Genetic population structure constrains local adaptation and probability of parallel evolution in sticklebacks

Petri Kemppainen[1,*], Zitong Li[1,2], Pasi Rastas[1,3], Ari Löytynoja[3], Bohao Fang[1], Baocheng Guo[1,5], Takahito Shikano[1], Jing Yang[1,3], Juha Merilä[1]

[1]Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, FI-00014 University of Helsinki, Finland

[2]CSIRO Agriculture & Food, GPO Box 1600, Canberra, ACT 2601, Australia;

[3]Institute of Biotechnology, FI-00014 University of Helsinki, Finland

[4]The Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

[5]Chinese Sturgeon Research Institute, Three Gorges Corporation, Yichang, 443100, China

*Corresponding author: juha.merila@helsinki.fi

Abstract

Repeated and independent adaptation to specific environmental conditions from a shared ancestral pool of standing genetic variation (parallel evolution) is common. However, if standing genetic variation is limited, local adaptation may be restricted to different and potentially less optimal solutions (convergent evolution), or prevented from happening altogether. Using a quantitative trait locus (QTL) mapping approach, we identified genomic regions responsible for the repeated pelvic reduction (PR) in three crosses between nine-spined stickleback (NSS) populations expressing full and reduced pelvic structures (PS). In one cross, PR mapped to linkage group 7 (LG7) containing the gene *Pitx1* known to cause PR. In the other two crosses, PR was polygenic and attributed to ten novel QTL, of which 90% were unique to specific crosses. Next we screened the genomes of individuals from 27 different NSS populations for deletions in the *Pitx1* regulatory element (*Pel*), which is known to be associated with repeated PR in three-spined sticklebacks (TSS). This revealed deletions only in the population where PR mapped to LG7. Using individual-based forward simulations parametrised with empirical data on NSS and TSS population structures, we show that access to the same ancestral variation promotes parallel evolution (as seen in TSS), whereas lack thereof restricts local adaptation and instead promotes emergence of convergent genetic architectures (as seen in NSS). Hence, the results predict that with an increasing degree of population subdivision, parallel phenotypic evolution becomes increasingly non-parallel at the genetic level and restricts local adaptation to less optimal solutions.

Keywords: convergent evolution, epistasis, local adaptation, pelvic reduction, Pitx1,

42 *Pungitius pungitius*

Introduction

Failure to evolve in response to changing environmental conditions may lead to extirpation or even extinction (Orr and Unckless 2008). If the heritable variation necessary for adapting to environmental change already exists in the form of standing genetic variation, genetic adaptation may proceed swiftly, at least compared to the time it would take for populations to adapt from novel mutations (Orr and Unckless 2008; Barret and Schluter 2008; Thompson *et al*. 2019). Therefore, when exposed to novel yet similar environments, populations derived from the same ancestral population – hence carrying the same pool of alleles – can often be expected to respond to similar selection pressures in a similar fashion. The resulting evolution of parallel genetic adaptations is known as parallel evolution (Arendt and Reznick 2007; Schluter and Conte 2009; Elmer and Meyer 2011; Stern 2013; Conte *et al*. 2015; Bolnick *et al*. 2018; Barghi *et al*. 2019; Hermisson and Pennings 2017). However, in genetically highly structured species, potentially advantageous rare alleles may be lost due to founder events and random genetic drift, preventing adaptation. Alternatively, adaptation to given selection pressures might be more easily gained by allelic substitutions in alternate loci (convergent evolution) influencing the same polygenic trait, even if they may significantly differ in their fitness effects (Cohan 1984; Merilä 2013, 2014; Rosenblum *et al.* 2014). Thus, the demographic history of populations may play a central role in determining the likelihood of parallel evolution.

The three-spined stickleback (*Gasterosteus aculeatus*) is a widely used model system to study adaptive evolution in the wild (Bell and Foster 1994; Gibson 2005). Its

ability to rapidly adapt to local environmental conditions has often been shown to stem from a global pool of ancestral standing genetic variation (Schluter and Conte 2009; Jones *et al*. 2012; Terekhanova *et al.* 2014, 2019). The nine-spined stickleback (*Pungitius pungitius*) has been emerging as another model system for the study of repeated evolution in the wild (Merilä 2013). In general, it differs from the three-spined stickleback by having smaller effective population sizes (N_e), reduced gene flow in the sea, and a tendency to occur in small landlocked ponds in complete isolation from other populations (Shikano *et al.* 2010; DeFaveri *et al*. 2012; Merilä 2013). Given their contrasting population demographic characteristics, three- and nine-spined sticklebacks can thus be expected to occupy different parts of a parallelism-convergence continuum with respect to local adaptation.

Regressive evolution of the pelvic complex has occurred in three (*viz*., *Gasterosteus*, *Pungitius*, and *Culaea*) of the five recognised stickleback genera since the last glacial period (reviewed in Klepaker *et al*. 2013). While marine populations of three- and nine-spined sticklebacks usually have complete pelvic structures with fully developed pelvic girdles and lateral pelvic spines, partial or even complete pelvic reduction is common in freshwater populations (Blouw and Boyd 1992; Shapiro *et al*. 2004, 2006; Herczeg *et al*. 2010; Klepaker *et al*. 2013). Several factors may contribute to this, including the absence of gape-limited predatory fish and limited calcium availability, as well as the presence of certain insect predators (Reist *et al*. 1980; Reimchen *et al*. 1983; Giles 1983; Bell *et al*. 1993; Karhunen *et al*. 2013; Chan *et al*. 2010). Collectively, sticklebacks provide an important model system to study genetic mechanisms underlying

87 the adaptive parallel pelvic reduction at both intra- and inter-specific levels, under a wide
88 range of population demographic settings. However, studies of parallel patterns of
89 marine-freshwater divergence in nine-spined sticklebacks are still scarce (Herczeg *et al.*
90 2010; Wang *et al*. 2020), especially at the genetic level, precluding any comprehensive
91 and conclusive comparison of the two species.

92 The genetic basis of pelvic reduction in the three-spined stickleback is well
93 understood. Quantitative trait locus (QTL) mapping studies have identified a single
94 chromosomal region containing the gene *Pituitary homeobox transcription factor 1*
95 (*Pitx1*) that explains more than two thirds of the variance in pelvic size in crosses
96 between individuals with complete pelvic girdles and spines, and pelvic-reduced
97 individuals (Cresko *et al*. 2004; Shapiro *et al*. 2004; Coyle *et al*. 2007). Pelvic loss in the
98 marine-freshwater three-spined stickleback model system is predominantly caused by
99 expression changes of the *Pitx1* gene due to recurrent deletions of the pelvic enhancer
100 (*Pel*) upstream of *Pitx1* (Chan *et al*. 2010; Xie *et al.* 2019). However, in benthic-limnetic
101 and lake-stream pairs of three-spined sticklebacks, the genetic architecture of pelvic
102 reduction appears to be more diversified (Peichel *et al.* 2001, 2017; Deagle *et al*. 2012;
103 Stuart *et al.* 2017). In the three independent nine-spined stickleback QTL studies
104 available to date, *Pitx1* in linkage group (LG) 7 was also identified as a major cause for
105 the reduction in pelvic structures in one Canadian (Shapiro *et al.* 2006) and one Finnish
106 (Shikano *et al.* 2013) population. Another large effect region in LG4 was found to be
107 associated with pelvic reduction in an Alaskan population (Shapiro *et al.* 2009). Similar
108 to the three-spined stickleback, pelvic spine and pelvic girdle sizes are strongly correlated

in the population from the Finnish pond Rytilampi, since *Pitx1* controls both phenotypes (Shikano *et al.* 2013; Fig. 1 and Supplementary Table 1). In contrast, a considerable amount of phenotypic variation with respect to these traits and their inter-correlations exists among different freshwater pond populations in northern Europe (Herczeg *et al.* 2010; Karhunen *et al.* 2013, 2014; Fig. 1). Given their high heritability (Blouw and Boyd 1992; Leinonen *et al.* 2011), the lack of correlation between spine and girdle lengths suggests that they can be independently controlled by different QTL. Thus, the genetic underpinnings of pelvic reduction in nine-spined sticklebacks could be more diversified than those in the marine-freshwater three-spined stickleback model system (Merilä 2013, 2014).

To investigate the possible genetic heterogeneity underlying pelvic reduction in different nine-spined stickleback populations, we analysed data from three F_2 generation inter-population crosses (279–308 progeny per cross) between pond and marine individuals to identify QTL responsible for pelvic reduction in different ponds. One of these populations was the previously studied Rytilampi cross (earlier analysed only with microsatellites, Shikano *et al.* 2013), now re-analysed along with two new populations (Bynästjärnen and Pyöreälampi) using > 75 000 SNPs. This data was subjected to a cutting-edge mapping approach (Li Z. *et al.* 2017, 2018) that can provide more information on the source of the QTL effects than has been previously possible. In addition, to assess the extent to which *Pel* is responsible for pelvic reduction in the nine-spined stickleback, we screened the whole genomes of individuals from 27 wild populations for deletions in the genomic region spanning the *Pel* element and the *Pitx1*

131 gene. Finally, we used empirical data to estimate realistic levels of population structuring
132 and genetic isolation by distance (IBD) in marine populations of three- and nine-spined
133 sticklebacks, and then used individual-based forward simulations to test how these factors
134 affect the prevalence of genetic parallelism when local adaptation proceeds from standing
135 genetic variation.
136
137 **Materials and Methods**
138 *Fish collection, crossing, and rearing*
139 For the QTL crosses, three different marine F_0 generation females from the Baltic Sea
140 (Helsinki, Finland; 60°13′N, 25°11′E) were crossed with a freshwater F_0 generation male
141 from either Bynästjärnen (64°27′N, 19°26′E), Pyöreälampi (66°15′N, 29°26′E) or
142 Rytilampi (66°23′N, 29°19′E) ponds. Fish crossing, rearing, and sampling followed the
143 experimental protocol used in the earlier study of the Rytilampi population (Shikano *et*
144 *al.* 2013; Laine *et al.* 2013). For Rytilampi, the F_0 generation fish were artificially mated
145 in the lab during the early breeding season of 2006 (Shikano *et al.* 2013), and the
146 resulting full-sib F_1-offspring were group-reared in aquaria until mature, as explained in
147 Shikano *et al.* (2013). Two randomly chosen F_1 individuals were mated repeatedly (seven
148 different clutches) to produce the F_2 generation between September and October 2008.
149 The offspring were reared in separate aquaria. The same procedure was followed for
150 Pyöreälampi (F_0 mating: Jun 2011; F_1 mating: Jul–Sep 2012; F_2 rearing: Jul 2012–Apr
151 2013; 8 different clutches) and Bynästjärnen (F_0 mating: Jun 2011; F_1 mating: Nov 2013–
152 Jan 2014, F_2 rearing: Nov 2013–Aug 2014; 6 different clutches). The F_2 fish were

153 euthanized at 187, 238 and 238 days post-hatch for Rytilampi, Pyöreälampi and
154 Bynästjärnen, respectively. At this point, the fish were on average 52.3 mm in standard
155 length (Rytilampi = 48.6 mm; Pyöreälampi = 52.3 mm; Bynästjärnen = 53.6 mm), and
156 weighed on average 1.34 g (wet weight; Rytilampi = 1.07 g; Pyöreälampi = 1.49 g;
157 Bynästjärnen = 1.44 g). In total, 308, 283 and 279 F_2 offspring were available for
158 analyses from Helsinki × Bynästjärnen (HEL × BYN), Helsinki × Pyöreälampi (HEL ×
159 PYÖ) and Helsinki × Rytilampi (HEL × RYT) crosses, respectively.
160 The experiments were conducted under licenses from the Finnish National
161 Animal Experiment Board (#STH379A and #STH223A). Experimental protocols were
162 approved by the ethics committee of the University of Helsinki, and all experiments were
163 performed in accordance with relevant guidelines and regulations.
164
165 *Morphological data*
166 To visualise bony elements, all of the F_2-progeny were stained with Alizarin Red S
167 following Pritchard and Schluter (2001). Pelvic spine and girdle lengths from both sides
168 of the body, as well as standard body length, were measured with digital calipers to the
169 nearest 0.01 mm. Although it is known that the left-right asymmetry of the pelvic girdle
170 is heritable in sticklebacks (Blouw and Boyd 1992; Bell *et al.* 2007; Coyle *et al.* 2007),
171 we did not specifically study this. To reduce the number of tests, the mean of the left and
172 the right side measurements was used (analyses conducted for the two sides separately
173 always yielded qualitatively similar results as the tests conducted with the averages;
174 results not shown). All measurements were made by the same person twice; the

175 repeatability (R; Becker 1984) ranged between 0.80 and 0.84 for girdle lengths, and
176 between 0.98 and 0.99 for spine lengths. The QTL analyses were performed on both
177 absolute and relative (scaled to the total body length) trait values, but for all of the
178 analyses that compared phenotypic data between populations (which also differ in body
179 sizes), only relative trait values were used. Previously published phenotypic data from 19
180 wild populations were obtained from Herczeg *et al*. (2010) and Karhunen *et al.* (2013).
181 These included data on pelvic spine and girdle lengths of wild-collected individuals from
182 ten pond populations (Abbortjärn = ABB, Bolotjone = BOL, Karilampi = KAR,
183 Kirkasvetinen lampi = KRK, Mashinnoje = MAS, Lil-Navartjärn = NAV, Hansmytjärn =
184 HAN, Rytilampi = RYT, Bynästjärnen = BYN), four lake populations (Iso Porontima =
185 POR, Riikojärvi = RII, Joortilojärvi = JOR, Västre-Skavträsket = SKA) and five marine
186 populations (Fiskebäckskil = FIS, Trelleborg = TRE, Bölesviken = BÖL, Helsinki = HEL
187 LEV = Levin Navolak), as well data on common garden-reared F_1 generation individuals
188 from two marine (HEL & LEV) and two pond (BYN & PYÖ) populations. Visibly
189 broken spines were treated as missing data. A map showing the geographic location of
190 these populations is provided in Supplementary Fig. 1.
191
192 *Genotyping and linkage map construction*
193 The RAD sequencing protocol used to obtain the SNP data was the same as in Yang *et al.*
194 (2016) and Li Z. *et al*. (2017). In short, genomic DNA was extracted from ethanol
195 preserved fin clips using the phenol-chloroform method (Taggart *et al*. 1992). DNA was
196 fragmented with PstI restriction enzyme and the resulting 300 to 500 bp long fragments

were gel purified. Illumina sequencing adaptors and library specific barcodes were ligated to the digested DNA fragments, and the barcoded RAD samples were pooled and sequenced on 24 lanes of the Illumina HiSeq2000 platform with 45 bp single-end strategy. RAD library construction and sequencing were conducted by BGI HONGKONG CO., LIMITED. After eliminating adapters and barcodes from reads, a quality check was done using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/).

Linkage mapping for the three crosses was conducted using Lep-MAP3 (LM3; Rastas 2017), as described in detail in Li H. *et al.* (2018). LM3 can infer the parental/grandparental phase based on dense SNP data, which allowed us to utilise the four-way cross QTL mapping method described below. Input data were generated by first mapping individual fastq files to the nine-spined stickleback reference genome using BWA mem (Li H. 2013) and SAMtools (mpileup; Li H. *et al.* 2009), followed by pileupParser.awk and pileup2posterior.awk scripts from the LM3 pipeline using default parameters.

The mapping was carried out following the basic LM3 pipeline and by combining the three crosses: parental genotypes were called by taking into account genotype information on offspring, parents and grandparents (module ParentCall2). Markers segregating more distortedly than what would be expected by chance (1:1,000 odds) were filtered out (module Filtering2). Loci were then partitioned into chromosomes (modules SeparateChromosomes2 and JoinSingles2) yielding 21 linkage groups with > 75 000 markers assigned to these groups. Finally, the markers were ordered within each linkage

group with the module OrderMarkers2, removing markers that were only informative in either the mother or father, respectively. This created two maps for each chromosome, one having more maternal markers and the other having more paternal markers, with an average of ⅔ of the markers shared between the two. Constructing two maps this way removes the effect of markers that are informative only in one parent, as markers informative in different parents are not informative when compared against each other. The phases were converted into grandparental phase by first evaluating the final marker orders and then matching the (parental) phased data with the grandparental data, inverting the parental phases when necessary. Orphan markers from map-ends based on scatter plots of physical and map positions were manually removed.

Dimensionality reduction by linkage disequilibrium network clustering

In QTL mapping, it is essential to correct for multiple testing in order to reduce the rate of false positives. Moreover, in large genomic datasets, physically adjacent SNPs – particularly those from experimental crosses – are often in linkage disequilibrium (LD), i.e. correlated. Since a group of SNPs in high LD explains similar amounts of genetic variation in a given trait, it is reasonable to apply a dimensional reduction procedure before QTL-mapping to exclude the redundant information from the data. Here we used linkage disequilibrium (LD) network clustering ("LDn-clustering") and PC regression as dimensionality reduction tools prior to single-locus QTL mapping (Li Z. *et al.* 2018). The first step of this approach involves an extension of a method developed by Kemppainen *et al.* (2015), which uses LD network analysis for grouping loci connected by high LD.

The second step involves principal component analysis (PCA) as a method for dimensionality reduction in each cluster of loci connected by high LD ("LD-clusters"). This was achieved by the function "LDnClustering" in R-package "LDna" (v.0.64; Li Z. *et al*. 2018). The method used here differs slightly from the original method described in Li Z. *et al.* (2018), and was implemented as follows. Initially all edges (representing LD-values as estimated by r^2 using the function "snpgdsLDMat" from the R-package "SNPRelate", Zheng *et al.* 2012) below the LD threshold value of 0.7 were removed. This resulted in many sub-clusters in which all loci were connected by at least a single edge. Second, for each sub-cluster, a complete linkage clustering was performed, where a cluster is defined by its smallest link. Starting from the root, additional sub-clusters were found recursively, where the median LD between all loci was > 0.9. This recursive step was not used in the original implementation, but was later found to increase computational speed and reduce the number of single locus clusters, i.e. result in more efficient complexity reduction. To facilitate computational speed, we only considered SNPs no further than 2000 SNPs away from each other (rather than considering all pairwise values within a chromosome at a time; parameter w2 = 2000), but nevertheless analysed all SNPs from each cluster at a time (rather than performing LD clustering in windows). Thus, each resulting LD-cluster represents sets of physically adjacent and highly correlated loci. We performed PCA regression based on each LD-cluster in which individuals are separated according to their LD-cluster multi-locus genotypes. Since the loci are highly correlated, most of the genetic variation (>90%) between individuals is explained by the first principal component. For each LD-cluster, we then replaced all the

SNP genotypes by the position of the individuals along the first PC axis after removing the sex-linked loci (Pearson's correlation coefficient >0.95 between the PC coordinates and sex). The input data for the LDn-clustering was comprised of the original co-dominant SNP data with individuals from all three crosses pooled. A custom R-code used for this dimensionality reduction is available from DRYAD (xxx).

QTL mapping: four-way crosses model

In some circumstances, such as a four-way cross (Xu 1996), F_1 hybrids of two heterozygous parents (Van Ooijen 2009), and an outbred F_2 design (Xu 2013a), it is possible that up to four different alleles – A and B from the dam, and D and C from the sire – segregate in the population. In such cases, a linear regression model for QTL analysis of the outbred F_2 data (Li Z. *et al.* 2018) is defined by:

$$y_i = \beta_0 + x_{dij}\beta_{dj} + x_{sij}\beta_{sj} + z_{ij}\gamma_j + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2), \quad (1)$$

where y_i is the phenotype value of individual i (i=1,…,n), x_{dij}, x_{sij}, z_{ij} are genotypes coded as

$$\begin{vmatrix} +1 & +1 & +1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & -1 & +1 \end{vmatrix} \begin{matrix} for genotype AC, \\ for AD, \\ for BC, \\ for BD. \end{matrix}, \quad (2)$$

(Xu 2013b), β_0 is the parameter of the population mean, β_{dj} is the substitution effect of alleles A and B of the dam (i.e. the grandfather in F_0) at the locus j (j=1,…,p; p is the total number of SNPs), β_{dj} is the substitution effect of alleles C and D of the sire (i.e. the grandmother in F_0), γ_j is the dominance effect, and ε_i is the residual error term mutually

following an independent normal distribution.

The model (1) requires the knowledge of the grandparental phase (produced by LM3) with the benefit that the source (*viz*. F_0 female or F_0 male) of the observed QTL effect can be inferred, as described in more detail in Supplementary File 1. A multiple correction on the basis of permutation tests was further conducted to control for false positives due to multiplicity (Li Z. *et al.* 2017) with 10 000 replicates.

Estimating the proportion of phenotypic variance explained by SNPs

The overall proportion of phenotypic variance (PVE) explained jointly by all SNPs (an approximation of the narrow sense heritability) was obtained using LASSO regression (Tibshirani 1996), which incorporates all the SNPs into a multi-locus model:

$$\frac{1}{2n}\sum_{i=l}^{n}\left(y_i-\beta_0-x_{dij}\beta_{dj}-x_{sij}\beta_{sj}-z_{ij}\gamma_j\right)+\lambda\sum_{j=l}^{p}\left(\left|\beta_{dj}\right|+\left|\beta_{sj}\right|+\left|\gamma_j\right|\right), \quad (3)$$

where the l_1 penalty term $\lambda\sum_{j=l}^{p}\left(\left|\beta_{dj}\right|+\left|\beta_{sj}\right|+\left|\gamma_j\right|\right)(\lambda>0)$ shrinks the regression parameters towards zero; all other symbols are defined in the same way as in Equation (1).

Following Sillanpää (2011), the PVE can be estimated by the formula:

$$PVE_{total} = \frac{var\left(\sum_{j=l}^{p} x_j \beta_j\right)}{var(y)} \approx \frac{\mathrm{var(y)} - \sigma_0^2}{var(y)}, (4)$$

where β_j is the effects of the SNPs, and σ_0^2 is the LASSO residual variance estimated by a cross-validation-based estimator introduced by Reid *et al*. (2016). The PVE explained by each linkage group was estimated on the basis of LASSO estimates using the following formula:

$$PVE_{LG} \approx PVE_{total} - \frac{var\left(\sum x_j \hat{\beta}_j\right)}{var(y)}, (5)$$

where $\hat{\beta}_j (j \notin G)$ represents all the effects estimated by the LASSO of the SNPs which do not belong to a set of SNPs (e.g. to a chromosome/linkage group). A similar approach was used to estimate the contribution of grandparental alleles and to evaluate the dominance component by treating the coding systems [x_{dij}, x_{sij}, z_{ij}] (2) as different groups of SNPs. A custom R-code for PVE estimation is available from DRYAD (XXX).

Scanning for Pel deletions in full-genome sequence data

A minimum of 20 samples from populations RYT, MAS, BOL, BYN and PYÖ, and 10–31 samples from an additional 22 populations (from Northern Europe and USA; Supplementary Fig. 1 and Supplementary Table 2) were sequenced to 10× coverage by BGI HONGKONG CO., LIMITED. Reads were mapped to the nine-spined stickleback reference genome (Varadharajan *et al*. 2020) with BWA mem (Li H. 2013), and site-wise sequencing coverages were computed with SAMtools (depth; Li H. *et al*. 2009). Relative sequence depths across the *H2afy-Pitx1* intergenic region were estimated for the five

322 focal populations by first computing the median depths for 1000 bp sliding windows, and
323 then normalising these by the median depth for the full intergenic region. The Pel-
324 2.5kb^{SALR} region was extracted from the original BAC contig (GenBank accession
325 number 'GU130433.1') and mapped to the nine-spined stickleback intergenic region with
326 minimap2 (Li H. 2018). The sequencing depths for the *Pel*-2.5kb^{SALR} were normalised by
327 dividing the mean depths of the *Pel* region by the mean depths of the full intergenic
328 region. Gene annotations and relative sequencing depths (average and confidence
329 intervals) were computed and visualised using R-package "Gviz" (Hahne and Ivanek
330 2016). Lastly, we scanned the literature for genes that are known to affect pelvic and hind
331 limb development and searched for potential matches in the nine-spined stickleback
332 genome (Varadharajan *et al*. 2020) in regions where significant QTL were found. The
333 custom R-code for these analyses is provided in DRYAD (XXX).
334
335 *Proof of concept simulations*
336 To obtain empirical data to inform our simulations, we analysed a subset of 4326 SNPs
337 from 236 nine-spined and 98 three-spined sticklebacks from 54 Fennoscandian
338 populations. This data was extracted from previously available RAD-seq and full genome
339 sequence data as described in Supplementary File 2. The purpose of this was to provide a
340 ballpark estimate of IBD and degree of freshwater-freshwater genetic differentiation for
341 the simulations. For IBD analyses, the pairwise genetic distance between populations
342 were regressed against the geographic distance separating the populations. We estimated
343 geographic distance in the World Geodetic System 84 reference ellipsoid (i.e. point

distance when taking into account the curvature of the planet) using the function "pointDistance" from the R-package "raster" (Hijmans and van Etten 2012). Genetic distances were calculated as $F_{ST}/(1-F_{ST})$ (linearised F_{ST}), with F_{ST} estimated as shown above for all pairwise population comparisons. Slope and confidence intervals of the regression lines were estimated by bootstrapping SNPs (1000 bootstrap replicates). As the geographic distances between the marine populations in the simulations were on an arbitrary scale (and uniform between all adjacent marine populations), we scaled the geographic distance in the simulated data such that the slope for a regression line (mean from bootstrapping) between linearised F_{ST} and geographic distance for the empirical and simulated data would be the same.

The repeated local adaptation in independently colonised freshwater stickleback populations is widely considered to be due to selection on standing genetic variation available in the sea, which in turn is maintained by recurrent gene flow from previously colonised freshwater populations (the "transporter hypothesis"; Schluter and Conte 2009). We used individual-based forward simulations (QuantiNemo2; Neuenschwander *et al.* 2018) to investigate how our estimated differences in population structure between nine- and three-spined sticklebacks under the transporter hypothesis influence the prevalence of parallel evolution, following post-glacial colonisation of freshwater habitats from the sea. For both species, the sea populations were simulated by a stepping-stone model comprised of ten sub-populations with carrying capacity, $K = 1000$ each (under mutation-drift equilibrium $K = N_e$). All adjacent marine populations exchanged 100 (TSS) or ten (NSS) migrants per generation (M, symmetrical gene flow) with some

366 long-distance migration also allowed between every second population at a rate of ten
367 times less than that between adjacent populations (Supplementary Fig. 3). To allow
368 frequencies of freshwater adapted alleles to build up and filter to the marine populations
369 as standing genetic variation, a refuge freshwater population (K = 10000) with high
370 frequencies of freshwater adapted alleles was allowed to exchange migrants with the two
371 most central of the ten marine populations (symmetric gene flow with rate M = 1;
372 Supplementary Fig. 3). After a burn-in, two focal freshwater populations were founded
373 from the two marine populations situated at the opposite ends of the stepping-stone chain,
374 after which symmetrical gene flow was allowed between the focal freshwater populations
375 and their closest marine populations at rates M = 1 for TSS and M = 0.2 for NSS
376 (Supplementary Fig. 3). Thus, the freshwater populations were founded from different
377 marine populations with ancestral freshwater adapted alleles stemming from the same
378 refuge freshwater population situated equally far from the two focal freshwater
379 populations, in agreement with the transporter hypothesis. The simulations were then run
380 for 5000 generations to simulate post glacial colonisation of freshwater habitats (10000
381 years ago, assuming a generation time of two; Baker 1994; DeFaveri and Merilä 2013;
382 DeFaveri *et al*. 2014). The above parameters generated patterns of IBD and population
383 structuring similar to what was estimated from the empirical data (see Results) and are
384 also consistent with previous studies (DeFaveri *et al*. 2014).
385 Three different genetic architectures of a single trait coded by five (independent)
386 bi-allelic loci under stabilising selection for different optima were simulated in the marine
387 (optimal phenotypic value = 0) and freshwater populations (optimal phenotypic value =

388 20). In architecture A, one large effect additive locus with the homozygote for allele 1
389 ("a1") yielding a genotypic value of zero (locally adapted to marine habitat) was
390 simulated; the same was done for the homozygote for allele 2 ("a2") yielding a genotypic
391 value of 20 (locally adapted for the freshwater habitat). The four remaining minor effect
392 additive QTL were given the allelic value of zero for a1 and allelic values [1,1,2,3] for
393 a2. An empirical example of this kind of architecture is the *Ectodysplasin* (*EDA*) gene in
394 the three-spined stickleback (Colosimo *et al*. 2005). In architecture B, the allelic values of
395 a2 alleles were [1,2,3,4,6] and the allelic values of the alternative alleles were zero, i.e. a
396 single large effect locus is lacking. In architecture C, the distribution of allelic values was
397 the same as in architecture A, but with a2 being recessive to a1. An empirical example of
398 this is the *Pitx1* locus controlling pelvic reduction in three- and nine-spined sticklebacks
399 (Cresko *et al*. 2004; Shapiro *et al*. 2004).

400 We simulated ten chromosomes of 100 centi Morgan (cM) with 100 neutral loci
401 (and the QTL) equally spaced across the genome (i.e. uniform recombination rate across
402 chromosomes, with no tight linkage between any two loci). Mutation rate (μ) was set to
403 10^{-8} for all loci, and simulations were initiated with allele frequencies for all loci set to
404 0.5. At the end of the simulations (generation 15000) we estimated allele frequency
405 differences as the fixation index, F_{ST} (Weir and Cocherham 1984), with the "stamppFst"
406 function in the R-package "StAMPP" (Pembleton *et al*. 2013). Further details of
407 parameter settings and the simulation code can be found in DRYAD (XXX).

408 We classified loci being involved in parallel evolution if the F_{ST} between the focal
409 freshwater populations (pooled) and all marine populations (pooled) was higher than 0.5

(marine-freshwater F_{ST}) and lower than 0.5 between the two focal freshwater populations (freshwater-freshwater F_{ST}). Any loci with both high marine-freshwater F_{ST} and freshwater-freshwater F_{ST} (above 0.5) were classified as being locally adapted, but not involved in parallel evolution.

Results

Heterogeneity of pelvic reduction in the wild

Re-analysis of previously published phenotypic data from the wild confirmed a high degree of variation among different populations with respect to pelvic spine and girdle lengths and their inter-correlations (Fig. 1, Supplementary Table 1 and Supplementary Fig. 2). For instance, while spines were absent and pelvic girdles strongly reduced (but not completely absent) in RYT, spines and girdles were completely lacking in the MAS population (Fig. 1 and Supplementary Table 1). Furthermore, in the BYN population spines were absent but pelvic girdles were only partially reduced, whereas in BOL (a population adjacent to MAS), large variation in both spine (SD = 0.037) and girdle (SD = 0.047) lengths was observed, although these two traits were strongly correlated ($r^2 = 0.61$, Fig. 1, Supplementary Table 1 and Supplementary Fig. 2). This suggests that a large effect locus affecting both pelvic spines and girdles segregated in this population at the time of sampling. In six pond populations (*viz*. ABB, KAR, KRK, NAV, HAN, and PYÖ; Fig. 1), relative spine (mean = 0.079) and girdle (mean = 0.15) lengths were only slightly smaller relative to the marine populations (0.11 and 0.16 for spine and girdle lengths, respectively; Supplementary Table 1). Lack of pelvic reduction (relative to marine

populations) was observed only in one pond population (KAR; Fig. 1, Supplementary Table 1 and Supplementary Fig. 2).

QTL mapping of pelvic reduction in Helsinki × Rytilampi cross

After LD-network based complexity reduction, all QTL mapping analyses were performed on 241 PCs (six sex-linked PCs were removed). Re-analyses of the 279 F_2 progeny from the HEL × RYT cross confirmed a single QTL region on LG7 for both pelvic spine and girdle lengths in the single-locus analyses (Fig. 2a, b and Table 1). In the multi-locus analyses (absolute trait values), LG7 explained 74–86% of the PVE for both spine and girdle lengths, while all other chromosomes individually explained less than 3% of the phenotypic variance (Table 2). An approximately equal amount of the phenotypic variance was explained by alleles inherited from the F_0 male (pond individual) and the F_0 female (marine individual; ~30% of the total variance for all traits; Table 1), respectively, with 15–21% of the phenotypic variance also explained by dominance effects. This is the expected outcome for a recessive QTL when the F_0 individuals are fixed for different large effect causal alleles, and when no additional smaller effect loci affect the trait (Klug and Cummings 2018).

QTL mapping of pelvic reduction in Helsinki × Bynästjärnen cross

Among the 308 F_2 progeny of the HEL × BYN cross, single-locus mapping analyses on pelvic spine lengths detected two significant QTL on LG15 (PVE = 8.9%) and LG16 (PVE = 13.7%; Fig. 2c and Table 1) for alleles deriving from the F_0 male (pond

454 individual). Thus, the causal alleles for these QTL segregated in the F_0 pond male (as
455 explained in Supplementary File 1). No dominance effects were detected for these QTL,
456 suggesting that the allelic effects were additive. One QTL on LG6 with an allelic effect
457 deriving only from the F_0 female was also detected (Fig. 2c and Table 1). The QTL
458 significance profiles, in particular for LG15 and LG16, spanned large genomic regions
459 with no obvious peaks (in contrast to LG7 in the HEL × RYT cross; Fig. 2). This
460 remained true when analysing all SNPs individually (fine mapping; Supplementary Fig.
461 4). The individual and multi-locus phenotypic effects on spine lengths for the QTL on
462 LGs 6, 15 and 16 (using the most significant QTL for each QTL region) are detailed in
463 Figure 3. In the absence of any large effect loci, and when all QTL are additive and
464 independent (as is the case here), phenotypes in the F_2 generation are expected to be
465 normally distributed (Klug and Cummings 2018). For some multi-locus genotypes (of the
466 QTL on LGs 6, 15 and 16), the distribution of spine lengths were approximately normally
467 distributed, except with long tails of highly reduced spine lengths, implicating that some
468 of these loci could be involved in epistatic interactions (this was investigated further in
469 Supplementary File 3). One significant male QTL on LG4 for girdle length was also
470 detected (Fig. 2d and Table 1). Results for relative trait values were highly similar to the
471 absolute trait values, except for an additional significant male and female QTL on LG1,
472 as well as another significant female QTL on LG16 (Supplementary Fig. 5, Table 1 and
473 Supplementary Table 3).

474 Multi-locus analyses (for absolute trait values) identified 11 LGs that accounted
475 for at least 3% of the phenotypic variance in pelvic spine or girdle lengths in the HEL ×

476 BYN cross (Table 2). The largest of these effects were found on LG15 and LG16, which
477 explained 9% and 12% of variation in pelvic spine lengths, respectively (Table 2). For
478 pelvic spine and girdle lengths, 39% and 32% of the PVE, respectively, were accounted
479 for by all SNPs in the dataset, which equates to the narrow sense heritability (h^2) also
480 accounting for dominance (but not epistatic interaction) effects. For spine lengths, 24% of
481 the total PVE was attributed to alleles deriving from the F_0 male and 17% was attributed
482 to alleles deriving from the F_0 female; only 1% was attributed to the dominance effect
483 (Table 2). For girdle lengths, the corresponding numbers were 16%, 6% and 13% (Table
484 2).
485
486 *QTL mapping of pelvic reduction in Helsinki × Pyöreälampi cross*
487 In the HEL × PYÖ cross (283 F_2 individuals), one QTL for spine length was found on
488 LG9, which was explained by alleles inherited from the F_0 male. Two significant QTL for
489 girdle length were found, one on LG19 (explained by alleles inherited from the F_0 male)
490 and the other on LG4 (explained by alleles inherited from the F_0 female; Fig. 2e, f, and
491 Table 1). In multi-locus analyses, the PVE for different LGs mirrored these results; the
492 LGs that contain significant QTL explain most of the PVE, while PVE for all other LGs
493 were <4 % (Table 2). In multi-locus analyses, PVE for pelvic traits was lower than in the
494 HEL × BYN cross; 14% and 16% for spine length and girdle length, respectively, with
495 <2% PVE due to dominance effects (Table 2). When analysing relative trait values, one
496 additional female QTL peak was found for both girdle and spine lengths on LG1
497 (Supplementary Fig. 5, Table 1 and Supplementary Table 3). Due to the large size of the

498 identified QTL regions, it is not possible to know whether this QTL and that on LG1
499 detected in the HEL × BYN cross are due to the same or different underlying causal
500 mutations (we consider this as a single QTL region).
501
502 *Trait correlations in the QTL crosses*
503 There was a strong correlation between relative pelvic spine and girdle lengths ($r^2 = 0.85$,
504 Fig. 1, Supplementary Fig. 2 and Supplementary Table 1) in the HEL × RYT cross as
505 expected, since pelvic reduction in this cross is controlled by a single large effect QTL
506 affecting both traits. However, in the HEL × BYN cross, correlation between relative
507 pelvic spine and girdle lengths was much weaker ($r^2 = 0.11$, Fig. 1, Supplementary Fig. 2
508 and Supplementary Table 1). This finding is consistent with pelvic spine and girdle
509 reductions being independently controlled by different QTL. Furthermore, of the 306 F_2
510 individuals, only four displayed complete lack of spines, despite the fact that the BYN
511 population is fixed for complete spine reduction in the wild (and the spine was absent in
512 the F_0 male). This is consistent with spine length being controlled by multiple additive
513 loci in the HEL × BYN cross. Among the F_2 individuals from the HEL × BYN cross,
514 relative girdle lengths were normally distributed with much smaller variances (SD =
515 0.012) compared to the HEL × RYT (SD = 0.041) with only two individuals with reduced
516 girdles (Fig. 1, Supplementary Fig. 2 and Supplementary Table 1). This is consistent with
517 the lower heritability of pelvic girdle lengths in the HEL × BYN than in the HEL × RYT
518 cross, with contributions from many small effect loci. In the HEL × PYÖ cross,
519 phenotypic variation was comparable to that in the HEL × BYN cross (SD = 0.012),

520 although the mean relative spine length was slightly smaller (0.078 *vs* 0.091;
521 Supplementary Table 1) and the mean for relative girdle length was slightly larger (0.175
522 *vs* 0.169; Supplementary Fig. 2; Supplementary Table 1).
523

524 *Scanning for Pel deletions in the full-genome sequence data*

525 In the full-genome re-sequencing data of wild collected individuals, a large deletion
526 upstream of *Pitx1* was fixed for all 21 individuals from Rytilampi, where pelvic reduction
527 maps to this region (Fig. 4). The deletion is around 3.5 kb in size and fully encompasses
528 the Pel-2.5kb^{SALR} region of the three-spined stickleback (Chan *et al*. 2010). No
529 comparable deletion was found in any other individuals in the dataset (Fig. 4b). This
530 included the two White Sea populations where either complete reduction of both the
531 pelvic spines and girdles was observed (MAS), or a putative large effect locus affecting
532 both spine and girdle length was found to be segregating (BOL; Fig. 1; Supplementary
533 Fig. 2 and Supplementary Table 1).
534

535 *Candidate genes*

536 Seven candidate genes or regulatory elements for pelvic reduction from the literature
537 (Supplementary Table 4) were found in LGs with significant QTL (Fig. 2). *Hif1a* (Mudie
538 *et al*. 2014), *Pel* (Chan *et al*. 2010) and *Pou1f1* (Kelberman *et al*. 2009) are known to
539 regulate the expression of *Pitx1,* whereas four genes (*Fgf8*, *Wnt8c*, *Wnt8b* and *Hoxd9*) are
540 involved in the pelvic fin/hind limb development downstream of *Pitx1* (Don *et al.* 2012,

Tanaka *et al*. 2005). However, aside from *Pel* (LG7) only three of these, *Wnt8c* (LG6), *Hif1a* (LG15) and *Pouf1f1* (LG16) were clearly within the significant QTL regions (Fig. 2). One candidate locus, *Hif1a*, is also on LG1 where significant QTL peaks were found when analysing relative (but not absolute) trait values (Table 1 and Supplementary Fig. 5).

Proof of concept simulations

There were no trends in QTL allele frequencies in the simulated populations prior to local adaptation, showing that the 10000 generation burn-in was adequate (Supplementary Fig. 6). In the empirical data from the marine populations, the slope of the regression line between linearised F_{ST} and geographic distance (point distance in km) was 2.13^{-7} (95% of the bootstrap replicates between 2.05^{-7} and 2.2^{-7}) and 1.53^{-8} (95% of the bootstrap replicates between 1.10^{-8} and 1.96^{-8}) for nine- and three-spined sticklebacks, respectively. Thus, the slope of the IBD regression line in this dataset was 13.9 times higher for nine-spined sticklebacks compared to three-spined sticklebacks (Fig. 5a). When scaling the geographic distance in the simulated data to the empirical data, the distance between the two simulated marine populations furthest away from each other (i.e. the populations from which the focal freshwater populations were founded) equalled 264 km and 352 km for three- and nine-spined sticklebacks, respectively (Fig. 5a). Thus, with comparable levels of IBD as in the empirical data, our simulations mimic the levels of parallel evolution that can be expected in three- and nine-spined sticklebacks at relatively short geographic scales (<400km). The difference in IBD between the two species is also close

563 to that in the empirical data (264/352 = 0.75). In the empirical data, genetic
564 differentiation between freshwater habitats for populations <400 km from each other
565 (mean F_{ST} = 0.19 and 0.49 for three- and nine-spined sticklebacks, respectively) was also
566 on par with the simulations (mean [across simulation replicates] F_{ST} = 0.21 and 0.58 for
567 three- and nine-spined sticklebacks, respectively). Notably, F_{ST} was >0.8 (linearised
568 F_{ST}>4) between Pyöreälampi (no pelvic reduction) and Rytilampi (pelvic reduction
569 controlled by *Pitx1*/*Pel*), although these ponds are situated only 15 km from each other
570 (Fig. 5a; max and median F_{ST} = 0.96 and 0.62, respectively, for all pairwise freshwater-
571 freshwater comparisons). This is higher than the F_{ST} between any pair of three-spined
572 stickleback populations in the data (max = 0.78 and median = 0.26 for all pairwise
573 freshwater-freshwater comparisons). Note that the benchmark empirical data stem from
574 Northern European populations and may thus not accurately represent interspecific
575 differences in population demographic parameters elsewhere.
576
577 *Do population demographic parameters influence local adaptation?*
578 In the simulations, the relationship between marine-freshwater F_{ST} and freshwater-
579 freshwater F_{ST} depended on both the species and genetic architecture (Fig. 5b). For
580 instance, when the genetic architecture included one additive large effect locus
581 (architecture A), this locus was often involved in parallel evolution in three-spined
582 sticklebacks (65% of replicates), but not in nine-spined sticklebacks (3% of the
583 replicates). In 20% of the replicates for both species, the freshwater allele for this locus
584 was fixed in only one of the focal freshwater populations (i.e. local adaptation, but not

parallel evolution). When the trait under selection was controlled by several medium effect loci (architecture B), parallel evolution was more common in both stickleback species, particularly for the locus with the largest allelic value (20% and 80% for nine- and three-spined sticklebacks, respectively; Fig. 5b, c). Local adaptation also occurred in 57% of the replicates in nine-spined sticklebacks, and in 18% in the three-spined sticklebacks; Fig. 5b, c). For the locus with the second highest allelic value, the cases of both parallel evolution and local adaptation collectively dropped to 38% and 43% for three- and nine-spined sticklebacks, respectively. With one non-additive large effect locus with recessive alleles locally adapted to freshwater (architecture C), this locus was less likely to be involved in parallel evolution in the three-spined sticklebacks (48%), compared to architecture A. However, results for the nine-spined sticklebacks were similar to architecture A (6% parallel evolution; Fig. 5b and c). Thus, parallel evolution is an expected outcome in three- but not nine-spined sticklebacks, at relatively short (<400 km) geographical distances, particularly when a single additive large effect locus is responsible for freshwater adaptation.

Does local adaptation depend on ancestral allele frequency?

In the simulations, the frequency of a2 allele (locally adapted to freshwater) in each freshwater population after 2000 generations post-colonisation from the sea was always correlated with the a2 allele frequency in the founding marine population, demonstrating that local adaptation depended on ancestral genetic variation (Supplementary Fig. 7). This dependency was the strongest for small effect loci (phenotypic effect of a2 = 1). In these

607 cases, the linear regression slope (β) ranged between 1 and 1.4, the y-intercept ranged
608 between 0 and 0.09, and Pearson's squared correlation coefficient (r^2) ranged between
609 0.33 and 0.52, and was lowest in three-spined sticklebacks for genetic architecture A (y-
610 intercept = 0.51; $\beta = 4.2$; $r^2 = 0.07$; Supplementary Fig. 7). In the latter case, standing
611 genetic variation (defined as a2 frequency > 0) was available in 89% of the simulation
612 replicates (considering each freshwater population independently), of which 76% resulted
613 in local adaptation (defined as a2 frequency > 0.5) in at least one of the freshwater
614 populations. The corresponding numbers for nine-spined sticklebacks were 41% and
615 39%, respectively. Here the correlation between ancestral a2 frequency and a2 frequency
616 post colonisation was also stronger than for three-spined sticklebacks (y-intercept = 0.04;
617 $\beta = 10.8$; $r^2 = 0.33$), indicating a stronger dependence between ancestral variation and
618 local adaptation. When standing genetic variation was available for local adaptation, three
619 factors likely contributed to lower levels of local adaptation and parallel evolution in
620 nine-spined sticklebacks: *i*) the smaller *K* in nine-spined stickleback freshwater
621 populations (500 vs. 1000), which is expected to result in faster genetic drift; *ii*) lower
622 (1/5th) post-colonisation gene-flow between adjacent marine and freshwater populations;
623 and *iii*) lower a2 frequency (when present as standing genetic variation) in nine-
624 compared to three-spined sticklebacks (2.9% vs. 3.4%). Thus, IBD in the sea likely limits
625 the probability that a freshwater adapted allele is maintained as standing genetic
626 variation, while smaller effective population sizes and reduced gene flow can further limit
627 local adaptation, even when ancestral variation in the founding population is present (in
628 low frequencies). The generally higher β's (3.20 vs. 2.15) and y-intercepts (0.11 vs. 0.05)

for QTL with a2<6 for nine-spined sticklebacks indicate that smaller effect QTL mostly have a stronger influence on local adaptation in nine- compared to three-spined sticklebacks (Supplementary Fig. 7), resulting in polygenic and/or incomplete local adaptation (Supplementary Fig. 6 and Supplementary Fig. 8).

Discussion

This study demonstrates that the genetic basis of pelvic reduction in nine-spined sticklebacks is more variable than in three-spined sticklebacks. This is consistent with the empirical evidence that IBD in the sea is stronger, and that freshwater populations are more isolated in nine- than three-spined sticklebacks. According to our simulations, such scenario would limit the pool of ancestral variation available for freshwater adaptation in nine-spined sticklebacks compared to three-spined sticklebacks. Thus, the level of ancestral variation available for local adaptation (and also for parallel evolution) is likely a function of population demographic parameters, with local adaptation being less likely in poorly connected species, as suggested by Merilä (2013, 2014). However, other non-mutually exclusive factors, such as the genetic architecture (i.e. dominance effects, heritability, mutation rates and numbers of causal loci involved), as well as non-parallelism in phenotypic selection optima are also likely to play roles. In the following, we discuss the possible causes of the discrepancy in pelvic structure development between nine- and three-spined sticklebacks, as well as their implications to our understanding of adaptive evolution in the wild.

Can the genetic architecture of pelvic reduction be explained by population demographic history?

Together with earlier QTL studies (Shapiro *et al.* 2006, 2009; Shikano *et al.* 2013), we show that multiple genomic regions (11 QTL, of which ten are novel to this study) are associated with pelvic reduction in nine-spined sticklebacks across their distribution range. Only one small effect QTL region (LG1; Table 1 and Supplementary Fig. 5) was shared between any two crosses (HEL × BYN and HEL × PYÖ), but even here it is not certain that the underlying causal mutations are the same. Support for genetic parallelism among marine-freshwater pairs of three-spined stickleback populations is strong (Colosimo *et al.* 2005; Jones *et al.* 2012; Terekhanova *et al.* 2014, 2019; Nelson and Cresko 2018). Although the high frequency of *Pel* deletions (disrupting pelvic armour development) means that most of the deletions associated with pelvic reduction in three-spined sticklebacks are independently derived (Xie *et al.* 2019), *Pitx1*/*Pel* is predominantly responsible for pelvic reduction in three-spined sticklebacks (Chan *et al.* 2010; Xie *et al.* 2019). On the other hand, almost all of the studied nine-spined stickleback populations to date display different QTL regions responsible for pelvic reduction (of varying QTL effect sizes and dominance relationships). One explanation for this could be that the independently derived *Pel* deletions (among other possible solutions for pelvic reduction) are usually the most favoured by selection, and when they are accessible as standing genetic variation, they quickly spread and become fixed in newly colonised freshwater populations. However, with a more limited pool of ancestral variation, freshwater adaptation in nine-spined sticklebacks could be restricted to less

673 optimal solutions. This is consistent with our results that showed only marginal, partial or
674 no pelvic reduction in seven out of the ten pond populations, and exemplified by the HEL
675 × PYÖ and HEL × BYN crosses where pelvic reduction was polygenic and much less
676 heritable than in the HEL × RYT cross. This hypothesis predicts that if the pelvic
677 reducing *Pel*-deletion from Rytilampi (or any other allele with comparable fitness effects)
678 would invade pond populations that currently do not display complete pelvic reduction,
679 this mutation would quickly spread and become fixed. Such rapid spread of freshwater
680 adapted *EDA* alleles has been well documented in the three-spined stickleback (Schluter
681 *et al*. 2010; Bell *et al*. 2004).

682 Using simulations with population demographic parameters similar to those
683 estimated from empirical data, we demonstrated that parallel evolution is more likely to
684 occur in three- than in nine-spined sticklebacks – that is, when independently founded
685 freshwater populations have access to the same ancestral standing genetic variation. This
686 is consistent with the large body of studies showing that parallel evolution among pairs of
687 marine-freshwater ecotypes in three-spined sticklebacks is commonly due to the repeated
688 use of the same ancestral variation from a global pool of standing genetic variation in the
689 sea, both at the global (Jones *et al.* 2012) and more local scales (Colosimo *et al.* 2005;
690 Jones *et al.* 2012; Terekhanova *et al.* 2014, 2019; Nelson and Cresko, 2018).
691 Furthermore, while pelvic reduction in freshwater populations of three-spined stickleback
692 populations is more or less the norm (Colosimo 2005; Klepaker *et al*. 2013), our survey
693 of phenotypic data from the wild demonstrates that full pelvic reduction in northern
694 European freshwater populations of nine-spined sticklebacks is rare. In fact, it only

695 occurs in two of the ten populations, with one additional population segregating for a
696 large effect locus and one showing reduction in spines but not in girdles. Consistent with
697 these results, when the access to ancestral standing genetic variation was restricted in the
698 simulations (by IBD in the sea and population structuring among freshwater populations
699 being similar to that observed in nine-spined sticklebacks), parallel evolution was much
700 less likely, and local adaptation was less complete or lacking altogether. Furthermore,
701 (partial) local adaptation was often polygenic and due to fixation of phylogenetically
702 independent (small effect) QTL, and not due to parallel evolution.

703 In our simulations, genetic parallelism in the three-spined stickleback was less
704 common when the large effect freshwater adapted allele was recessive (architecture C),
705 compared to when it was additive (architecture A). This may also partly explain why the
706 *Pitx1/Pel* deletions responsible for pelvic reduction in three-spined sticklebacks are
707 mostly due to repeated independent deletions rather than phylogenetically related alleles
708 (Xie *et al.* 2019). In contrast, the additive *EDA* gene controlling lateral armour plate
709 numbers in three-spined sticklebacks is a classic example of local adaptation from
710 ancestral standing variation (Colosimo *et al.*, 2005; Schluter and Conte 2009).
711 Furthermore, if a single large effect QTL already leads to a near-optimal phenotype, any
712 additional minor or medium effect loci could lead individuals farther from the optimum,
713 i.e. maladaptation (Bolnick *et al*. 2018). If so, selection is expected to prune such
714 variants, leaving only large effect QTL in the population. This could explain why other
715 small or medium effect loci for pelvic reduction are rare – but not completely absent – in
716 three-spined sticklebacks as well as in Rytilampi nine-spined sticklebacks. Consistent

717 with this, the single large effect locus available for freshwater adaptation (architectures A
718 and C) was the most commonly used allele for local adaptation at the expense of medium
719 effect loci; using both would have led to overshooting the phenotypic optimum (in both
720 species; Supplementary Fig. 8; Rogers *et al*. 2012).
721
722 *Geographic heterogeneity in selection optima*
723 Another explanation for the high heterogeneity in the genetic architecture of pelvic
724 reduction in the nine-spined stickleback resides within habitat environmental variation
725 resulting in different selection optima in different pond populations (cf. Stuart *et al.* 2017;
726 Thompson *et al.* 2019). The small differences between pelvic reduced phenotypes in our
727 study (e.g. in BYN and MAS/BOL spines are completely absent, while in RYT they are
728 only strongly reduced; Fig. 1, Supplementary Table 1 and Supplementary Fig. 2) could in
729 fact indicate different selection optima in the different populations (Stuart *et al*. 2017;
730 Thompson *et al*. 2019). It is possible to use available phenotypic data to estimate the
731 phenotypic optima of pelvic morphology (hypersphere) in each of the populations using
732 Fisher's geometric model (Stuart *et al.* 2017; Thompson *et al*. 2019), where a strong
733 overlap would suggest a higher probability of genetic parallelism (Thompson *et al.* 2019).
734 However, this assumes that the populations have access to exactly the same ancestral
735 variation and are free to evolve and reach their optima, which is at odds with the results
736 presented here. Without detailed environmental data or direct estimates of strength of
737 selection on pelvic phenotypes, disentangling the effects of gene flow and within habitat
738 environmental variation (assuming this leads to non-parallel angles of selection) is not

739 possible (Stuart *et al.* 2017).

740 In a recent simulation study, Thompson *et al.* (2019) showed that genetic
741 parallelism from standing genetic variation rapidly declines as selection starts to change
742 from fully parallel (optima angle of 0°) to divergent (optima angle of 180°), especially
743 when the trait is polygenic. However, although selection was fully parallel in our
744 simulations, we failed to observe strong genetic parallelism for smaller effect loci (with
745 allelic effects < 6) in both the three- and nine-spined stickleback-like scenarios (Fig. 5a,
746 c). This suggests that the effects of the underlying genetic architecture on parallelism (in
747 conjunction with some IBD and population structuring) can be independent of the angle
748 of optimal phenotypes between two habitats. In other words, non-parallelism in selection
749 optima is not necessary to explain non-parallelism at the phenotypic and genetic levels.
750 Thus, it is important not to disregard the population demographic setting as a factor that
751 could severely restrict heritability for adaptation and/or constrain adaptation to less
752 optimal solutions. Evolutionary studies of species with population demographic
753 parameters comparable to those typical for vulnerable or endangered species/populations
754 (such as the nine-spined stickleback) would be valuable to gain a better understanding of
755 how such species may respond to environmental changes and urbanisation (Thompson *et*
756 *al.* 2018).

757

758 *The limited role of* Pitx1/Pel *in the nine-spined stickleback pelvic reduction*

759 Thus far, *Pitx1* has been found to be responsible for pelvic reduction in nine-spined
760 sticklebacks in both Canada (Shapiro *et al*. 2006) and Finland (Shikano *et al*. 2013; this

761 study), but we do not know to what extent the underlying mutations are identical by
762 descent. By scanning whole-genome sequence data of 27 populations, we found a
763 deletion in the genomic region expected to contain *Pel* (LG7) only in Rytilampi – where
764 pelvic reduction maps to this gene. Similar to marine-freshwater pairs of three-spined
765 sticklebacks, we would have expected *Pitx1/Pel* deletions to exist in several other nine-
766 spined stickleback pond populations as well, most notably in the MAS and BOL
767 populations, where phenotypic data suggest large effect QTL to be responsible for pelvic
768 reduction. Since no deletions were found in the genomic region spanning *Pel* and *Pitx1* in
769 MAS and BOL, *Pitx1*/*Pel* is likely not responsible for the pelvic reduction in these two
770 populations. This provides further evidence that the genomic basis of pelvic reduction in
771 the nine-spined stickleback is more heterogeneous than that in the three-spined
772 stickleback. The loci, most likely major effect, responsible for pelvic reduction in these
773 ponds are yet to be identified.
774
775 *Pelvic reduction outside marine-freshwater study systems*
776 While the evidence for genetic parallelism at large geographical scales in the marine-
777 freshwater stickleback model system is extensive (Colosimo *et al.* 2005; Jones *et al.*
778 2012; Terekhanova *et al.* 2014, 2019; Nelson and Cresko 2018; Fang et al. 2019), the
779 level of parallelism in lake-stream and pelagic-benthic ecotype pairs of three-spined
780 sticklebacks is much more diverse (Peichel *et al*. 2001, 2017; Conte *et al*. 2012, 2015;
781 Stuart *et al*. 2017). For instance, Conte *et al*. (2015) found that among benthic-limnetic
782 three-spined stickleback pairs from Paxton and Priest lakes (Vancouver Island, BC,

783 Canada), 76% of 42 phenotypic traits diverged in the same direction, whereas only 49%
784 of the underlying QTL evolved in parallel in both lakes. For highly parallel traits in two
785 other pairs of benthic-limnetic sticklebacks, only 32% of the underlying QTL were shared
786 (Conte *et al*., 2012). Similarly, Stuart *et al*. (2017) found that among 11 independent
787 evolutionary replicate pairs of lake-stream three-spined stickleback populations
788 (Vancouver Island, BC, Canada), both within habitat variation and constraints to gene
789 flow contributed to the observed variation in levels of phenotypic parallelism. Different
790 lakes and streams do not likely have similar access to the same global pool of ancestral
791 variation as pairs of marine-freshwater three-spined sticklebacks, where gene flow in the
792 sea is high. This is consistent with the notion that more heterogeneous access to ancestral
793 variation can indeed limit genetic parallelism. This is also true for one example of
794 marine-freshwater three-spined stickleback divergence among isolated insular freshwater
795 populations in the Haida Gwaii archipelago off the northern Pacific coast of Canada
796 (Deagle *et al.* 2013). Here, similar to the nine-spined sticklebacks in this study, several
797 freshwater populations did not display any reduction in pelvic armour. However, those
798 populations that were fully plated were also genetically more similar to adjacent marine
799 individuals, suggesting that recent marine-freshwater admixture and/or selection
800 favouring plated freshwater individuals could explain this pattern. Other morphological
801 traits, such as lateral plating (coded by the *EDA* locus) were instead more consistent with
802 the same global variants being repeatedly re-used at the regional scale. Thus, with respect
803 to access to ancestral variation available for freshwater adaptation, nine-spined
804 sticklebacks are likely closer to the three-spined stickleback lake-stream and benthic-

805 limnetic study systems than to the three-spined stickleback marine-freshwater study
806 system. The only notable exception is the lake-stream three-spined sticklebacks
807 mentioned above, where genetic structuring also is high.
808
809 *Epistatic control of pelvic reduction?*
810 For traits with more than one additive QTL of equal effect sizes, the F_2 phenotypes are
811 expected to be normally distributed with a mean close that of the mean for the parents
812 (Klug and Cummings 2018). This was not the case for spine length in the HEL × BYN
813 cross, which was controlled by three QTL with similar effect sizes. In this case, the bulk
814 of the phenotypic values was centered around the mean, but also had a long tail of
815 individuals with strongly reduced pelvic spines (Fig. 1, Fig. 3 and Supplementary Fig. 2).
816 This skew in the phenotype distribution could be caused by epistatic interactions among
817 loci controlling pelvic spine length (Wolf *et al*. 2000). Consistent with this hypothesis,
818 complete spine reduction was most likely when the allele responsible for spine reduction
819 for the LG6 QTL was combined with at least one allele causing spine reduction from the
820 LG15 and LG16 QTL (Fig. 3b and Supplementary File 3). If a threshold number of
821 alleles are needed for complete pelvic reduction, this could also explain how standing
822 genetic variation in the sea is maintained, as the necessary multi-locus genotypes that
823 cause sub-optimal phenotypes in the sea are rarely formed, due to overall lower
824 frequencies of spine-reducing alleles in the sea. This is analogous to "epistatic shielding"
825 that can contribute to the persistence of disease alleles in populations (Phillips and
826 Johnsson 1998; Phillips 2008). Consistent with this possibility, LG6 of the F_0 female of

827 the HEL × BYN cross (from the sea) was polymorphic for the pelvic spine QTL effect
828 (Fig. 2) – evidently a single pelvic-reducing allele alone in this female was not enough to
829 cause any pelvic reduction at all (this female had a complete pelvis).
830
831 *Candidate genes*
832 While the QTL peak for the *Pitx1*/*Pel* region in the HEL × RYT cross was narrow, this
833 was not the case for the other QTL we detected. Hence, due to the large QTL regions
834 detected by four-way single-mapping analyses, it was not meaningful to perform gene
835 ontology enrichment (GO) analyses – the QTL regions would have contained possibly
836 thousands of genes. Instead, we searched the literature for known candidate genes for
837 pelvic reduction and found three (excluding *Pitx1*/*Pel*) that were clearly contained within
838 the identified QTL regions (Fig. 2 and Supplementary Table 4). Due to the
839 aforementioned large QTL regions, these can only be considered as highly putative
840 candidate genes for pelvic reduction and will not be discussed further. However, further
841 studies of pelvic reduction might find these candidates worthy of attention.
842
843 *Conclusions*
844 Our results show that the repeated parallel reduction in pelvic structures in freshwater
845 populations of nine-spined sticklebacks is due to a diverse set of genetic changes: only
846 one small effect QTL for pelvic reduction was shared between the three experimental
847 crosses in this study. In one cross, pelvic reduction was mapped to the previously
848 identified *Pitx1*/*Pel* regions, but in the two other crosses, the genetic basis of pelvic

849 reduction was polygenic, and mapped to many different chromosomes. In addition to
850 these, yet another large effect QTL different from the *Pitx1*/*Pel* locus likely segregates in
851 one nine-spined stickleback population waiting to be identified. The results also shed
852 light on the possible drivers of the observed genetic heterogeneity underlying pelvic
853 reduction; as shown by simulations, heterogeneous genetic architectures are more likely
854 to emerge when access to ancestral variation is limited by strong isolation by distance and
855 population structuring. This reinforces the role of the nine-spined sticklebacks as a useful
856 model system, alongside the three-spined stickleback, to study adaptive evolution in the
857 wild. Furthermore, since the population demographic characteristics of nine-spined
858 sticklebacks are similar to small and endangered species/populations, it is also likely to
859 be a well-suited model to study genetics of adaptation in populations of conservation
860 concern.

861

862 **Literature cited**

863 Arendt, J., and D. Reznick. 2007. Convergence and parallelism reconsidered: what have
864 we learned about the genetics of adaptation? Trends Ecol. Evol. 23: 26–32.
865 https://doi.org/10.1016/j.tree.2007.09.011

866 Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., … C.
867 Schlötterer. 2019. Genetic redundancy fuels polygenic adaptation in Drosophila.
868 PLoS Biol. 17: 1–31. https://doi.org/10.1371/journal.pbio.3000128

869 Baker, J.A. 1994. Life history variation in female threespine stickleback. In The
870 evolutionary biology of the threespine stickleback. M.A. Bell MA, Foster SA, (eds).
871 Oxford University Press, New York. pp. 144–187.

872 Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation.
873 Trends Ecol. Evol. *23*: 38–44. https://doi.org/10.1016/j.tree.2007.09.008

874 Becker, W. A. 1984. Manual of quantitative genetics (Fourth ed.). Academic enterprises,
875 Washington.

876 Bell, M. A., and S. A. Foster (Eds). 1994. The evolutionary biology of the threespine
877 Stickleback. Oxford University Press, Oxford.

878 Bell, M. A., Khalef, V. and M. Travis. 2007. Directional asymmetry of pelvic vestiges in
879 threespine stickleback. J. Exp. Zool. Part B, Mol. Dev. Evol. 308B: 189–199.
880 https://doi.org/10.1002/jez.b

881 Bell, M. A., Orti, G., Walker, J. A., and J.P. Koenings. 1993. Evolution of pelvic
882 reduction in threespine stickleback fish: a test of competing hypotheses. Evolution
883 47: 906–914. https://doi.org/10.2307/2410193

884 Bell, M. A., Windsor, E. A., Buck, N. J. 2004. Twelve years of contemporary evolution
885 in a threespine stickleback population. *Evolution* 58:814-24. 10.1554/03-373.

886 Blouw, D. M., and G.J. Boyd. 1992. Inheritance of reduction, loss, and asymmetry of the
887 pelvis in Pungitius pungitius (ninespine stickleback). Heredity, 68: 33-42.
888 https://doi.org/10.1038/hdy.1992.4

889 Bolnick, D. I., Barrett, R. D. H., Oke, K. B., Rennison, D. J., and Y.E. Stuart. 2018.
890 (Non) Parallel Evolution. Ann. Rev. Ecol. Evol. Syst. 12: 303–330.

891 Chan, Y. F., Marks, M. E., Jones, F. C., Jr, G. V., Shapiro, M. D., Brady, S. D., … D.M.
892 Kingsley. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent
893 deletion of a Pitx1 enhancer. Science 327: 302–305.

894 Cohan, F. M. 1984. Can uniform selection retard random genetic divergence between
895 isolated conspecific populations? Evolution 38: 495-504.
896 https://doi.org/10.2307/2408699

897 Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, H.,
898 Grimwood, J., … D.M. Kingsley. 2005. Widespread parallel evolution in
899 sticklebacks by repeated fixation of ectodysplasin alleles. Science 307: 1928–1933.
900 https://doi.org/10.1126/science.1107239

901 Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., …
902 C.L. Peichel. 2015. Extent of QTL reuse during repeated phenotypic divergence of
903 sympatric threespine stickleback. Genetics 201: 1189–1200.
904 https://doi.org/10.1534/genetics.115.182550

905 Conte, G. L., Arnegard, M. E., Peichel, C. L., and D. Schluter. 2012. The probability of
906 genetic parallelism and convergence in natural populations. Proc. R. Soc B. 279:
907 5039–5047. https://doi.org/10.1098/rspb.2012.2146

908 Coyle, S. M., Huntingford, F. A., and C.L. Peichel. 2007. Parallel evolution of Pitx1
909 underlies pelvic reduction in Scottish Threespine Stickleback (Gasterosteus
910 aculeatus). Genetics *98*: 581–586. https://doi.org/10.1093/jhered/esm066

911 Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., … J.H.
912 Postlethwait. 2004. Parallel genetic basis for repeated evolution of armor loss in
913 Alaskan threespine stickleback populations. Proc. Natl. Acad. Sci. USA 101: 6050–
914 6055.

915 Deagle, B. E., Jones, F., Chan, Y. F., Absher, D. M., Kingsley, D. M.,and T. E.
916 Reimchen. 2012. Population genomics of parallel phenotypic evolution in
917 stickleback across stream–lake ecological transitions. Proc. R. Soc. B 279:1277–
918 1286 doi:10.1098/rspb.2011.1552

919 Deagle, B. E., Jones, F. C., Absher, D. M., Kingsley, D. M., and T. E. Reimchen. 2013.
920 Phylogeography and adaptation genetics of stickleback from the Haida Gwaii
921 archipelago revealed using genome-wide single nucleotide polymorphism
922 genotyping. *Mol. Ecol. 22:* 1917–1932. https://doi.org/10.1111/mec.12215

923 DeFaveri, J., Shikano, T., Ghani, N. I. A., and J. Merilä. 2012. Contrasting population
924 structures in two sympatric fishes in the Baltic Sea basin. Mar. Biol. 159: 1659–
925 1672. https://doi.org/10.1007/s00227-012-1951-4

926 DeFaveri, J., and J. Merilä. 2013. Variation in age and size in Fennoscandian three-spined
927 sticklebacks (*Gasterosteus aculeatus*). PLoS One 8: e80866.
928 https://doi.org/10.1371/journal.pone.0080866

929 DeFaveri, J., Shikano, T., and J. Merilä. 2014. Geographic variation in age structure and
930 longevity in the nine-spined stickleback (*Pungitius pungitius*). PLoS One 9:
931 e102660. https://doi.org/10.1371/journal.pone.0102660

932 Don, E. K., Currie, P. D., and N.J. Cole. 2012. The evolutionary history of the
933 development of the pelvic fin/hindlimb. J. Anatomy 222: 114–133.
934 https://doi.org/10.1111/j.1469-7580.2012.01557.x

935 Elmer, K. R., and A. Meyer. 2011. Adaptation in the age of ecological genomics: Insights
936 from parallelism and convergence. Trends Ecol. Evol. *26*: 298–306.
937 https://doi.org/10.1016/j.tree.2011.02.008

938 Fang, F., Kemppainen, P., Momigliano, P., and J. Merilä. 2019. Oceans apart:
939 Heterogeneous patterns of parallel evolution in sticklebacks, *BioRxiv*, doi:
940 https://doi.org/10.1101/826412

941 Gibson, G. 2005. The synthesis and evolution of a super-model. Science 307: 1890-1891

942 Giles, N. 1983. The possible role of environmental calcium levels during the evolution of
943 phenotypic diversity in outer Hebridean populations of the Three-spined stickleback,
944 *Gasterosteus aculeatus*. J. Zool. 199: 535–544. https://doi.org/10.1111/j.1469-
945 7998.1983.tb05104.x

946 Hahne, F. and R. Ivanek. 2016. Statistical Genomics: Methods and Protocols. In Mathé E,
947 Davis S (eds.), chapter Visualizing Genomic Data Using Gviz and Bioconductor,
948 335–351. Springer, New York.

949 Herczeg, G., Turtiainen, M., and J. Merilä. 2010. Morphological divergence of North-
950 European nine-spined sticklebacks (*Pungitius pungitius*): Biol. J. Linn. Soc. 101:
951 403–416.

952 Hermisson, J., and P.S. Pennings. 2017. Soft sweeps and beyond: understanding the
953 patterns and probabilities of selection footprints under rapid adaptation. Methods
954 Ecol. Evol. 8: 700–716. https://doi.org/10.1111/2041-210X.12808

955 Hijmans, R.J., and J. van Etten, 2012. raster: Geographic analysis and modeling with
956 raster data. R package version 2. 0-12. http://CRAN.R-project.org/package=raster

957 Karhunen, M., Merilä, J., Leinonen, T., Cano, J. M., and O. Ovaskainen. 2013. driftsel:
958 An R package for detecting signals of natural selection in quantitative traits. Mol.
959 Ecol. Res. 13: 746–754. https://doi.org/10.1111/1755-0998.12111

960 Karhunen, M., Ovaskainen, O., Herczeg, G., and Merilä, J. 2014. Bringing habitat
961 information into statistical tests of local adaptation in quantitative traits: A case
962 study of nine-spined sticklebacks. Evolution 68: 559-568.

963 Kelberman, D., Rizzoti, K., Lovell-Badge, R., Robinson I. C. A. F., and M. T. Dattani.
964 2009. Genetic Regulation of Pituitary Gland Development in Human and Mouse.
965 *Endocrine Reviews* 30:790–829, https://doi.org/10.1210/er.2009-0008

966 Kemppainen, P., Knight, C. G., Sarma, D. K., Hlaing, T., Prakash, A., Maung Maung, Y.
967 N., … C. Walton. 2015. Linkage disequilibrium network analysis (LDna) gives a
968 global view of chromosomal inversions, local adaptation and geographic structure.
969 Mol. Ecol. *15*: 1031–1045. https://doi.org/10.1111/1755-0998.12369

970 Klepaker, T., Østbye, K., and M. A. Bell. 2013. Regressive evolution of the pelvic
971 complex in stickleback fishes: a study of convergent evolution. Evol. Ecol. Res. 15:
972 413–435.

973 Klug, W. S., and M.R. Cummings. 2018. Concepts of genetics. Pearson Education, Inc,
974 New Jersey.

975 Laine, V. N., Shikano, T., Herczeg, G., Vilkki, J., and J. Merilä. 2013. Quantitative trait
976 loci for growth and body size in the nine-spined stickleback *Pungitius pungitius* L.
977 Mol. Ecol. 22: 5861–5876. https://doi.org/10.1111/mec.12526

978 Leinonen, T., Cano, J. M., and J. Merilä. 2011. Genetics of body shape and armour
979 variation in threespine sticklebacks. J. Evol. Biol. 24: 206–218.
980 https://doi.org/10.1111/j.1420-9101.2010.02161.x

981 Li, H. 2018. Sequence analysis Minimap2 : pairwise alignment for nucleotide sequences.
982 Bioinformatics 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

983 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … T. Sam. 2009.
984 The Sequence Alignment / Map format and SAMtools. Bioinformatics 25: 2078–
985 2079. https://doi.org/10.1093/bioinformatics/btp352

986 Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-
987 MEM. arXiv:1303.3997v2 [q-bio.GN]

988 Li, Z., Guo, B., Yang, J., Herczeg, G., Gonda, A., Balázs, G., … J. Merilä. 2017.
989 Deciphering the genomic architecture of the stickleback brain with a novel
990 multilocus gene-mapping approach. Mol. Ecol. 26: 1557–1575.
991 https://doi.org/10.1111/mec.14005

992 Li, Z., Kemppainen, P., Rastas, P., and J. Merilä. 2018. Linkage disequilibrium
993 clustering-based approach for association mapping with tightly linked genomewide
994 data. Mol. Ecol. Res. 18: 809–824. https://doi.org/10.1111/1755-0998.12893

995 Merilä, J. 2013. Nine-spined stickleback (*Pungitius pungitius*): An emerging model for
996 evolutionary biology research. *Ann. NY. Acad. Sci. 1289*: 18–35.
997 https://doi.org/10.1111/nyas.12089

998 Merilä, J. 2014. Lakes and ponds as model systems to study parallel evolution. J. Limnol.
999 73: 33–45. https://doi.org/10.4081/jlimnol.2014.805

1000 Mudie, S., Bandarra, D., Batie, M., Biddlestone, J., Ortmann, B., Shmakova, A., … S.
1001 Rocha. 2014. PITX1, a specificity determinant in the HIF-1α- mediated

1002 transcriptional response to hypoxia Sharon. Cell Cycle 13: 3878–3891.
1003 https://doi.org/10.4161/15384101.2014.972889

1004 Nelson, T. C., and W.A. Cresko. 2018. Ancient genomic variation underlies repeated
1005 ecological adaptation in young stickleback populations. Evol. Lett. 2: 9–21.
1006 https://doi.org/10.1002/evl3.37

1007 Neuenschwander, S., Michaud, F., and J. Goudet. 2018. QuantiNemo 2: a Swiss knife to
1008 simulate complex demographic and genetic scenarios, forward and backward in
1009 time. Bioinformatics 35:886-888. https://doi.org/10.1093/bioinformatics/bty737

1010 Orr, H. A., and L. R. Unckless. 2008. Population Extinction and the Genetics of
1011 Adaptation. *Am. Nat.* 172:2, 160-169

1012 Peichel, C. L., Nereng, K. S., Ohgi, K. A., Cole, B. L. E., Colosimo, P. F., Buerkle, C. A.,
1013 …, and D.M. Kingsley. 2001. The genetic architecture of divergence between
1014 threespine stickleback species. Nature 414: 901–905.

1015 Pembleton, L.W., Cogan, N.O.I., and J.W. Forster. 2013. StAMPP: an R package for
1016 calculation of genetic differentiation and structure of mixed-ploidy level
1017 populations. Mol. Ecol. Res. 13: 946-952

1018 Phillips, P. C. 2008. Epistasis — the essential role of gene interactions in the structure
1019 and evolution of genetic systems. Fundamental Concepts in Genetics 9: 855–867.
1020 https://doi.org/10.1038/nrg2452

1021 Phillips, P. C., and N.A. Johnson. 1998. The population genetics of synthetic lethals.
1022 Genetics 150: 449–458.

1023 Pritchard, J. R., and D. Schluter. 2001. Declining interspecific competition during
1024 character displacement: Summoning the ghost of competition past. Evol. Ecol. Res.
1025 3: 209–220.

1026 Rastas, P. 2017. Lep-MAP3: Robust linkage mapping even for low-coverage whole
1027 genome sequencing data. Bioinformatics 33: 3726–3732.
1028 https://doi.org/10.1093/bioinformatics/btx494

1029 Reid, S., Tibshirani, R., and J. Friedman. 2016. A study of error variance estimation in
1030 lasso regression. Statistica Sinica 26: 35-67. https://www.jstor.org/stable/24721190

1031 Reimchen, T. E. 1983. Structural Relationships Between Spines and Lateral Plates in
1032 Threespine Stickleback (Gasterosteus aculeatus). Evolution 37: 931–946.
1033 https://doi.org/10.2307/2408408

1034 Reist, J. D. 1980. Predation upon pelvic phenotypes of brook stickleback, *Culaea*
1035 *inconstancs*, by selected invertebrates. Can. J. Zool. 58: 1253-1258.

1036 Rogers, S. M., Tamkee, P., Summers, B., Balabahadra, S., Marks, M., Kingsley, D. M.,
1037 and D. Schluter. 2012. Genetic signatures of adaptive peakshift in threespine
1038 sticklebacks. *Evolution* 66: 2439–2450.

1039 Rosenblum, E. B., Parent, C. E., and E.E. Brandt. 2014. The molecular basis of
1040 phenotypic convergence. Annu. Rev. Ecol. Syst. 45: 203-226. doi:10.1146/annurev-
1041 ecolsys-120213-091851

1042 Schluter, D., and G.L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad.
1043 Sci USA, 106: 9955–9962. https://doi.org/10.1073/pnas.0901264106

1044 Schluter, D., Marchinko, K.B., Barrett, R.D., and S.M. Rogers. 2010. Natural selection
1045 and the genetics of adaptation in threespine stickleback. Philos Trans R Soc Lond B
1046 Biol Sci. 365(1552):2479–2486. doi:10.1098/rstb.2010.0036

1047 Shapiro, M. D., Bell, M. A., and D.M. Kingsley. 2006. Parallel genetic origins of pelvic
1048 reduction in vertebrates. Proc. Natl. Acad. Sci USA 103: 13753–13758.
1049 https://doi.org/10.1073/pnas.0604706103

1050 Shapiro, M. D, Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jo, B., …
1051 D.M. Kingsley. 2004. Genetic and developmental basis of evolutionary pelvic
1052 reduction in threespine sticklebacks. Nature 428: 717–724.

1053 Shapiro, Michael D, Summers, B. R., Balabhadra, S., Aldenhoven, J. T., Miller, A. L.,
1054 Cunningham, C. B., … and D.M. Kingsley. 2009. The genetic architecture of
1055 skeletal convergence and sex determination in ninespine sticklebacks. Curr. Biol. 19:
1056 1140–1145. https://doi.org/10.1016/j.cub.2009.05.029.The

1057 Shikano, T., Laine, V. N., Herczeg, G., Vilkki, J., and J. Merilä. 2013. Genetic
1058 architecture of parallel pelvic reduction in ninespine sticklebacks. *3G: Genes,*
1059 *Genomes, Genetics*, *3*, 1833–1842. https://doi.org/10.1534/g3.113.007237

1060 Shikano, T., Shimada, Y., Herczeg, G., and J. Merilä. 2010. History vs. habitat type:
1061 Explaining the genetic structure of European nine-spined stickleback (Pungitius
1062 pungitius) populations. Mol. Ecol. 19: 1147–1161. https://doi.org/10.1111/j.1365-
1063 294X.2010.04553.x

1064 Sillanpää, M. J. 2011. On statistical methods for estimating heritability in wild
1065 populations. Mol. Ecol. 20: 1324–1332. https://doi.org/10.1111/j.1365-
1066 294X.2011.05021.x

1067 Stern, D. L. 2013. The genetic causes of convergent evolution. Genetics 14: 751–764.
1068 https://doi.org/10.1038/nrg3483

1069 Stuart, Y. E., Veen, T., Weber, J. N., Hanson, D., Ravinet, M., Lohman, B. K., … D.I.
1070 Bolnick. 2017. Contrasting effects of environment and genetics generate a
1071 continuum of parallel evolution. Nature Ecol. Evol. 1:0158.
1072 https://doi.org/10.1038/s41559-017-0158

1073 Taggart, J. B., Hynes, R. A., and A. Ferguson. 1992. A simplified protocol for routine
1074 total DNA isolation from salmonid fishes. J. Fish. Biol 40: 963–965.

1075 Tanaka, M., Hale, L. A., Amores, A., Yan, Y., Cresko, W. A., Suzuki, T., and J.H.
1076 Postlethwait. 2005. Developmental genetic basis for the evolution of pelvic fin loss
1077 in the pufferfish Takifugu rubripes. Dev. Biol. 281: 227–239.
1078 https://doi.org/10.1016/j.ydbio.2005.02.016

1079 Terekhanova, N. V, Barmintseva, A. E., Kondrashov, A. S., Bazykin, G. A., and N.S.
1080 Mugue. 2019. Architecture of parallel adaptation in ten lacustrine threespine
1081 stickleback populations from the White Sea area. Genome Biol. Evol. 11: 2605–
1082 2618. https://doi.org/10.1093/gbe/evz175

1083 Terekhanova, N. V, Logacheva, M. D., Penin, A. A., Neretina, T. V, Barmintseva, A. E.,
1084 Bazykin, G. A., … N.S. Mugue. 2014. Fast evolution from precast bricks :
1085 Genomics of young freshwater populations of threespine stickleback Gasterosteus
1086 aculeatus. PLoS Genet. 10: e1004696. https://doi.org/10.1371/journal.pgen.1004696

1087 Thompson, K. A., Osmond, M. M., and D. Schluter. 2019. Parallel genetic evolution and
1088 speciation from standing variation. Evol. Lett. 3:129-141.
1089 https://doi.org/10.1002/evl3.106

1090 Thompson, K. A., Rieseberg, L. H., and D. Schluter. 2018. Speciation and the City.
1091 *Trends Ecol. Evol* 33:815-826

1092 Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc.
1093 *Series. B* 58:267-288

1094 Varadharajan, S., Rastas, P., Löytynoja, A., Matschiner, M., Calboli, F.C.F., Guo, B., …
1095 J. Merilä. 2019. A high-quality assembly of the nine-spined stickleback (Pungitius
1096 pungitius) genome. Genome Biol. Evol. 11:3291-3308. evz240,
1097 https://doi.org/10.1093/gbe/evz240

1098 Van Ooijen, J. W. 2009. MapQTL v. 6.0: Software for the mapping of quantitative trait
1099 loci in experimental populations of diploid species. Kyazma BV, Wageningen, The
1100 Netherlands.

1101 Wang, Y., Zhao, Y., Wang, Y., Li, Z., Baocheng Guo, B. and J. Merilä. 2020. Population
1102 transcriptomics reveals weak parallel genetic basis in repeated marine and
1103 freshwater divergence in nine-spined sticklebacks. Mol. Ecol.
1104 https://doi.org/10.1111/mec.15435

1105 Weir, B. S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of
1106 population structure. Evolution 38: 1358–1370.

1107 Wolf, J. B., Brodie, E. D., and M.J. Wade. 2000. Epistasis and the Evolutionary Process.
1108 Oxford University Press, Oxford.

1109 Xie, K. T., Wang, G., Thompson, A. C., Wucherpfennig, J. I., Reimchen, T. E., Maccoll,
1110 A. D. C., … and P. Lake. 2019. DNA fragility in the parallel evolution of pelvic
1111 reduction in stickleback fish. Science 84: 81–84.

1112 Xu, S. 1996. Mapping quantitative trait loci using four-way crosses. Genet. Res. 68: 175–
1113 181.

1114 Xu, S. 2013a. Genetic mapping and genomic selection using recombination breakpoint
1115 data. Genetics 195: 1103–1115. https://doi.org/10.1534/genetics.113.155309

1116 Xu, S. 2013b. Principles of Statistical Genomics. Springer, New York.
1117 https://doi.org/DOI 10.1007/978-0-387-70807-2

1118 Yang, J., Guo, B., Shikano, T., Liu, X., and J. Merilä. 2016. Quantitative trait locus
1119 analysis of body shape divergence in nine-spined sticklebacks based on high-density
1120 SNP-panel. Sci. Rep. *6*:26632. https://doi.org/10.1038/srep26632

1121 Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and B.S. Weir. 2012. A
1122 high-performance computing toolset for relatedness and principal component
1123 analysis of SNP data. Bioinformatics 28: 3326–3328.
1124 https://doi.org/10.1093/bioinformatics/bts606

Table 1 | Summary of QTL-mapping analyses. Each row corresponds to a significant (arbitrarily numbered) QTL region, with the proportion variance explained (PVE), jointly estimated for all PCs (one for each significant LD-cluster) extracted from such regions. Coding indicates whether the QTL was significant for the alleles inherited from the F_1 female (♀), the F_1 male (♂) or the dominance effect (dom). Effect size (β) is based on the first PC extracted from all SNPs from each significant QTL region. "Std." indicates whether the trait was standardised (Yes) or not (No). *P* and P_{COR} represent nominal and corrected *P*-values from single-mapping four-way analyses, respectively. Results for standardised trait values are only shown for QTL that were not also significant for absolute trait values.

Cross	Trait	Coding	LG	QTL	Std.	*P*	P_{COR}	PVE_{TOT}	β
HEL × RYT	Spine	male	7	1	No	1.31e-04	0.019	0.285	1.654
HEL × RYT	Spine	female	7	1	No	1.26e-04	0.026	0.311	1.573
HEL × RYT	Girdle	male	7	1	No	3.66e-04	0.040	0.311	2.041
HEL × RYT	Girdle	female	7	1	No	1.08e-04	0.024	0.263	1.711
HEL × RYT	Spine	dom	7	1	No	1.08e-07	<0.001	0.213	1.432
HEL × RYT	Girdle	dom	7	1	No	1.35e-04	0.028	0.154	1.463
HEL × BYN	Spine	male	15	2	No	6.12e-06	0.001	0.095	0.540
HEL × BYN	Spine	male	16	3	No	1.16e-04	0.010	0.086	0.493
HEL × BYN	Spine	male	21	4	No	3.36e-04	0.031	0.037	0.360
HEL × BYN	Spine	female	6	5	No	1.37e-04	0.018	0.070	0.500
HEL × BYN	Girdle	male	14	6	No	2.49e-04	0.024	0.039	0.307
HEL × PYÖ	Spine	male	9	7	No	8.96e-11	<0.001	0.108	0.352
HEL × PYÖ	Girdle	male	19	8	No	1.56e-05	0.001	0.056	0.325
HEL × PYÖ	Girdle	female	4	9	No	1.32e-04	0.020	0.045	0.301
HEL × BYN	Spine	female	16	10	Yes	2.53e-04	0.030	0.047	0.007
HEL × BYN	Girdle	female	1	11	Yes	2.13e-05	0.002	0.077	0.006
HEL × BYN	Girdle	male	1	11	Yes	1.41e-04	0.016	0.026	0.005
HEL × PYÖ	Spine	male	1	11	Yes	4.75e-06	<0.001	0.075	0.005
HEL × PYÖ	Girdle	male	1	11	Yes	9.69e-05	0.016	0.044	0.005

Table 2 | Proportion phenotypic variance explained (PVE) in pelvic traits. Shown are the percentages of total phenotypic variance explained by different linkage groups (LG), by all SNPs (Tot), by loci inherited from females (♀) and males (♂), as well as the dominance effect (Dom). Shown are results for each cross and trait separately and for absolute trait values.

LG	HEL x RYT		HEL x BYN		HEL x PYÖ	
	Spine	Girdle	Spine	Girdle	Spine	Girdle
1	-	0.01	0.03	0.05	-	-
2	-	-	0.01	0.03	-	0.02
3	-	-	-	0.01	-	-
4	-	0.02	0.01	0.01	-	0.03
5	-	-	-	-	-	-
6	-	-	0.03	-	0.02	-
7	0.85	0.82	0.02	0.02	-	-
8	-	-	0.01	0.01	-	-
9	-	-	-	0.02	0.11	-
10	-	-	-	0.06	-	-
11	-	-	0.02	-	-	-
12	0.02	0.02	0.03	0.01	0.01	-
13	-	-	-	0.01	-	0.02
14	-	-	-	0.05	0.01	0.02
15	-	-	0.09	0.01	-	-
16	-	-	0.12	0.01	-	-
17	-	-	0.01	0.01	-	-
18	-	-	-	0.04	-	0.01
19	-	-	0.02	0.06	-	0.05
20	-	-	-	0.01	-	-
21	-	-	0.03	-	0.01	-
Tot	0.8	0.76	0.39	0.32	0.14	0.16
♂	0.3	0.38	0.24	0.16	0.14	0.08
♀	0.29	0.27	0.17	0.06	-	0.09
Dom	0.23	0.17	0.01	0.13	0.01	-

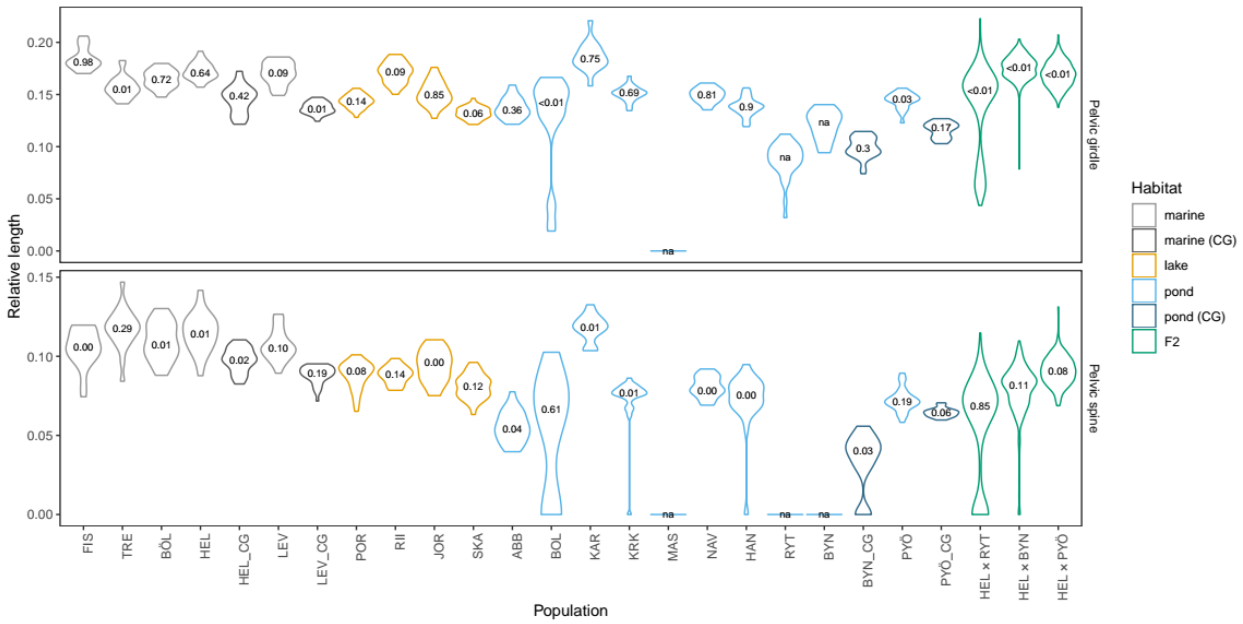

Figure 1 | Summary of previously published nine-spined stickleback pelvic phenotypes from the wild and common garden experiments. Violin plots depict relative (to standard body length) pelvic spine and girdle lengths according to population and habitat type. Numbers in the top panel show *P*-values for Pearson's product moment correlation between relative pelvic spine and relative pelvic girdle lengths, and the bottom panel the respective squared correlation coefficients for each population/cross. "na" indicates data with no variation in spine or girdle lengths. Further details can be found in Supplementary Fig. 1 and Supplementary Table 1. Phenotypic data for F_2 individuals from the current study are also presented (HEL x RYT, HEL x BYN and HEL x PYÖ), are also shown for comparison. Details of sample locations and data collection can be found from Herczeg *et al*. (2010) and Karhunen et al. (2013). Data from common garden experiments is indicated by the suffix "_CG" in x-axis labels.

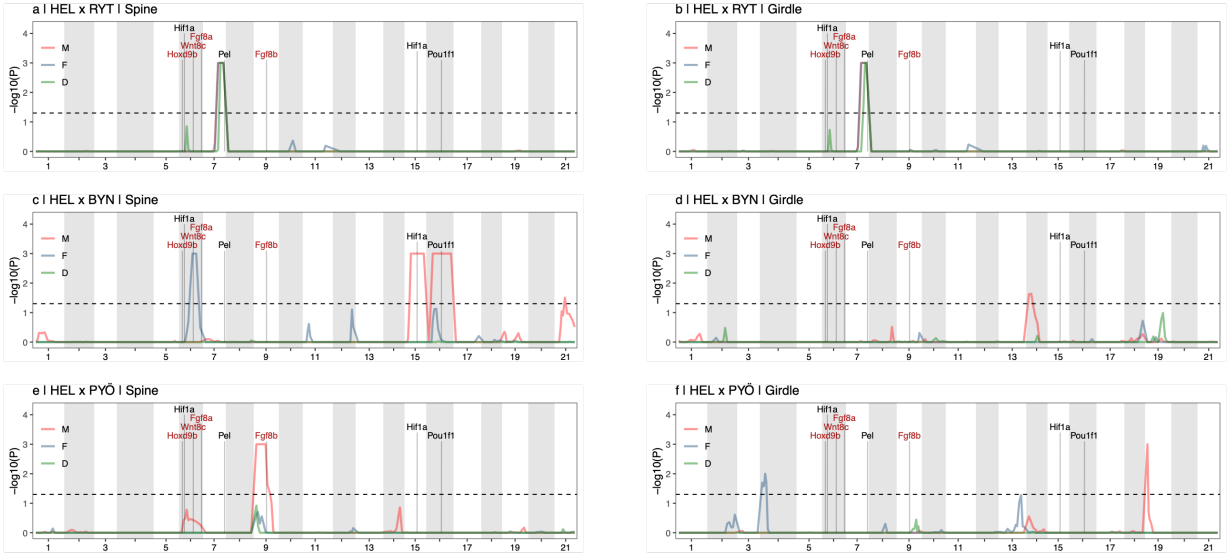

Figure 2 | Quantitative trait locus mapping of pelvic reduction in three independent stickleback crosses. Single-mapping four-way analyses of four morphological traits associated with pelvic reduction in (a-b) HEL × RYT cross, (c-d) HEL × BYN cross and (e-f) HEL × PYÖ cross. Shown are QTL for pelvic spine length and girdle length with x-axis indicating position in centi Morgans (cM). Results are based on permutation and dotted vertical line indicates genome wide significance at $\alpha = 0.05$. Results are shown separately for alleles inherited from the male F_1 (M), the female F_1 (F) together with the dominance effect (D) according to legend. Absence of a dominance effect indicates that the trait inheritance is additive, whereas a peak only for M or F indicates that the allelic effect was segregating in the F_0 male or the F_0 female, respectively (see Supplementary File S1 for details). Candidate genes involved in pelvic development are indicated with black text representing genes that affect expression of *Pitx1* and red text indicating genes that affect pelvic development downstream of *Pitx1* expression. Results for analyses based on relative trait values can be found in Supplementary Figure 4.

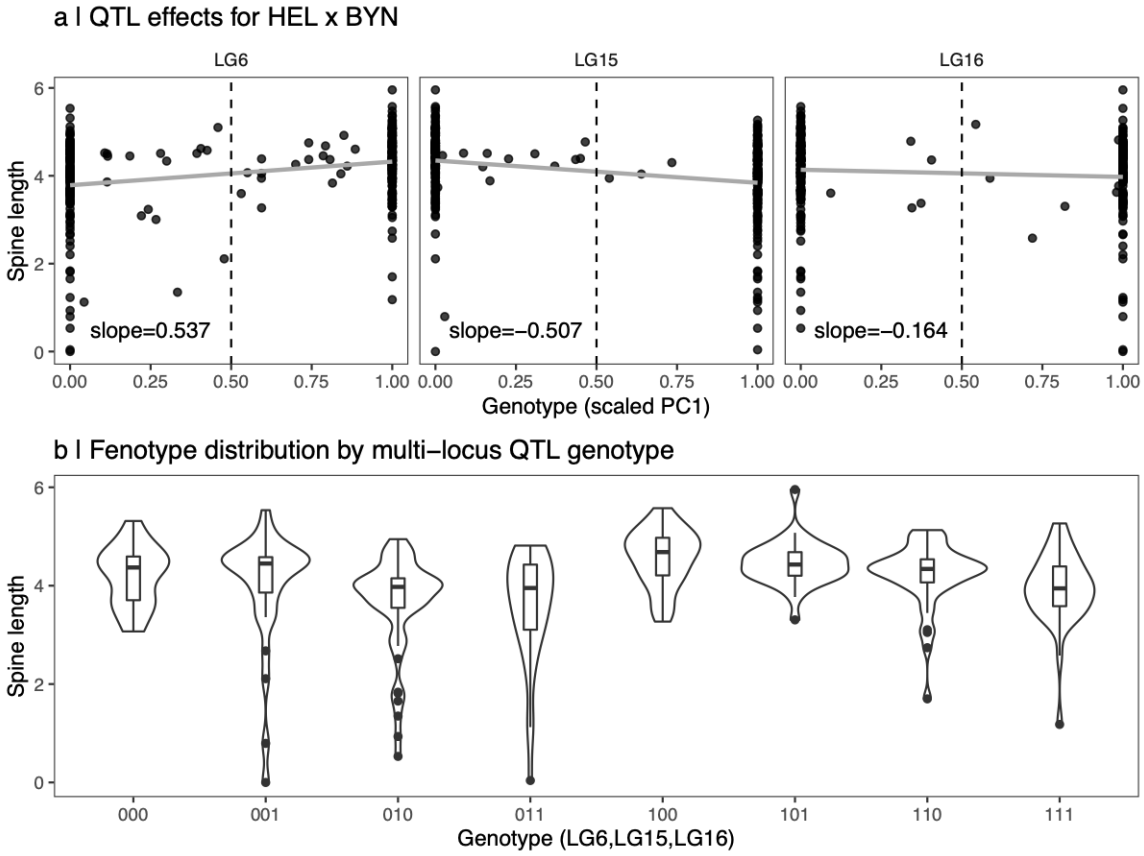

Figure 3 | Epistatic interactions in pelvic spine development for HEL × BYN cross. (a) individual genotype effects where the genotype is given by the first PC (scaled between 0 and 1) from the cluster of SNPs that was the most significant for spine length on LG6, LG15 and LG16 (see Fig. 2), respectively. Genotypes were further based on the genotypes [x_{dij}, x_{sij}] depending on which of these were significant for the QTL effects (x_{dij} for LG6 and x_{sij} for LG15 and LG16). Some individuals have genotypes between 0 and 1 only because the genotype is based on the first PC of large sets of highly but not perfectly correlated SNPs. Slope of the regression line (grey) is shown. (b) distribution of spine lengths for all multi-locus genotypes from (a). The multi-locus genotype was based on rounding PC1 coordinates from (a) where values below 0.5 (left of vertical dashed line) were considered as "Allele 0" or below or equal to 0.5 were considered as "Allele 1". The first digit for genotypes in (b) represents thus the genotype of the QTL on LG6, followed by LG15 and LG16, respectively.

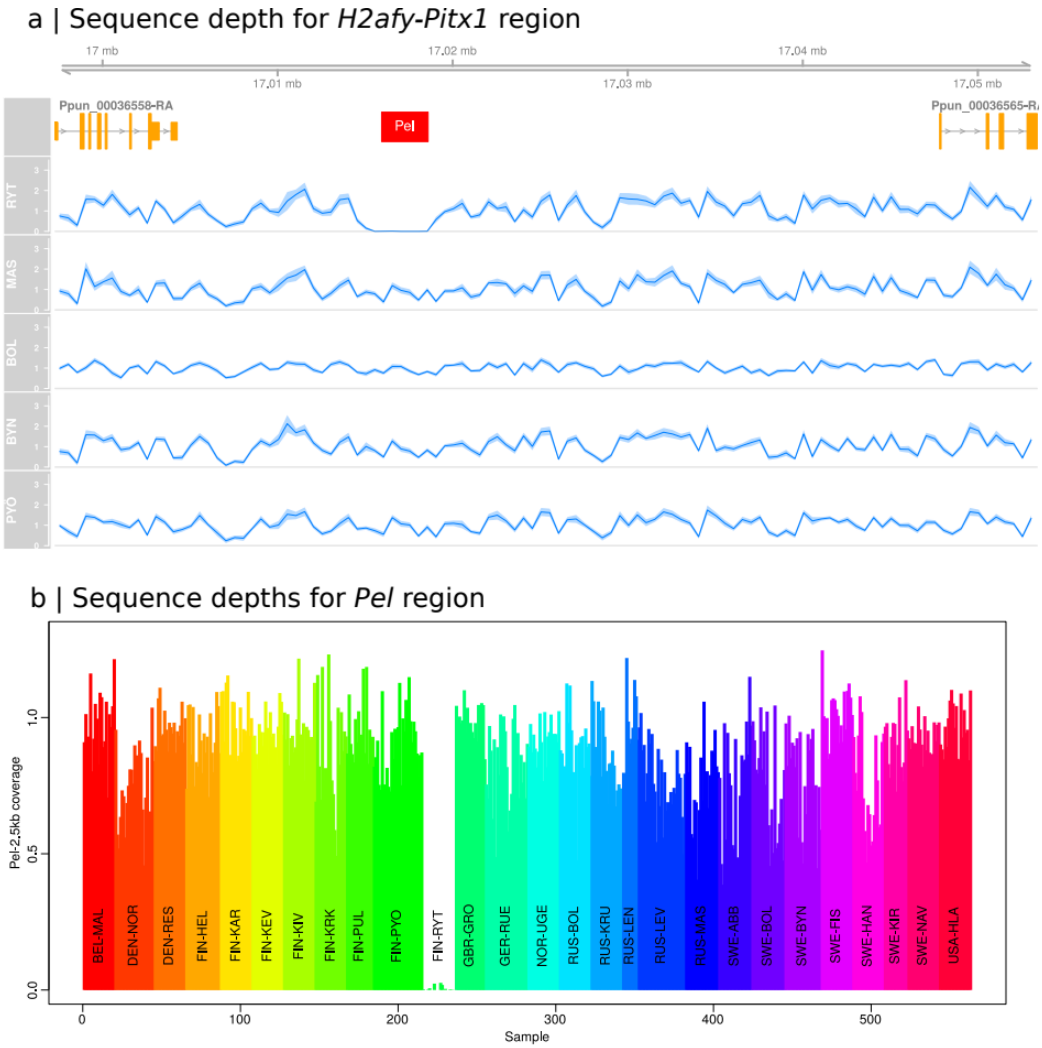

Figure 4 | Sequence depths for *H2afy-Pitx1* region. (a) Relative sequencing depth across the *H2afy-Pitx1* intergenic region for 10 individuals from five different populations. Blue line and shading indicate population average and 95% confidence intervals, respectively. The 3.5kb deletion in RYT, seen as a dip sequencing depth, fully encloses the Pel-2.5kb^{SALR} of three-spined stickleback indicated by the red box. (b) Normalised sequencing depths for the Pel-2.5kb^{SALR} region for 563 samples from 27 populations show that the complete deletion of the *Pel* region is unique to RYT. Phenotypic data in the Russian populations MAS and BOL suggest a large effect locus is responsible for pelvic reductions, and BYN and PYÖ are populations analysed here where pelvic reduction does not map to LG7.

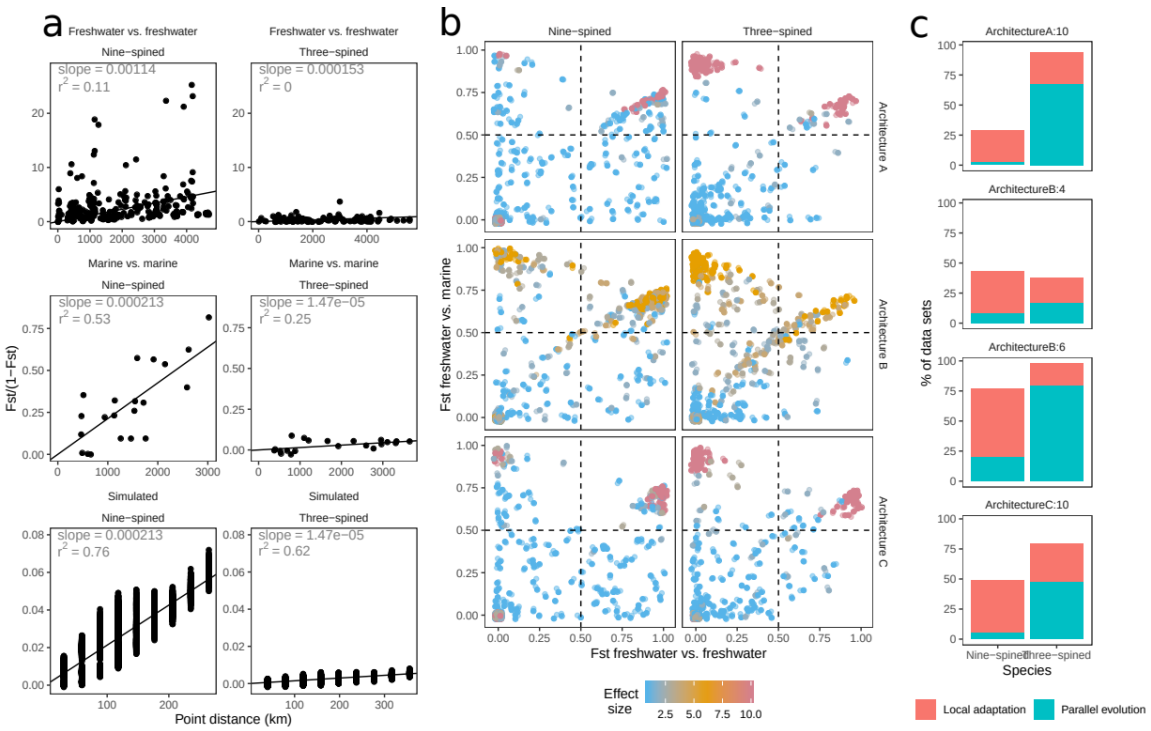

Figure 5 | Simulation results. Linearised F_{ST} against geographic distance (IBD) for empirical and simulated data (a), with slope and squared Pearson's product moment correlation coefficient indicated. Geographic distance for simulated data (IBD in the sea) is scaled to match the slope for the IBD-plot in the sea in empirical data. Freshwater-freshwater F_{ST} against marine-freshwater F_{ST} from all QTL from all simulated data (n = 100), with effect sizes indicated as shown in legend (b). Loci in upper left quadrant are classified as being involved in parallel evolution, and loci in the upper right quadrant are loci that are involved in local adaptation in only one freshwater population. This data is summarised in (c) focusing on the four largest effect loci, with genetic architecture and effect sizes indicated by the figure titles.