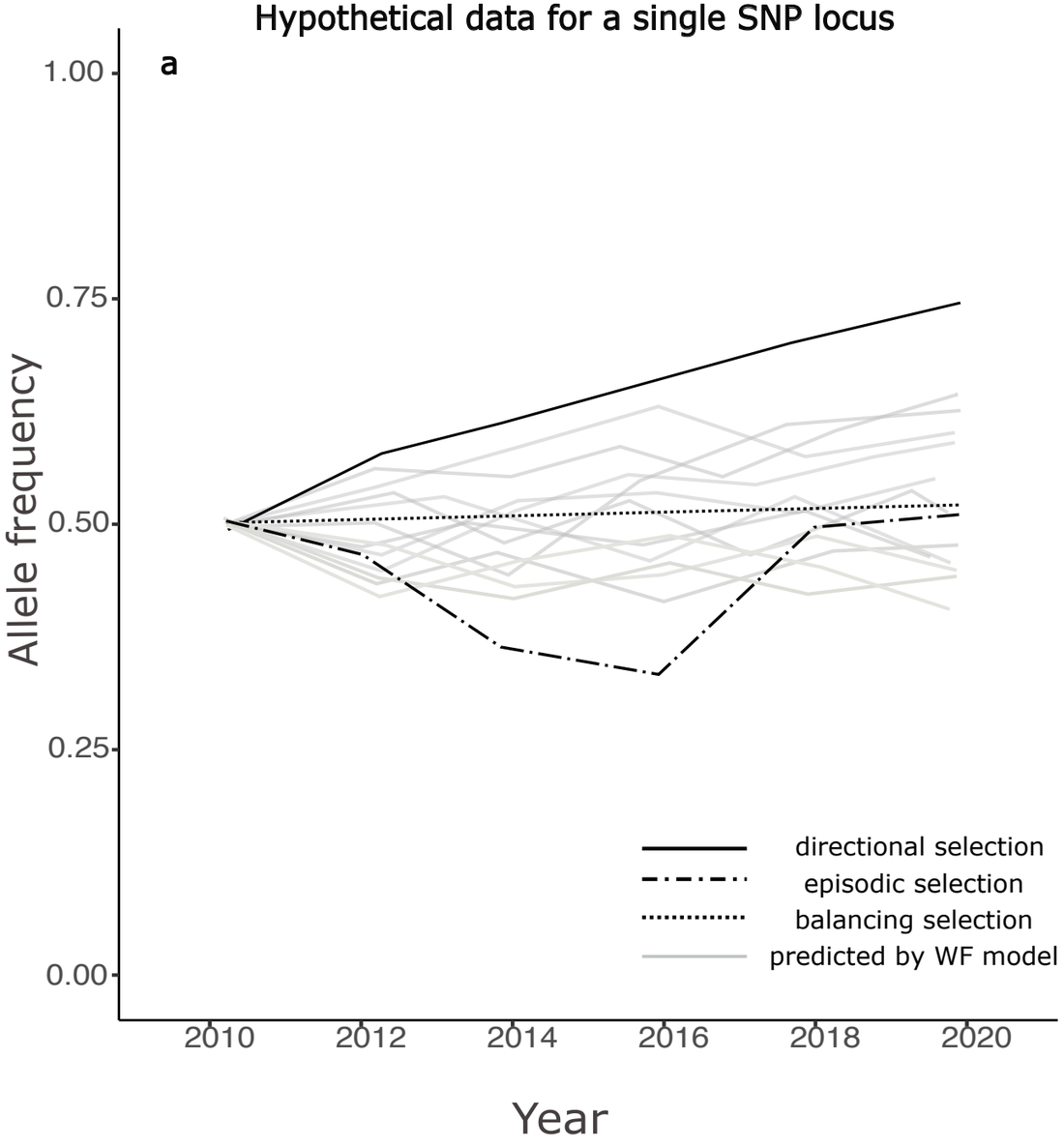
b

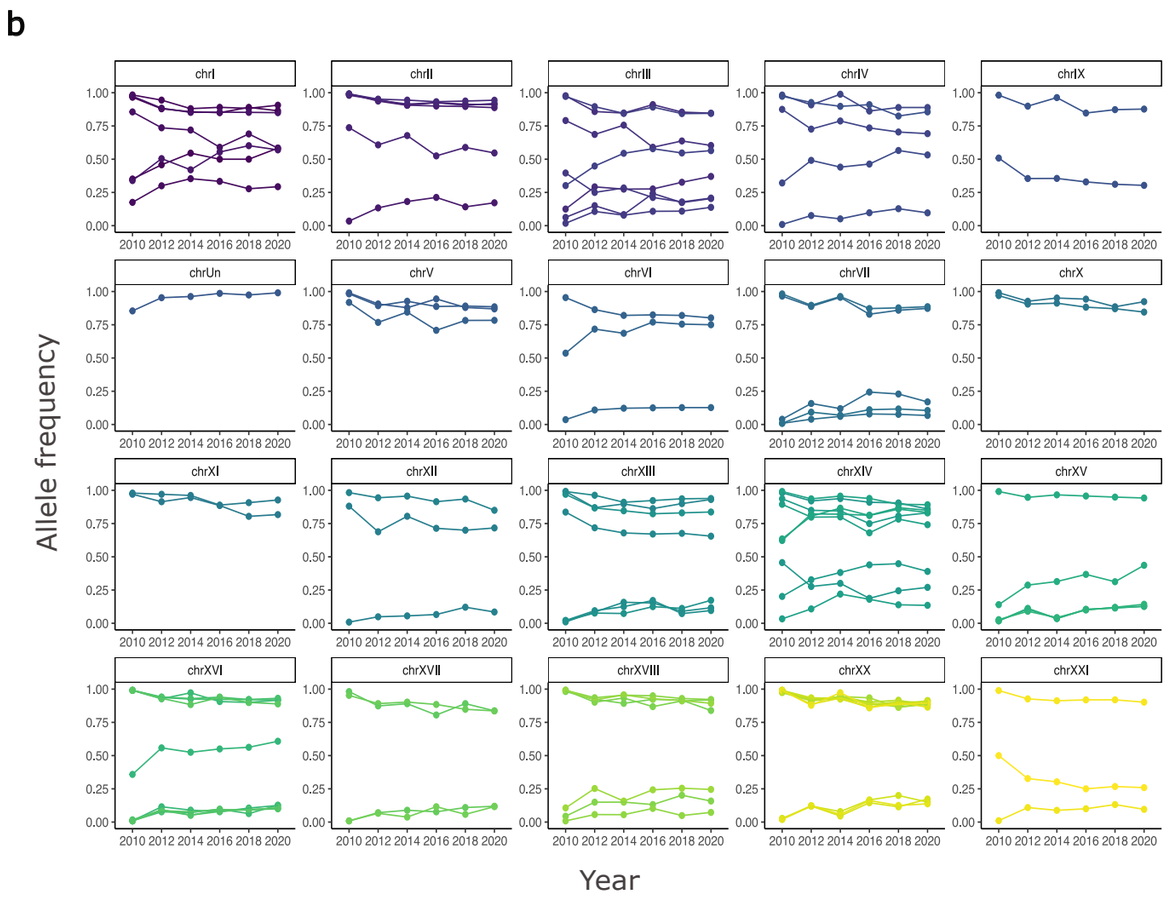


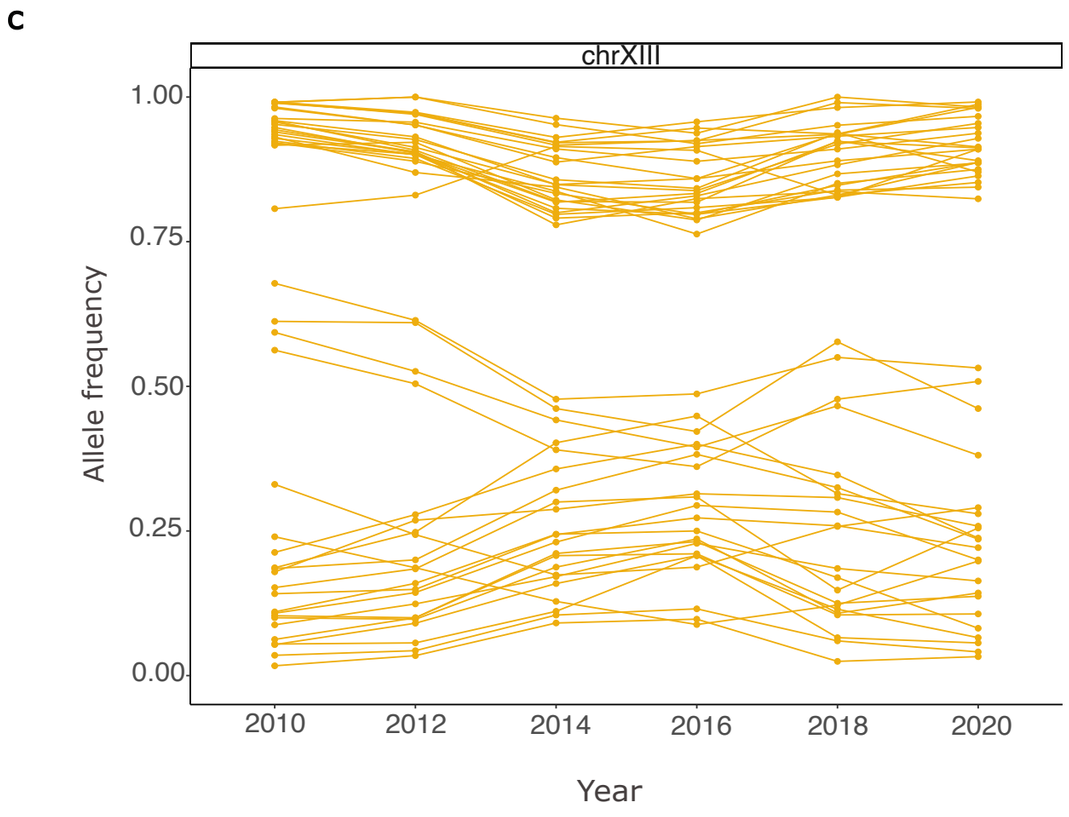
* Longer and fewer gill rakers with narrower gaps
* Longer guts
* Fewer plates and longer spine lengths

c




Hypothetical data for a single SNP locus
a
1.00
0.75
0.50
0.25
0.00
Allele frequency
2010
2012
2014
2016
2018
2020
Year
directional selection
episodic selection
balancing selection
predicted by WF model


b
chrI
chrII
chrIII
chrIV
chrIX
chrUn
chrV
chrVI
chrVII
chrX
chrXI
chrXII
chrXIII
chrXIV
chrXV
chrXVI
chrXVII
chrXVIII
chrXX
chrXXI
1.00
0.75
0.50
0.25
0.00
2010 2012 2014 2016 2018 2020
Allele frequency
Year


c
chrXIII
1.00
0.75
0.50
0.25
0.00
Allele frequency
2010
2012
2014
2016
2018
2020
Year