Convergent genetic architecture underlies parallel pelvic reduction in genetically highly structured stickleback species

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, Research Programme in Organismal and Evolutionary Biology, 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]*JY, Chinese Sturgeon Research Institute, Three Gorges Corporation, Yichang, 443100, China*

*corresponding author: juha.merila@helsinki.fi

Abstract

Repeated and independent evolution of the same phenotypes in similar environments is common, but less is known about the repeatability of the underlying genetic mechanisms, particularly in species with small effective population sizes and strong population structuring. We identified genomic regions responsible for the repeated reduction of pelvic spines and girdles in three crosses between pond (reduced pelvic apparatus) and marine (full apparatus) nine-spined sticklebacks (*Pungitius pungitius*) using quantitative trait locus (QTL) mapping. In one cross both traits mapped to linkage group 7 (LG7), where the gene *Pituitary homeobox transcription factor 1* (*Pitx1*) is known to be associated with pelvic reduction. In the other two crosses, pelvic reduction was polygenic and the nine out of the ten novel QTL (explaining 3-10% of the total phenotypic variation) were all unique to one of the three crosses. We further screened whole genomes from 27 different populations for the presence of deletions in *Pitx1* regulatory element (*Pel*) which, in the related *Gasterosteus aculeatus*, is associated with a repeated pelvic reduction. No deletions were found in any other locations, except in the population where pelvic reduction mapped to LG7, including ponds where phenotypic data suggested fixation of large effect loci causing complete pelvic reduction. These results, which are backed up by simulations, are consistent with the notion that in small populations where founder events and strong genetic drift prevail, parallel phenotypic evolution can be expected to be highly non-parallel at the genetic level. The results contribute to our understanding of genetic architecture of adaptation highlighting the fact that a diversity of

genetic architectures may underline expression of similar phenotypes in species characterised by high degree of population structuring.

Keywords: convergent evolution, epistasis**,** *Gasterosteus aculeatus*, parallel evolution, pelvic reduction, Pitx1, *Pungitius pungitius,* QTL

Introduction

A fundamental question in evolutionary biology is to what extent is adaptation reliant on different mutations (convergent evolution) and to what extent can we expect alleles that are identical by descent to contribute to adaptation in multiple independent populations experiencing similar selection pressures (parallel evolution)? Perhaps even more intriguing is the question under what circumstances and demographic scenarios is one or the other process expected to dominate? Although genetic mapping has revealed traits under the control of single genes/genomic regions – particularly in the case of discrete phenotypes – most quantitative variation is believed to be polygenic in nature (Mackay *et al*. 2009; Hill & Kirkpatrick 2010). Moreover, much of the selection responses in polygenic quantitative traits is thought to be based on standing genetic variation (Barrett & Schluter 2007; Thompson *et al*. 2019). Therefore, when exposed to novel but yet similar environments, populations derived from the same ancestral population and carrying the same pool of alleles can often be expected to respond to similar selection pressures in a similar fashion, and evolve parallel genetic adaptations (Arendt & Reznick 2007; Schluter & Conte 2009; Elmer & Meyer 2011; Stern 2013; Conte *et al*. 2015; Bolnick *et al*. 2018; Barghi *et al*. 2019; Hermisson & Pennings 2017). However, in small populations, wherein founder events and random genetic drift prevail, potentially advantageous rare alleles (if even present within founder groups) may be lost, and/or adaptation to given selection pressures might be more easily gained by allelic substitutions in alternate loci influencing the same polygenic trait (Cohan 1984; Merilä

2013, 2014; Rosenblum *et al.* 2014). Thus, without detailed insight into the genetic architecture of such traits, much of the genetic and evolutionary diversity inherent in the wild may go undetected.

The three-spined stickleback (*Gasterosteus aculeatus*) is a widely used model system to study adaptive evolution in the wild (Bell & 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 as explained by the "transporter hypothesis" (Schluter & 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). It differs from the three-spined stickleback by usually having smaller effective population sizes (N_e), reduced gene flow in the sea, and its 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 opposite ends of a parallelism-convergence continuum with respect to local adaptation.

A striking example of repeated parallel evolution in sticklebacks (Gasterosteidae) is the reduction of pelvic skeleton. Regressive evolution of the pelvic complex has occurred in freshwater populations of three (*viz*., *Gasterosteus*, *Pungitius*, and *Culaea*) out 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 a complete pelvic apparatus with fully developed pelvic girdles and lateral pelvic spines, partial or even complete pelvic reduction is common in freshwater populations (Blouw & Boyd 1992; Shapiro et al. 2004, 2006; Herczeg *et al*. 2010; Klepaker *et al*. 2013). Several factors may contribute to the loss of pelvic girdles in freshwater habitats, including the absence of gape-limited predatory fish and limited calcium availability, as well as 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). Thus and collectively, sticklebacks provide an important model system to study genetic mechanisms underlying the adaptive parallel pelvic reduction at both intra- and inter-specific levels, under a wide range of population demographic settings. However, as compared to three-spined sticklebacks, studies of parallel patterns of marine-freshwater divergence in nine-spined sticklebacks are scarce, both at the empirical and the theoretical levels, precluding any comprehensive and conclusive comparative studies of the two species.

Genetic basis of pelvic reduction in the three-spined stickleback is well understood. Independent quantitative trait locus (QTL) mapping studies have identified a single chromosomal region containing the gene *Pituitary homeobox transcription factor 1* (*Pitx1*) explaining more than two thirds of the variance in pelvic size in crosses between individuals with complete pelvic girdles and spines, and pelvic-reduced individuals (Cresko *et al*. 2004; Shapiro *et al*. 2004; Coyle *et al*. 2007). It has further been demonstrated that pelvic loss in the marine-freshwater three-spined stickleback model

system is predominantly caused by expression changes of *Pitx1* gene due to recurrent deletions of the pelvic enhancer (*Pel*) upstream of *Pitx1* (Chan *et al*. 2010; Xie *et al.* 2019), although the genetic underpinnings of pelvic reduction in benthic-limnetic and lake-stream pairs of three-spined sticklebacks appear to more diversified (Peichel *et al.* 2001, 2017; Stuart *et al.* 2017). So far the genomic region controlling pelvic reduction has been identified in only three pelvic-reduced nine-spined stickleback populations (Shapiro et al. 2006, 2009; Shikano *et al.* 2013). Although *Pitx1* in linkage group (LG) 7 was identified as a major cause for the pelvic apparatus reduction in one Canadian (Shapiro *et al.* 2006) and in one Finnish (Shikano *et al.* 2013) population, another large effect region in LG4 was found to be associated with pelvic reduction in an Alaskan population (Shapiro *et al.* 2009). Similarly to three-spined sticklebacks, pelvic spine and the pelvic girdle sizes are strongly correlated in the Finnish pond Rytilampi since *Pitx1* on LG7 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; Fig. 1). Given their high heritability (Blouw & Boyd 1992; Leinonen *et al.* 2011), the lack of correlation between spine and girdle lengths suggests that these traits can also independently be controlled by different QTL. Thus, the genetic underpinnings of the 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 parallel 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 aside of the two new populations (Bynästjärnen and Pyöreälampi) using over 89 000 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 that *Pel* could be responsible for pelvic reduction also in the nine-spined stickleback, we screened whole genomes of 27 wild populations for deletions (or lack thereof) in the genomic region spanning the *Pel* element and the *Pitx1* gene. In addition, as a more diversified genetic architecture underlying pelvic reduction in the nine- compared to the three-spined stickleback could be due to stronger population structuring and reduced N_e in freshwater populations of the former species, we used individual based forward simulations to test how realistic levels (as estimated from empirical data) of population structuring and genetic isolation by distance (IBD) in the sea affects the prevalence of genetic parallelism when local adaptation proceeds from standing genetic variation.

Materials and Methods

Fish collection, crossing, and rearing

For the QTL crosses, three different marine grandparental females (F_0) from the Baltic Sea (Helsinki, Finland; 60°13′N, 25°11′E) were crossed with a male grandparental pond individual from Bynästjärnen (64°27′N, 19°26′E), Pyöreälampi (66°15′N, 29°26′E) or Rytilampi (66°23′N, 29°19′E), respectively. Fish crossing, rearing, and sampling followed the experimental protocol used in the earlier study of the Rytilampi population (Shikano *et al.* 2013; Laine *et al.* 2013). For Rytilampi, the F_0 parental fish were artificially mated in the lab during the early breeding season of 2006 (Shikano *et al.* 2013), and the resulting full-sib F_1-offspring were group-reared in aquaria as explained in Shikano *et al.* (2013) until mature. Two randomly chosen F_1-individuals were mated to produce the F_2 generation between September and October 2008, with the offspring reared in separate aquaria. The same procedure was followed for Pyöreälampi (mating: July-Sep 2012, F_2 breeding: July 2012-Apr 2013) and Bynästjärnen (mating: Nov 2013-Jan 2014, F_2 breeding: Nov 2013-Aug 2014). In total 308, 283 and 279 F_2-offspring were available for analyses from Helsinki × Bynästjärnen (HEL × BYN), Helsinki × Pyöreälampi (HEL × PYÖ) and Helsinki × Rytilampi (HEL × RYT) crosses, respectively.

The experiments were conducted under licenses from the Finnish National Animal Experiment Board (#STH379A and #STH223A). Experimental protocols were approved by the ethics committee of University of Helsinki, and all experiments were performed in accordance with relevant guidelines and regulations.

Morphological data

To visualise bony elements, all the F_2-progeny were stained with Alizarin Red S following Pritchard & Schluter (2001). Pelvic spine and girdle lengths from both sides of the body, as well as standard body length, were measured with digital callipers to the nearest 0.01 mm. Although it is known that the left-right asymmetry of the pelvic girdle is also heritable in sticklebacks (Blouw & Boyd 1992; Bell *et al.* 2007; Coyle *et al.* 2007), we did not specifically study this here. To reduce the number of tests, the mean of the left and the right-side measurements were used. All measurements were made by the same person twice, and the repeatability (R; Becker 1984), ranged between 0.80 and 0.84 for girdle lengths and between 0.98 and 0.99 for spine lengths. The QTL analyses were performed on both absolute and relative (scaled to the total body length) trait values, but for all analyses comparing phenotypic data between populations (which also differ in body sizes), only relative trait values were used. Previously published phenotypic data from wild populations were obtained from Herczeg *et al*. (2010) and Karhunen *et al.* (2013). Visibly broken spines were treated as missing data.

Genotyping and linkage map construction

The RAD sequencing protocol used to obtain the SNP data was the same as in Yang *et al.* (2016) and Li *et al*. (2017). In short, Genomic DNA was extracted from ethanol preserved fin clips using the phenol-chloroform method (Taggart *et al*. 1992). DNA was

fragmented by the restriction enzyme PstI and DNA fragments of 300 to 500 bp were gel purified. Illumina sequencing adaptors and library specific barcodes were ligated to the digested DNA fragments and barcoded RAD samples were then pooled and sequenced on 24 lanes of the Illumina HiSeq2000 platform with 45 bp single-end strategy. RAD library construction and sequencing were performed by BGI HONGKONG CO., LIMITED. Adapters and barcodes were eliminated from reads and quality was checked 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 the dense SNP data, which allowed us to utilise the four-way cross QTL mapping (see below). Input data were generated by first mapping individual fastq files to the nine-spined stickleback reference genome using BWA mem (Li 2013) and SAMtools (mpileup; Li *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), and markers segregating in a more distorted fashion than what would be expected by 1:1,000 odds by chance were filtered out (module Filtering2). Loci were then partitioned into chromosomes (modules SeparateChromosomes2 and JoinSingles2) yielding 21 linkage groups with a total of over 89,000 markers assigned to these groups. Finally, the markers

were ordered within each linkage group with (module OrderMarkers2) removing either markers only informative in the mother or father, respectively. This created two maps for each chromosome, one having more maternal markers and the other having more paternal markers, on average sharing ⅔ of the markers. Constructing two maps like this removes the effect of markers 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, 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, conducting corrections for multiple testing is essential to reduce the rate of false positives. In addition, in large genomic data sets adjacent SNPs (along chromosomes), particularly from experimental crosses, are often in linkage disequilibrium (LD), i.e. correlated. As a group of SNPs in high LD explain 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 by function "snpgdsLDMat" from R-package "SNPRelate", Zheng *et al.* 2012) below the LD threshold value of 0.7 were removed, which resulted in many sub-clusters where all loci were connected by at least a single edge. Second, for each sub cluster, a complete linkage clustering (where a cluster is defined by its smallest link) was performed and, 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 LD-estimates between adjacent 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 analysed nevertheless 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. For loci in each LD-cluster, we then applied a PCA and extracted the first principal component that captured

the largest portion of variation. The original SNP data in the QTL-model were then replaced with coordinates from these PCs, with sex-linked loci (Pearson's correlation coefficient >0.95 between the PC coordinates and sex) removed. The input data for the LDn-clustering comprised of the original co-dominant SNP data with individuals from all three crosses pooled.

QTL mapping: four-way crosses model

In some circumstances, such as in the case of a four-way cross (Xu 1996), F_1 offspring of a hybrid cross generated from two heterozygous parents (Van Ooijen 2009), and in the case of 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 a case a linear regression model for QTL analysis of the outbred F_2 data (Li Z. *et al.* 2018) is defined by:

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

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

$$\left|\begin{array}{l|l} +1+1+1 & forgenotypeAC, \\ +1-1-1 & forAD, \\ -1+1-1 & forBC, \\ -1-1+1 & forBD. \end{array}\right. , (2)$$

(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 (estimable with LM3) and has the benefit that it can be used to infer the source of the observed QTL effect as described in more detail in Supplementary Information File1. A multiple correction on the basis of permutation tests was further conducted to control false positives due to multiplicity (Li *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, and all the other symbols are defined in the same way as in Equation (1).

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

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

where β_j is the effects of the SNPs, 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:

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

where $\beta_j(j \notin G)$ 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 dominance component by treating the coding systems [x_{dij}, x_{sij}, z_{ij}] (2) as different groups of SNPs.

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 22 other populations were sequenced to 10X coverage by BGI HONGKONG CO., LIMITED. Reads were mapped to the nine-spined stickleback reference genome (Varadharajan *et al*. 2020) with BWA mem (Li 2013) and site-wise sequencing coverages were computed with SAMtools (depth; Li *et al*. 2009). Relative sequence depths across the *H2afy-Pitx1* intergenic region were estimated for the five focal populations by first computing the median depths for 1000 bp sliding windows and then normalising these by the median depth for the full intergenic region. The Pel-2.5kb^{SALR} region was extracted from the original BAC contig (GenBank accession number 'GU130433.1') and mapped to the nine-spined stickleback intergenic region with minimap2 (Li 2018). The sequencing depths for the *Pel*-2.5kb^{SALR} were normalised by dividing the mean depths of the *Pel* region by the mean depths of the full intergenic region. Gene annotations and relative sequencing depths (average and confidence intervals) were computed and visualised using R-package "Gviz" (Hahne & Ivanek 2016). Lastly, we scanned the literature for genes that are known to affect pelvic and hind limb development and searched for potential matches in the nine-spined stickleback genome (Varadharajan *et al*. 2020) in regions where significant QTL were found.

Proof of concept simulations

QuantiNemo2 (Neuenschwander *et al.* 2008) was used to perform individual based

forward simulations of out-crossing diploid individuals with non-overlapping generations to investigate how differences in population structure influence the prevalence of parallel evolution. Two scenarios were investigated. The first corresponded to three-spined stickleback like situation in which genetic isolation by distance (IBD) in the sea is low, and two post-glacial freshwater populations are connected by some gene flow from the sea. The second scenario mimicked the nine-spined stickleback like population structure, where isolation by distance in the sea was ~10 times stronger, carrying capacity (K) of the marine population was halved and gene flow between the freshwater and marine populations was 1/5th compared to the first scenario. These values are within the range of what previously has been reported from empirical data (DeFaveri *et al.* 2012) and our analyses (see below).

The sea populations were simulated by a stepping-stone model comprised of ten subpopulations with carrying capacities (K) of 1000 each (which under mutation drift equilibrium equates to N_e). All adjacent marine population were exchanging 100 (scenario 1) or ten (scenario 2) migrants per generation (M, symmetrical gene flow) with some long-distance migration allowed also between every second population at a rate of ten times less than that between adjacent populations (Supplementary Fig. 2). A refuge freshwater population (K=10000), where freshwater adapted alleles are allowed to thrive, exchanges migrants with the two most central of the ten marine populations at a rate of M=1 in both directions (Supplementary Fig. 2). While this is not a realistic scenario, the sole purpose of this was to allow frequencies of freshwater adapted alleles build up (see

below) and filter to the marine populations as standing genetic variation. Two focal freshwater populations were then founded from the two marine populations situated at the opposite ends of the stepping-stone chain (Supplementary Fig. 2). Thus, the freshwater populations were founded from different marine populations with ancestral freshwater adapted alleles stemming from the same refuge freshwater population situated equally far from the two focal freshwater populations, mimicking the "transporter hypothesis" of Schluter & Conte (2009).

The probability of the same ancestral alleles being used for local adaptation in two or more freshwater populations likely also depends on the allelic effect sizes and the strength of selection for and against a given allele in freshwater and in marine populations, respectively. Hence, three different genetic architectures of a single trait coded by five (independent) bi-allelic loci under stabilising selection for different optima in the marine (optimal phenotypic value is 0) and freshwater populations (optimal phenotypic value is 20) were simulated. In architecture A, we simulated one large effect additive locus with the homozygote for allele 1 yielding a genotypic value of zero (locally adapted to marine habitat) and the homozygote for allele 2 yielding a genotypic value of 20 (locally adapted for the freshwater habitat). The four remaining minor effect additive QTL were given the allelic value of zero for allele 1 and allelic values [1,1,2,3] for allele 2. An empirical example of this kind of architecture is the Ectodysplasin (EDA) gene in the three-spined stickleback (Colosimo *et al*. 2004, 2005). In architecture B, the allelic values of the freshwater adapted alleles were [1,2,3,4,6] and the allelic values of

the alternative alleles were zero, i.e. a single large effect locus is lacking. In architecture C, the distribution of allelic values was the same as in architecture A, but with the allele 2 being recessive to allele 1. An empirical example of this from three- and nine-spined sticklebacks is provided by the *Pitx1* locus (see Introduction).

The selection intensity was set to 100 for the freshwater habitat and 200 for the marine habitat, lower values translating to stronger selection intensity (these values do not easily translate to something that has real biological meaning, see quantiNemo manual for details). After trial and error, these values allowed sufficiently high frequencies of allele 2 in the sea as standing genetic variation and sufficiently rapid subsequent local adaptation in newly founded freshwater populations. We simulated ten chromosomes of 100 centi Morgan (cM) and 100 equally spaced neutral loci each, with the QTL for the locally adapted trait also evenly distributed in the genome (i.e. uniform recombination rate across chromosomes). Mutation rate (μ) was set to $10e^{-8}$ for all loci and simulations were initiated with allele frequencies for all loci set to 0.5. We first allowed a 10,000 generations burn-in, after which the freshwater populations were founded (K changed from 0 to 1000 or 500 in the focal three- or nine-spined stickleback freshwater populations, respectively) by allowing these populations to be founded from the sea during a single generation (with M=K). Following this, we allowed M=1 between the focal freshwater populations and their closest marine populations (in both directions) for scenario 1 and M=0.2 for scenario 2 (Supplementary Fig. 2). The simulations were then run for 5000 generations to simulate post glacial colonisation of freshwater habitats

(10000 years ago, assuming a generation time of two; Baker 1994; DeFaveri & Merilä 2013; DeFaveri et al. 2014). Parameter settings for the refuge population remained unaltered throughout the simulations. One hundred replicate simulations were run for each parameter combination. Full simulation details are available from DRYAD (doi: XX).

At generation 15000 we estimated allele frequency differences as the fixation index, F_{ST} (Weir & Cocherham 1984), by function "stamppFst" in R-package "StAMPP" (Pembleton 2013). We classified loci being involved in parallel evolution if the F_{ST} between the focal 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}; the refuge population was not analysed further). Any loci with both high marine-freshwater F_{ST} and marine-freshwater F_{ST} (above 0.5) was classified as being locally adapted (but not involved in parallel evolution).

To obtain empirical data to inform our simulations, we also analysed a subset of 4,326 single nucleotide polymorphism from 236 nine-spined and 98 three-spined sticklebacks samples from 54 Fennoscandian populations. This data was extracted from previously available RAD-seq and full genome sequence data as described in Supplementary File 3. The sole purpose of this was to provide a ballpark estimate of IBD and degree of freshwater-freshwater genetic differentiation for the simulations. For IBD analyses pairwise genetic distance between populations was regressed against the

geographic distance separating the populations. We estimated geographic distance in the World Geodetic System 84 reference ellipsoid (i.e. point distance when taking into account the curvature of the planet) using function "pointDistance" from R-package "raster" (Hijmans & van Etten 2012), and genetic distance as $F_{ST}/(1-F_{ST})$ (linearised F_{ST}), with F_{ST} estimated as shown above, for all pair-wise population comparisons. Slope and confidence intervals of the regression lines were estimated by bootstrapping SNPs (n=1000). As the geographic distances between the marine populations in the simulations are 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.

Results

Heterogeneity of pelvic reduction in the wild

Re-analysis of previously published phenotypic data from wild populations confirmed a high degree of variation in pelvic spine and girdle lengths among different northern European freshwater populations, as well as in their inter-correlations (Fig. 1, Supplementary Table 1 and Supplementary Fig. 1). For instance, while spines were absent and the pelvic girdles strongly reduced (but not completely absent) in Rytilampi (RYT), complete lack of both spines and girdles characterised the Mashinnoje population (MAS; Fig. 1 and Supplementary Table 1). Furthermore, the Bynästjärnen (BYN)

population shows complete lack of spines, but only partially reduced pelvic girdles, whereas the Bolotnoje (BOL; a population adjacent to MAS) shows large variation both in spine and girdle lengths with a strong correlation between the two traits (Fig. 1, Supplementary Table 1 and Supplementary Fig. 1). Six additional pond populations (*viz*. ABB, KAR, KRK, NAV, HAN, and PYÖ; Fig. 1) show no dramatic pelvic reduction although several of these (the notable exception being the KAR population) display some reductions in relative spine and girdle lengths relative to marine populations (Fig. 1, Supplementary Table 1 and Supplementary Fig. 1).

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, uncovered a significant QTL on LG7 for both pelvic spine and girdle lengths in single-locus analyses (Fig. 2a, b and Table 1). In 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). Approximately equal amount of the phenotypic variance was explained by alleles inherited from the grandfather (pond individual) and the grandmother (marine individual; ~30% of the total variance for all traits; Table 1), respectively, with appreciable phenotypic variance also explained by dominance effects (15-21%). This is the expected pattern 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 & 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 male F_1 (pond) individual. Thus, the causal alleles for these QTL segregated in the F_0 pond grandfather (as explained in Supplementary File 1). No dominance effect was detected for these QTL suggesting that the allelic effects were additive. One QTL, on LG6 with allelic effect deriving only from the F_1 female was also detected (Fig. 2c and Table 1). The QTL significance profiles, in particular for LG15 and LG16, spanned large genomic regions (with no obvious peaks, in contrast to for LG7 in the HEL × BYN cross; Fig. 2), and this remained true also when analysing all SNPs individually (fine-mapping; Supplementary Fig. 3). The individual and multi-locus phenotypic effects on spine lengths for the QTL on LGs 6, 15 and 16 (using the most significant QTL for each QTL region) are detailed in Figure 3. In the absence of any large effect loci and when all QTL are additive and independent (as is the case here), phenotypes in the F_2 generation are expected to be normally distributed (Klug & Cummings 2018). For some multi-locus genotypes of the QTL on LGs 6, 15 and 16, the distribution of spine lengths were approximately normally distributed except with long tails of highly reduced spine lengths, implicating some of

these loci are involved in epistatic interactions (this was investigated further in Supplementary File 2). One significant male QTL on LG4 for girdle length was also detected (Fig. 2d and Table 1). Results for relative trait values were highly similar to the absolute trait values, except for an additional significant male and female QTL on LG1, as well as an additional significant female QTL on LG16 (Supplementary Fig. 4, Table 1 and Supplementary Table 2).

Multi-locus analyses (for absolute trait values) identified 11 LGs that accounted for at least 3% of the phenotypic variance in pelvic spine or girdle lengths in the HEL × BYN cross (Table 2). Largest of these effects were found on LG15 and LG16, which explained 9% and 12% of variation in pelvic spine lengths, respectively (Table 2). For pelvic spine and girdle lengths 39% and 32% of the PVE, respectively, were accounted for by all SNPs in the data set, which equates to the narrow sense heritability (h^2) accounting also for dominance (but not epistatic interaction) effects. For spine lengths 24% of the total PVE was attributed to alleles deriving from the F_1 male and 17% was attributed to alleles deriving from the F_1 female, and only 1% was attributed to the dominance effect (Table 2). For girdle lengths the corresponding numbers were 16%, 6% and 13% (Table 2).

QTL mapping of pelvic reduction in Helsinki × Pyöreälampi cross

In the HEL × PYÖ cross (283 F_2 individuals) we found a QTL for spine length on LG9 explained by alleles inherited from the F_1-male, and two significant QTL for girdle length

on LG19 explained by alleles inherited from the F_1-male and on LG4 explained by alleles inherited from the F_1-female (Fig. 2e, f, and Table 1). In multi-locus analyses, the PVE for different LGs mirrored these results, with LGs containing significant QTL explaining most of the PVE while PVE for all other LGs were < 4 % (Table 2). In multi-locus analyses, PVE for pelvic traits was lower than in the HEL × BYN cross; 14% and 16% for spine length and girdle length, respectively with <2% PVE due to dominance effects (Table 2). When analysing relative trait values, one additional female QTL peak was found for both girdle and spine lengths on LG1 (Supplementary Fig. 4, Table 1 and Supplementary Table 2).

Trait correlations in the QTL crosses

There was a strong correlation between relative pelvic spine and girdle lengths (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1) in the HEL × RYT cross, as expected since pelvic reduction in this cross is controlled by a single large effect QTL affecting both traits. However, in the HEL × BYN cross, correlation between relative pelvic spine and girdle lengths was much weaker (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1), consistent with pelvic spine and girdle reductions independently being controlled by different QTL. Furthermore, only 4/306 of the F_2 individuals displayed complete lack of spines, despite that the BYN population is fixed for complete spine reduction in the wild (and the spine was absent in the F_0 male), consistent with multiple additive loci controlling spine length in the HEL × BYN cross. Among the F_2 individuals

from HEL × BYN cross, relative girdle lengths were normally distributed with much smaller variances (SD=0.012) compared to the HEL × RYT (SD=0.041) with only two individuals with reduced girdles (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1). This is consistent with the lower heritability of pelvic girdle lengths in the HEL × BYN than in HEL × RYT cross with contributions from many small effect loci. In the HEL × PYÖ cross phenotypic variation was comparable to that in the HEL × BYN cross (SD=0.012) with the mean relative spine length being slightly smaller (0.078 vs 0.091; Supplementary Table 1) and the mean for relative girdle length being slightly larger (0.175 *vs* 0.169; Supplementary Table 1).

Scanning for Pel deletions in full-genome sequence data

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

Candidate genes

Seven candidate genes or regulatory elements for pelvic reduction from the literature (Supplementary Table 1) were found in LGs with significant QTL (Fig. 2). *Hif1a* (Mudie *et al*. 2014), *Pel* (Chan *et al*. 2010) and *Pouf1* (Kelberman *et al*. 2009) are known to regulate the expression of *Pitx1,* whereas four genes (*Fgf8*, *Wnt8c*, *Wnt8b* and *Hoxd9*) are involved in the pelvic fin/hindlimb development downstream of *Pitx1* (Don *et al.* 2012, Tanaka *et al*. 2005). However, next to *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 trait values (but not when analysing absolute trait values; Table 1 and Supplementary Fig. 4).

Proof of concept simulations

Allele frequencies for all QTL in the marine and focal freshwater populations showed no obvious trends prior to local adaptation, showing that the 10,000 generation burn-in was adequate (Supplementary Fig. 5). In the empirical data of marine populations, the slope of regression of linearised F_{ST} against geographic distance (point distance in km) was 2.13e-7 (95% of the bootstrap replicates between 2.05e-7 and 2.2e-7) and 1.53e-8 (95% of the bootstrap replicates between 1.10e-8 and 1.96e-8) for nine- and three-spined

sticklebacks, respectively. Thus, IBD in this data set was 13.9 times nine-spined sticklebacks compared to three-spined sticklebacks (Fig. 5a). When scaling the geographic distance in the simulations to match IBD in the empirical data, the geographic distance between the two most distant simulated marine populations (from which the focal freshwater populations were colonised from) equaled to 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), with the difference in IBD between the two species also close to that in the empirical data (the ratio between 264 and 352 km is 0.75). In the empirical data, genetic differentiation between freshwater habitats for populations < 400 km from each other (mean F_{ST} = 0.19 and 0.49 for three- and nine-spined sticklebacks, respectively) was also in par with the simulations (mean [across simulation replicates] F_{ST} = 0.21 and 0.58 for three- and nine-spined sticklebacks, respectively).

The relationship between marine-freshwater F_{ST} and freshwater-freshwater F_{ST} in the simulated data depended on both the species and genetic architecture (Fig. 5b). For instance, when the genetic architecture includes one additive large effect locus (architecture A), this locus was often (65% of replicates) involved in parallel evolution in three-, but not in nine-spined sticklebacks (3% of the replicates). In 20% of the replicates (for both species) the freshwater allele for this locus 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 sticklebacks species, particularly for the locus with the largest effect size (six; 20% and 80%, for nine and three-spined sticklebacks, respectively; Fig. 5b, c) with local adaptation also occurring in 57% of the replicates in nine-spined sticklebacks (18% in three-spined sticklebacks; Fig. 5b, c). For the locus with effect size of four, 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 the alleles locally adapted to freshwater being recessive; architecture C) this locus was less likely to be involved in parallel evolution in the three-spined sticklebacks (48%), compared to architecture A, with results for the nine-spined sticklebacks being similar to architecture A (6% parallel evolution; Fig. 5b & c). Thus, particularly when a single additive large effect locus is responsible for freshwater adaptation, parallel evolution is an expected outcome in three- but not in nine-spined sticklebacks, at relatively short (<400 km) geographical distances, when selection optima in the freshwater populations are the same.

Discussion

The most salient findings of this study include demonstration that the genetic architecture of pelvic reduction in nine-spined sticklebacks appears to be far more variable than that in three-spined sticklebacks. Our results, together with findings of earlier studies in nine-spined sticklebacks (Shapiro *et al.* 2006; Shikano *et al.* 2013), suggest that multiple

genomic regions (11 QTL, of which ten are novel to this study) are responsible for pelvic reduction in this species across its distribution range. This is in stark contrast to three-spined sticklebacks, in which the vast majority of cases of pelvic reduction can be traced back to the *Pitx1/Pel* deletion, particularly among marine-freshwater ecotype pairs (Chan *et al.* 2010; Xie et al. 2019). As shown by our simulations anchored to mimic the contrasting demographic and population genetic structures of nine- and three-spined sticklebacks, adaptation based on parallel genetic changes is to be expected when independent freshwater populations have access to the same ancestral standing genetic variation. This is consistent with the large body of studies showing that parallel evolution among pairs of marine-freshwater ecotypes in three-spined sticklebacks is commonly due to the repeated use of the same ancestral variation from a global pool of standing genetic variation in the sea, both at the global scale (Jones *et al.* 2012) and also more locally (Colosimo *et al.* 2005; Jones *et al.* 2012; Terekhanova *et al.* 2014, 2019; Nelson & Cresko, 2018). However, given the 13.9 times stronger IBD in the sea, and 2.4 times stronger genetic differentiation (as measured by F_{ST}) between freshwater populations in the nine-spined stickleback compared to the three-spined stickleback, one can expect that the possibilities for local adaptation from standing genetic variation in nine-spined sticklebacks to be much more limited than in three-spined sticklebacks. Indeed, our results are consistent with strong IBD limiting the spread of freshwater adapted alleles across larger geographic areas in the nine-spined stickleback. This was also confirmed by simulations where strong IBD in the sea and high population structuring between

freshwater populations did not allow locally adapted alleles to reach freshwater populations restricting local adaptation and thus also genetic parallelism. However, it should be noted that pelvic reduction in three-spined sticklebacks is also due to independently derived mutations in many different freshwater populations made possible by the high mutation rate of the *Pitx1/Pel* deletion (Xie *et al.* 2019), in contrast to the otherwise strong support for genetic parallelism among marine-freshwater pairs of three-spined stickleback populations (Colosimo *et al.* 2005; Jones *et al.* 2012; Terekhanova *et al.* 2014, 2019; Nelson & Cresko 2018). While it is clear that most of the *Pel* deletions in three-spined sticklebacks are independently derived, it is noteworthy that still predominantly *Pel* is responsible for pelvic reduction in three-spined sticklebacks, when almost all studied nine-spines stickleback populations 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 the most favoured by selection (when available as standing genetic variation), and given enough gene flow, they are repeatedly used in almost all instances of pelvic reducing in the marine-freshwater study system. In nine-spined sticklebacks, however, freshwater adaptation seems to be limited to less optimal solutions (e.g. due to lack of standing genetic variation stemming from strong IBD in the sea and population structuring), for instance allowing only spine (but not girdle) reduction in the HEL × RYT cross.

Can isolation by distance and population structuring limit parallel evolution?

The nine-spined stickleback has several key attributes that could explain the high observed level of convergent genetic basis for evolution of pelvic reduction at relatively small spatial geographic scales. In addition to generally smaller N_e and stronger IBD in the sea in nine- than in three-spined sticklebacks, the populations of the former species also show exceptionally high degree of genetic differentiation even among adjacent ponds (e.g. Pyöreälampi and Rytilampi; DeFaveri *et al*. 2012; Shikano *et al*. 2010). In our empirical population genomic data, F_{ST} reached >0.8 (linearised F_{ST}<4) between these two ponds situated 1.5 km from each other (Fig. 5a). All these factors likely limit homogenization of locally adapted alleles among the different freshwater populations through gene flow, even at extremely short geographic scales. As proof of concept, the simulations (where levels of IBD and genetic differentiation between freshwater populations are comparable to those in empirical data) show that parallel evolution can, under certain circumstances readily be expected for three-spined sticklebacks, but not for nine-spined sticklebacks. We also often observed the fixation of freshwater adapted alleles in one focal freshwater population, while no such alleles reached the other freshwater population leading to non-parallel local adaptation (Fig. 5), particularly in the nine-spined stickleback (up to 57% of the replicates for architecture B). This is consistent with our re-analyses of previously published nine-spined stickleback morphological data from Fennoscandia and the White Sea area (Fig. 1) showing large variation in the degree of pelvic reduction among the sampled pond populations, with many populations

displaying none or only marginally reduced pelvic apparatuses (e.g. populations ABB, KAR, NAV and PYÖ in Fig. 1 and Supplementary Table 1). Furthermore, our study demonstrated that genetic loci controlling pelvic reduction in three independent QTL-crosses between marine (fully developed pelvis) and pond (varying degrees of pelvic reduction) individuals were different, and that pelvic reduction in nine-spined sticklebacks can be achieved by heritable variation in at least 11 different QTL-regions of varying effect sizes (Fig. 2, Table 1 and Supplementary Table 2; excluding the previously known QTL on LG4 found in Alaska [Shapiro *et al.* 2009] and the possible unknown large effect QTL in the BOL and MAS populations). In the Swedish pond Bynästjärnen showing strong reduction in pelvic spine lengths (but not in pelvic girdle lengths), spine and girdle lengths are independently controlled by different QTL. This is an unusual finding as previous studies focused on pelvic reduction have only found large effect QTL that simultaneously affects both girdle and spine lengths (Cresko *et al*. 2004; Shapiro *et al*. 2006; Coyle *et al*. 2007; Chan *et al*. 2010, Shikano *et al*. 2013; Xie *et al*. 2019). Only one small effect QTL region (LG1; relative trait values; Table 1 and Supplementary Fig. 3) was shared between any two crosses (HEL × BYN and HEL × PYÖ). However, due to large size of identified QTL regions, it is not possible to know whether these are due to the same or different underlying causal mutations (we here consider this as a single QTL region). In the HEL × PYÖ cross, where both spines and girdles are only marginally smaller (relative to body size) than in the sea, both traits were also heritable (when crossed with a marine individual with a complete pelvis), but to a lesser extent than in the

HEL × RYT and HEL × BYN crosses. Also here our analyses suggest that spine and girdle length are independently controlled by small effect additive QTL.

Interestingly, five out of seven significant QTL in HEL × BYN and HEL × PYÖ crosses indicated that the QTL effect was segregating in grandfather (from the pond). This suggests that, despite the low levels of genetic variation in these ponds (Shikano *et al*. 2010; Bruneaux *et al*. 2013), some additive genetic variance is retained, thus potentially also allowing future evolution of the pelvic apparatus in these ponds. In Rytilampi, where LG7 explained ~80% of the phenotypic variation in both spine and girdle lengths, no small effect QTL were detected, despite that such loci were found in the other two crosses. This could be due to chance, with local adaptation being restricted to whatever freshwater adapted alleles that were available when the populations were founded (Merilä 2014; this study). Alternatively, this could be due to limitations to how many and which locally adapted alleles can simultaneously coexist in a population. For instance, if a single large effect QTL already leads to near optimal phenotype, any additional minor or medium effect loci could lead individuals further from the optimum, i.e. maladaptation (Bolnick *et al*. 2018). If so, selection is expected to prune such variants leaving only large-effect QTL in the population.

Pelvic reduction: comparison of nine- and three-spined sticklebacks

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

study), but we do not know to what extent the underlying mutations are identical by descent. By scanning whole-genome sequence data, we found a deletion in the genomic region expected to contain *Pel* (LG7) only in Rytilampi (where pelvic reduction maps to this gene) out of the 27 populations screened. Similar to marine-freshwater pairs of three-spined sticklebacks, we would have expected *Pitx1/Pel* deletions to exists in several other nine-spined stickleback pond populations too, most notably in the MAS and BOL populations where phenotypic data suggest large effect QTL to be responsible for pelvic reduction (Fig. 1 and Fig. 4). Since no deletions were found in the genomic region spanning *Pel* and *Pitx1* in MAS and BOL, *Pitx1*/*Pel* is likely not responsible for the pelvic reduction in these two populations. This provides further evidence that the genomic basis of pelvic reduction in the nine-spined stickleback is more heterogeneous than that in the three-spined stickleback - the loci (most likely major effect) responsible for pelvic reduction in these ponds are yet to be identified.

Another explanation to high heterogeneity of the genetic mechanisms underlying pelvic reduction in nine-spined sticklebacks is within habitat environmental variation resulting in different selection optima in different pond populations (cf. Stuart *et al.* 2017; Thompson *et al.* 2019). Indeed, the small differences between pelvic reduced phenotypes in our study (e.g. in BYN and MAS/BOL spines are completely absent while in RYT they are only strongly reduced; Fig. 1, Supplementary Table 1 and Supplementary Fig. 1) could, in itself, indicate different selection optima in the different populations (Stuart *et al*. 2017; Thompson *et al*. 2019). It could be possible to use available phenotypic data to

estimate the phenotypic optima of pelvic morphology (hypersphere) in each of the populations using Fisher's geometric model (Stuart *et al.* 2017; Thompson *et al*. 2019), where a strong overlap would suggest a higher probability of genetic parallelism (Thompson *et al.* 2019). However, this assumes that the populations have access to exactly the same ancestral variation and are free to evolve and reach their optima, which, as we argue here (with support from simulations), is most likely not the case. Without detailed environmental data or direct estimates of strength selection on pelvis phenotypes, disentangling the effects gene flow and within habitat environmental variation (assuming this leads to non-parallel angles of selection) is not possible (Stuart *et al.* 2017).

From our simulations, genetic parallelism in the three-spines stickleback was the most common outcome when the trait was controlled by a single additive large effect locus, and less so for loci with additive effect sizes <6 (i.e. loci where a homozygote for the freshwater adapted allele would increase the phenotype to 60% of the optimal value in freshwater habitats) or when the (large effect) freshwater adapted allele was recessive (parallelism was always rare in nine-spined sticklebacks; Fig. 5c, d). This may partly explain why the *Pitx1/Pel* deletions responsible for pelvic reduction also in three-spined sticklebacks are mostly due to repeated independent deletions (and not genetic parallelism; Xie *et al.* 2019). In a contrast, the additive Ectodysplasin gene (*EDA*) controlling lateral armour plate numbers in three-spined sticklebacks is a classical example of local adaptation from ancestral standing variation (Colosimo *et al.*, 2005; Schluter & Conte 2009). In a recent simulation study, Thompson *et al.* (2019) showed

that genetic parallelism from standing genetic variation rapidly declines as selection starts to change from fully parallel (optima angle of 0°) to divergent (optima angle of 180°), especially when the trait is polygenic. However, despite that selection was fully parallel in our simulations we failed to observe strong genetic parallelism for smaller effect loci (with allelic effects < 6) in both three- and nine-spined sticklebacks (Fig. 5a, c). This suggests that the effects of the underlying genetic architecture on parallelism (in conjunction with some IBD and population structuring) can be independent of the angle of optimal phenotypes between two habitats.

While the evidence for genetic parallelism at large geographical scales in the marine-freshwater stickleback model system is extensive (Colosimo *et al.* 2005; Jones *et al.* 2012; Terekhanova *et al.* 2014, 2019; Nelson & Cresko 2018; Fang et al. 2019), the level of parallelism in lake-stream and pelagic-benthic ecotype pairs of three-spined sticklebacks is much more diverse (Peichel *et al*. 2001, 2017; Conte *et al*. 2012, 2015; Stuart *et al*. 2017). For instance, Conte *et al*. (2015) found that among benthic-limnetic three-spined stickleback pairs from Paxton and Priest lakes (Vancouver Island, BC, Canada) 76% of 42 phenotypic traits diverged in the same direction, whereas only 49% of the underlying QTL evolved in parallel in both lakes. For highly parallel traits in two other pairs of benthic-limnetic sticklebacks, only 32% of the underlying QTL were shared (Conte *et al*., 2012). Similarly, Stuart et al. (2017) found that among 11 independent evolutionary replicate pairs of lake-stream three-spined stickleback populations (Vancouver Island, BC, Canada) indeed both within habitat variation and constraints to

gene flow contributed to the observed variation in levels of phenotypic parallelism. Different lakes and streams do not likely have similar access to the same global pool of ancestral variation as pairs of marine-freshwater three-spined sticklebacks (where gene flow in the sea is high) consistent with the notion that more heterogeneous access to ancestral variation can indeed limit genetic parallelism. This is also true for one example of marine-freshwater three-spined stickleback divergence among isolated insular freshwater populations in the Haida Gwaii archipelago off the northern Pacific coast of Canada (Deagle *et al.* 2013). Here, similarly to the nine-spines sticklebacks in this study, several freshwater populations did not display any reduction in pelvic armour. However, those populations that were fully plated, were also genetically more similar to adjacent marine individuals, suggesting that recent marine-freshwater admixture and/or selection favouring plated freshwater individuals could explain this pattern. Other morphological traits, such as lateral plating (*EDA*) were instead more consistent with the same global variants being repeatedly re-used at the regional scale. Thus, with respect to access to ancestral variation available for freshwater adaptation, nine-spined sticklebacks are likely closer to the three-spined stickleback lake-stream and benthic-limnetic study systems, than the three-spined stickleback marine-freshwater study system, with the one notable exception above (lake-stream three-spined sticklebacks, where genetic structuring also is high).

Epistatic control of pelvic reduction?

For a trait that is controlled by three additive QTL of equal effect sizes, 1/64 (1.6%) of the F_2-individuals are expected to show the extreme phenotypes of the parents (Klug & Cummings 2018). This is close to what was observed in the HEL × BYN cross with respect to pelvic spine lengths, as only 4/306 individuals (1.3%) of the F_2 individuals in this cross lacked spines. However, in such crosses, the phenotypes are also expected to be normally distributed (with a mean close that of the mean for the parents; Klug & Cummings 2018). This was not the case for spine length in the HEL × BYN cross, where the bulk of the phenotypic values was centered around the mean, but with a long tail of individuals with strongly reduced pelvic spines (Fig. 1, Fig. 3 and Supplementary Table 1). This skew in the phenotype distribution could be caused by epistatic interactions among loci controlling pelvic spine length (Wolf et al. 2000). For instance, a threshold number of alleles (from different QTL) might be needed to complete pelvic reduction. Consistent with this hypothesis, the phenotypic data for spine lengths from the HEL × BYN cross suggests that complete spine reduction was most likely when the allele responsible for spine reduction for the LG6 QTL was combined with at least one allele (causing spine reduction) from the LG15 and LG16 QTL (Fig. 3b and Supplementary File 3). If a threshold number of alleles are needed for complete pelvic reduction, this could also explain how standing genetic variation in the sea is maintained, as the necessary multi-locus genotypes that cause sub-optimal phenotypes in the sea are rarely formed (due to overall lower frequencies of spine reducing alleles in the sea). This is analogous to “epistatic shielding” that can contribute to the persistence of disease alleles

in populations (Phillips & Johnsson 1998; Phillips 2008). Consistent with this possibility, LG6 of the grandmother of the HEL × BYN cross (from the sea) was polymorphic for the pelvic spine QTL effect (Fig. 2) – evidently a single pelvic reducing allele alone in this female was not enough to cause any pelvic reduction at all (this female had a complete pelvis).

Only a small proportion of the individuals with the three-way bi-allelic genotypes (*viz*. 011, 010 or 001; Fig 3b) that were associated with spine reduction, had fully reduced spines. Since Bynästjärnen is fixed for the reduced spine phenotype (at the time of sampling; Fig 1), this likely means that spine reduction also depends on either genotype × environment interactions or epistatic interactions involving more than the three significant QTL detected in our study. It is possible to detect epistatic QTL without individual effects, but the power to detect such QTL depends on the number of individuals in the genotype classes on which the analysis is based (Carlberg & Haley 2004). Thus, given the small number of individuals with reduced spines in the HEL × BYN F_2 generation, this was not possible.

Candidate genes

While the QTL peak for the *Pitx1*/*Pel* region in the HEL × RYT cross was narrow, this was not the case for the other QTL we detected. Hence, due to the large QTL regions detected by four-way single-mapping analyses, it was not meaningful to perform gene ontology enrichment (GO) analyses – the QTL regions would have contained possibly

thousands of genes. Instead, we searched the literature for known candidate genes for pelvic reduction and found three (excluding *Pitx1*/*Pel*) that were clearly contained within the identified QTL regions (Fig. 2 and Supplementary Table 1). Due to the aforementioned large QTL regions, these can only be considered as highly putative candidate genes for pelvic reduction, and will not be discussed further. However, further studies of pelvic reduction might find these candidates worthy of attention.

Conclusions

Our results show that the repeated parallel reduction of the pelvic apparatus in freshwater populations of nine-spined sticklebacks is due to a diverse set of genetic changes: only one small effect QTL for pelvic reduction was shared between the three experimental crosses in this study. In one cross, pelvic reduction was caused by the same large effect QTL that has also been associated with pelvic reduction in most three-spined stickleback populations studied so far. In contrast, in the two other crosses the genetic basis of pelvic reduction was found to be polygenic, and map to different chromosomes. In addition to these, yet another large effect QTL different from the *Pitx1*/*Pel* locus likely segregates in one nine-spined stickleback population waiting to be identified. Evidence was also found that epistatic interactions play a role in determining pelvic phenotype in one of the crosses. The results also shed light on the possible drivers of the observed genetic heterogeneity underlying pelvic reduction: as shown by simulations, heterogenous genetic architectures are more likely to emerge in presence of strong isolation by distance

and limited gene flow. This together with the demonstrated stronger IBD and more profound population structuring in nine- than three-spined sticklebacks, suggests that parallel genetic responses to similar selection pressures are far more common in the latter, than in the former species. This reinforces the role of the nine-spined sticklebacks as a useful model system, along side the three-spined stickleback, to study adaptive evolution in the wild. Furthermore, since the population demographic characteristics of nine-spined sticklebacks are similar to small and endangered species/populations, it is also likely to be a well-suited model to study genetics of adaptation in populations of conservation concern.

Acknowledgments

We thank Alexandre Budria, Abigel Gonda, Chris Eberlein, Gabor Herczeg, Sami Karja, Heini Natri and Ismo Rautiainen for help in fish collection and/or rearing. Thanks are also due to Kirsi Kähkönen and Miinastiina Issakainen for help in laboratory, and Oulanka Biological Station (University of Oulu) for logistic support. We are grateful for the computing resource support from CSC – the Finnish IT Center for Science Ltd administered by the Ministry of Education and Culture, Finland. B.G. thanks support from CAS Pioneer Hundred Talents Program and the National Natural Science Foundation of China (31672273). This work is supported by the Academy of Finland (grant numbers, 129662, 134728 and 218343 to J.M.), a grant from Helsinki Institute of Life Science (HiLIFE to JM) and a personal grant to Petri Kemppainen from the Finnish

Cultural Foundation (#00190489).

References

Arendt, J., & Reznick, D. (2007). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology and Evolution*, *23*(1), 26–32. https://doi.org/10.1016/j.tree.2007.09.011

Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., … Schlötterer, C. (2019). Genetic redundancy fuels polygenic adaptation in Drosophila. *PLoS Biology*, *17*(2), 1–31. https://doi.org/10.1371/journal.pbio.3000128

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

Barrett, R. D. H., & Schluter, D. (2007). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Becker, Walter, A. (1984). *Manual of quantitative genetics* (Fourth ed.). Washington: Academic enterprises.

Bell, M. A. and Foster, S. A. (Eds) (1994) The Evolutionary Biology of the Threespine Stickleback, Oxford, UK, Oxford University Press.

Bell, M. A., Khalef, V., & Travis, M. (2007). Directional asymmetry of pelvic vestiges in threespine stickleback. Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution, 308B, 189–199. https://doi.org/10.1002/jez.b

Bell, M. A., Orti, G., Walker, J. A., & Koenings, J. P. (1993). Evolution of pelvic reduction in threespine stickleback fish: a test of competing hypotheses. *Evolution*, *47*(3), 906–914. https://doi.org/10.2307/2410193

Blouw, D. M., & Boyd, G. J. (1992). Inheritance of reduction, loss, and asymmetry of the pelvis in Pungitius pungitius (ninespine stickleback). *Heredity*, *68*(33). Retrieved from https://doi.org/10.1038/hdy.1992.4

Bolnick, D. I., Barrett, R. D. H., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non) Parallel Evolution. *Annual Review of Ecology, Evolution, and Systematics*, *12*(41), 303–330.

Bruneaux, M., Johnston, S. E., Herczeg, G., Merilä, J., Primmer, C. R., & Vasemägi, A. (2013). Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. *Molecular Ecology*, *22*, 565–582. https://doi.org/10.1111/j.1365-294X.2012.05749.x

Carlborg, Ö., & Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nature*, *5*, 618–625.

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

Cohan, F. M. (1984). Can uniform selection retard random genetic divergence between isolated conspecific populations? *Evolution*, *38*(3), 495. https://doi.org/10.2307/2408699

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

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biology, 2(5), 635–641. https://doi.org/10.1371/journal.pbio.0020109

Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., … Peichel, C. L. (2015). Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. Genetics, 201(3), 1189–1200. https://doi.org/10.1534/genetics.115.182550

Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. 279, 5039–5047. https://doi.org/10.1098/rspb.2012.2146

Coyle, S. M., Huntingford, F. A., & Peichel, C. L. (2007). Parallel vevolution of Pitx1 underlies pelvic reduction in Scottish Threespine Stickleback (Gasterosteus aculeatus). Genetics *98*(6), 581–586. https://doi.org/10.1093/jhered/esm066

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., … Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. JOURNAL? *101*(16), 6050–6055.

DeFaveri, J., Shikano, T., Ghani, N. I. A., & Merilä, J. (2012). Contrasting population structures in two sympatric fishes in the Baltic Sea basin. *Marine Biology*, *159*, 1659–1672. https://doi.org/10.1007/s00227-012-1951-4

DeFaveri, J., & Merilä, J. (2013) Variation in age and size in Fennoscandian three-spined sticklebacks (*Gasterosteus aculeatus*). PLoS One 8(11): e80866. https://doi.org/10.1371/journal.pone.0080866

DeFaveri, J., Shikano, T., & Merilä, J (2014) Geographic variation in age structure and longevity in the nine-spined stickleback (*Pungitius pungitius*). PLoS One 9(7): e102660. https://doi.org/10.1371/journal.pone.0102660

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

Elmer, K. R., & Meyer, A. (2011). Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends in Ecology and Evolution*, *26*(6), 298–306. https://doi.org/10.1016/j.tree.2011.02.008

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

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

Giles, N. (1983). The possible role of environmental calcum levels during the evolution of phenotypic diversity in Outer Hebridean populations of the Three-spined stickleback, *Gasterosteus aculeatus*. *Journal of Zoology*, *199*(4), 535–544. https://doi.org/10.1111/j.1469-7998.1983.tb05104.x

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

Herczeg, G., Turtiainen, M., & Merilä, J. (2010). Morphological divergence of North-European nine-spined sticklebacks (*Pungitius pungitius*): *Biological Journal of the Linnean Society*, 101(2), 403–416.

Hermisson, J., & Pennings, P. S. (2017). Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, *8*(6), 700–716. https://doi.org/10.1111/2041-210X.12808

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

Hill, W. G., & Kirkpatrick, M. (2010). What animal breeding has taught us about evolution. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 1–19. https://doi.org/10.1146/annurev-ecolsys-102209-144728

Karhunen, M., Merilä, J., Leinonen, T., Cano, J. M., & Ovaskainen, O. (2013). driftsel: An R package for detecting signals of natural selection in quantitative traits. *Molecular Ecology Resources*, *13*(4), 746–754. https://doi.org/10.1111/1755-0998.12111

Kemppainen, P., Knight, C. G., Sarma, D. K., Hlaing, T., Prakash, A., Maung Maung, Y. N., … Walton, C. (2015). Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. *Molecular Ecology*, *15*(5), 1031–1045. https://doi.org/10.1111/1755-0998.12369

Klepaker, T., Østbye, K., & Bell, M. A. (2013). Regressive evolution of the pelvic complex in stickleback fishes: a study of convergent evolution. *Evolutionary Ecology Research*, *15*, 413–435.

Klug, W. S., & Cummings, M. R. (2018). Concepts of genetics. *Pearson Education, Inc, New Jersey*

Laine, V. N., Shikano, T., Herczeg, G., Vilkki, J., & Merilä, J. (2013). Quantitative trait loci for growth and body size in the nine-spined stickleback *Pungitius pungitius* L. *Molecular Ecology*, *22*(23), 5861–5876. https://doi.org/10.1111/mec.12526

Leinonen, T., Cano, J. M., & Merilä, J. (2011). Genetics of body shape and armour variation in threespine sticklebacks. *Journal of Evolutionary Biology*, *24*(1), 206–218. https://doi.org/10.1111/j.1420-9101.2010.02161.x

Li, H. (2018). Sequence analysis Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows – Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Sam, T. (2009). The Sequence Alignment / Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

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

Li, Z., Guo, B., Yang, J., Herczeg, G., Gonda, A., Balázs, G., … Merilä, J. (2017). Deciphering the genomic architecture of the stickleback brain with a novel multilocus gene-mapping approach. *Molecular Ecology*, *26*(6), 1557–1575. https://doi.org/10.1111/mec.14005

Li, Z., Kemppainen, P., Rastas, P., & Merilä, J. (2018). Linkage disequilibrium clustering-based approach for association mapping with tightly linked genomewide data. *Molecular Ecology Resources*, *18*(4), 809–824. https://doi.org/10.1111/1755-0998.12893

MacKay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: Challenges and prospects. *Nature Reviews Genetics*, *10*(8), 565–577. https://doi.org/10.1038/nrg2612

Merilä, J. (2013). Nine-spined stickleback (*Pungitius pungitius*): An emerging model for evolutionary biology research. *Annals of the New York Academy of Sciences*, *1289*(1), 18–35. https://doi.org/10.1111/nyas.12089

Merilä, J. (2014). Lakes and ponds as model systems to study parallel evolution. *Journal of Limnology*, *73*(1), 33–45. https://doi.org/10.4081/jlimnol.2014.805

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

transcriptional response to hypoxia Sharon. *Cell Cycle*, *13*(24), 3878–3891. https://doi.org/10.4161/15384101.2014.972889

Nelson, T. C., & Cresko, W. A. (2018). Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters*, *2*(1), 9–21. https://doi.org/10.1002/evl3.37

Neuenschwander, S., Michaud, F., & Goudet, J. (2018). QuantiNemo 2: a Swiss knife to simulate complex demographic and genetic scenarios, forward and backward in time. Bioinformatics, 1, 1–3. https://doi.org/10.1093/bioinformatics/bty737

Peichel, C. L., & Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(20150486). https://doi.org/10.1098/rstb.2015.0486

Peichel, C. L., Nereng, K. S., Ohgi, K. A., Cole, B. L. E., Colosimo, P. F., Buerkle, C. A., … Kingsley, D. M. (2001). The genetic architecture of divergence between threespine stickleback species. *Nature* 414(20/27), 901–905.

Pembleton, L.W., Cogan, N.O.I., Forster, J.W. (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* 13(5), 946-952

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

Phillips, P. C., & Johnson, N. A. (1998). The population genetics of synthetic lethals. *Genetics*, *150*, 449–458.

Pritchard, J. R., & Schluter, D. (2001). Declining interspecific competition during character displacement: Summoning the ghost of competition past. *Evolutionary Ecology Research, 3*, 209–220.

Rastas, P. (2017). Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics*, *33*(23), 3726–3732. https://doi.org/10.1093/bioinformatics/btx494

Reid, S., Tibshirani, R., Friedman, J. (2016). A study of error variance estimation in lasso regression. *Statistica Sinica*, 26(1), 35-67. https://www.jstor.org/stable/24721190

Reimchen, T. E. (1983). Structural Relationships Between Spines and Lateral Plates in Threespine Stickleback (Gasterosteus aculeatus). *Evolution*, *37*(5), 931–946. https://doi.org/10.2307/2408408

Reist, J. D. (1980). Predation upon pelvic phenotypes of brook stikleback, *Culaea inconstancs*, by selected invertebrates. *Canadian Journal of Zoology*, 1980, 58(7), 1253-1258.

Roesti, M., Kueng, B., Moser, D., & Berner, D. (2015). The genomics of ecological vicariance in threespine stickleback fish. Nature Communications, 6, 1–14. https://doi.org/10.1038/ncomms9767

Rosenblum, E. B., Parent, C. E., & Brandt, E. E. (2014). The Molecular Basis of Phenotypic Convergence. Annual Review of Ecology, Evolution, and Systematics, 45, 203-226. doi:10.1146/annurev-ecolsys-120213-091851

Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences USA*, *106*(1), 9955–9962. https://doi.org/10.1073/pnas.0901264106

Shapiro, M. D., Bell, M. A., & Kingsley, D. M. (2006). Parallel genetic origins of pelvic reduction in vertebrates. *Proceedings of the National Academy of Sciences USA*, *103*(37), 13753–13758. https://doi.org/10.1073/pnas.0604706103

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

Shapiro, Michael D, Summers, B. R., Balabhadra, S., Aldenhoven, J. T., Miller, A. L., Cunningham, C. B., … Kingsley, D. M. (2009). The genetic architecture of skeletal convergence and sex determination in ninespine sticklebacks. *Current Biology, 19*(13), 1140–1145. https://doi.org/10.1016/j.cub.2009.05.029.The

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

Shikano, T., Shimada, Y., Herczeg, G., & Merilä, J. (2010). History vs. habitat type: Explaining the genetic structure of European nine-spined stickleback (Pungitius pungitius) populations. *Molecular Ecology*, *19*(6), 1147–1161. https://doi.org/10.1111/j.1365-294X.2010.04553.x

Sillanpää, M. J. (2011). On statistical methods for estimating heritability in wild populations. *Molecular Ecology*, 20, 1324–1332. https://doi.org/10.1111/j.1365-294X.2011.05021.x

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

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

Taggart, J. B., Hynes, R. A., & Ferguson, A. (1992). A simplified protocol for routine total DNA isolation from salmonid fishes. *Journal of Fish Biology*, *40*, 963–965.

Tanaka, M., Hale, L. A., Amores, A., Yan, Y., Cresko, W. A., Suzuki, T., & Postlethwait, J. H. (2005). Developmental genetic basis for the evolution of pelvic fin loss in the pufferfish Takifugu rubripes. *Developmental Biology*, *281*, 227–239. https://doi.org/10.1016/j.ydbio.2005.02.016

Terekhanova, N. V, Barmintseva, A. E., Kondrashov, A. S., Bazykin, G. A., & Mugue, N. S. (2019). Architecture of parallel adaptation in ten lacustrine threespine stickleback populations from the White Sea area. Genome Biology and Evolution, 11(9), 2605–2618. https://doi.org/10.1093/gbe/evz175

Terekhanova, N. V, Logacheva, M. D., Penin, A. A., Neretina, T. V, Barmintseva, A. E., Bazykin, G. A., … Mugue, N. S. (2014). *Fast Evolution from Precast Bricks : Genomics of Young Freshwater Populations of Threespine Stickleback Gasterosteus aculeatus*. *10*(10). https://doi.org/10.1371/journal.pgen.1004696

Thompson, K. A., Osmond, M. M., & Schluter, D. (2019). Parallel genetic evolution and speciation from standing variation. *Evolution Letters*, *0*(0), 1–13. https://doi.org/10.1002/evl3.106

Varadharajan, S., Rastas, P., Löytynoja, A., Matschiner, M., Calboli, F.C.F., Guo, B., … Merilä, J. (2019) A high-quality assembly of the nine-spined stickleback (Pungitius pungitius) genome. *Genome Biology and Evolution*, evz240, https://doi.org/10.1093/gbe/evz240

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

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370.

Wolf, J. B., Brodie, E. D., & Wade, M. J. (2000). *Epistasis and the Evolutionary Process*. Oxford: Oxford University Press.

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

Xu, S. (1996). Mapping quantitative trait loci using four-way crosses. *Genetics Research*, *68*, 175–181.

Xu, S. (2013a). Genetic mapping and genomic selection using recombination Breakpoint Data. *Genetics*, *195*, 1103–1115. https://doi.org/10.1534/genetics.113.155309

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

Yang, J., Guo, B., Shikano, T., Liu, X., & Merila, J. (2016). Quantitative trait locus analysis of body shape divergence in nine-spined sticklebacks based on high-density SNP-panel. *Scientific Reports*, *6*(26632). https://doi.org/10.1038/srep26632

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. 28(24), 3326–3328. 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 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 female F_1 (♀), the male F_1 (♂) 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	-

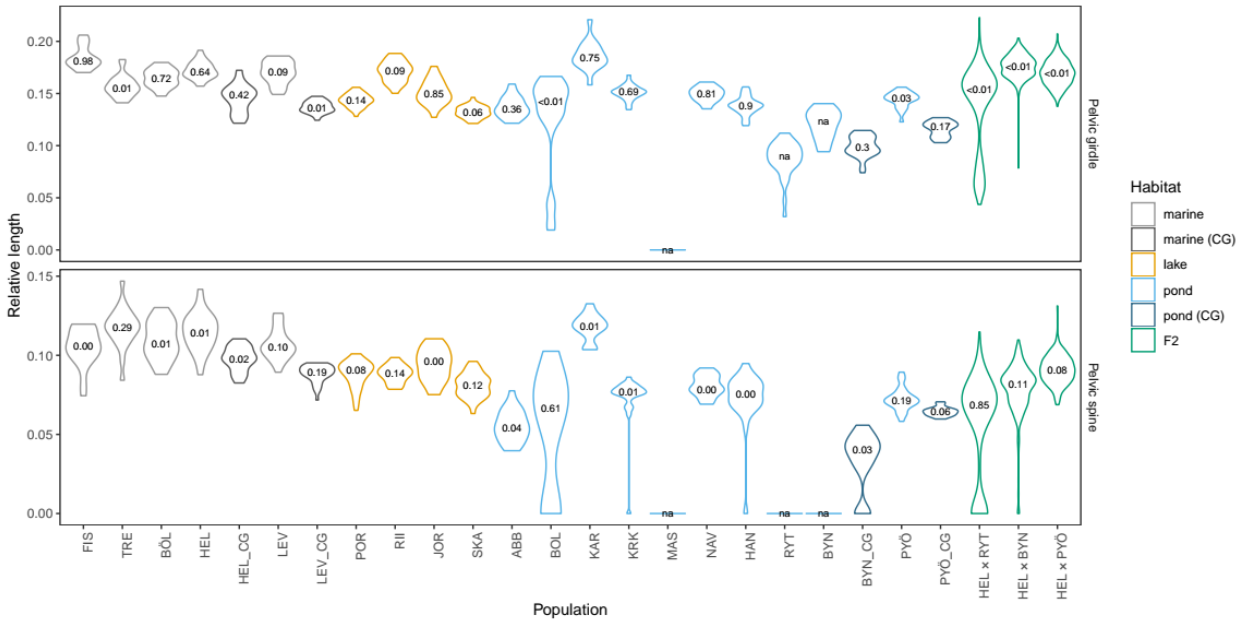

Figure 1 | Summary of previously published nine-spined stickleback pelvic phenotypes from the wild and common garden experiments. Violin plots depict relative (to standard body length) pelvic spine and girdle lengths according to population and habitat type. Numbers in the top panel show *P*-values for 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.

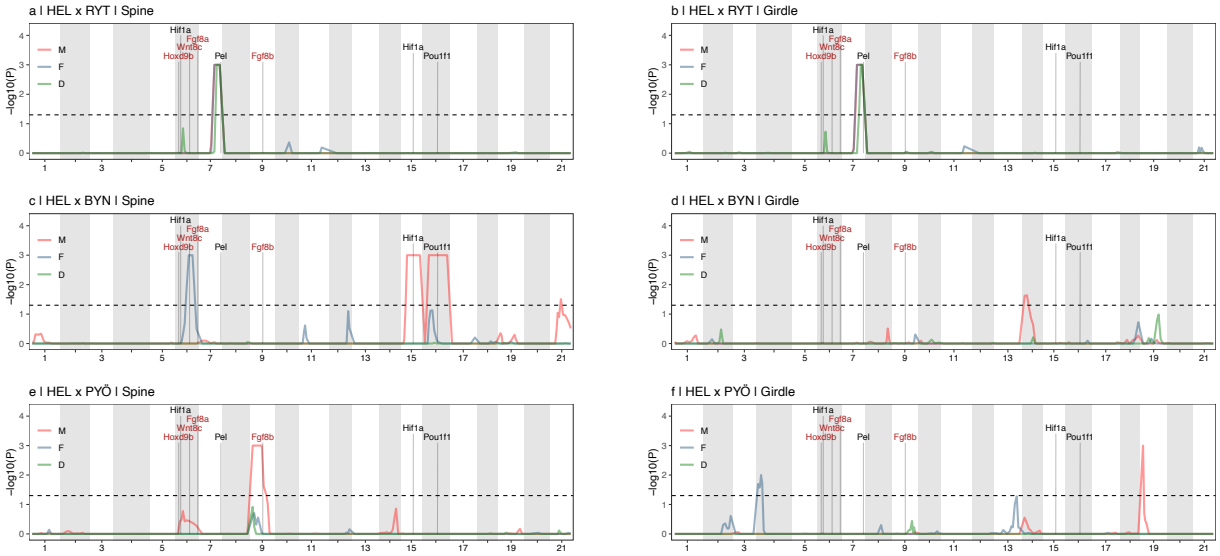

Figure 2 | Quantitative trait locus mapping of pelvic reduction in three independent stickleback crosses. Single-mapping four-way analyses of four morphological traits associated with pelvic reduction in (a-b) HEL × RYT cross, (c-d) HEL × BYN cross and (e-f) HEL × PYÖ cross. Shown are QTL for pelvic 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 α=0.05. Results are shown separately for alleles inherited from the male F_1 (M), the female F_1 (F) together with dominance effect (D) according to legend. Absence of a dominance indicates that the trait inheritance is additive, whereas a peak only for M or F indicates that the allelic effect was segregating in the grandfather or the grandmother, respectively (see Supplementary Information 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 Figures 4.

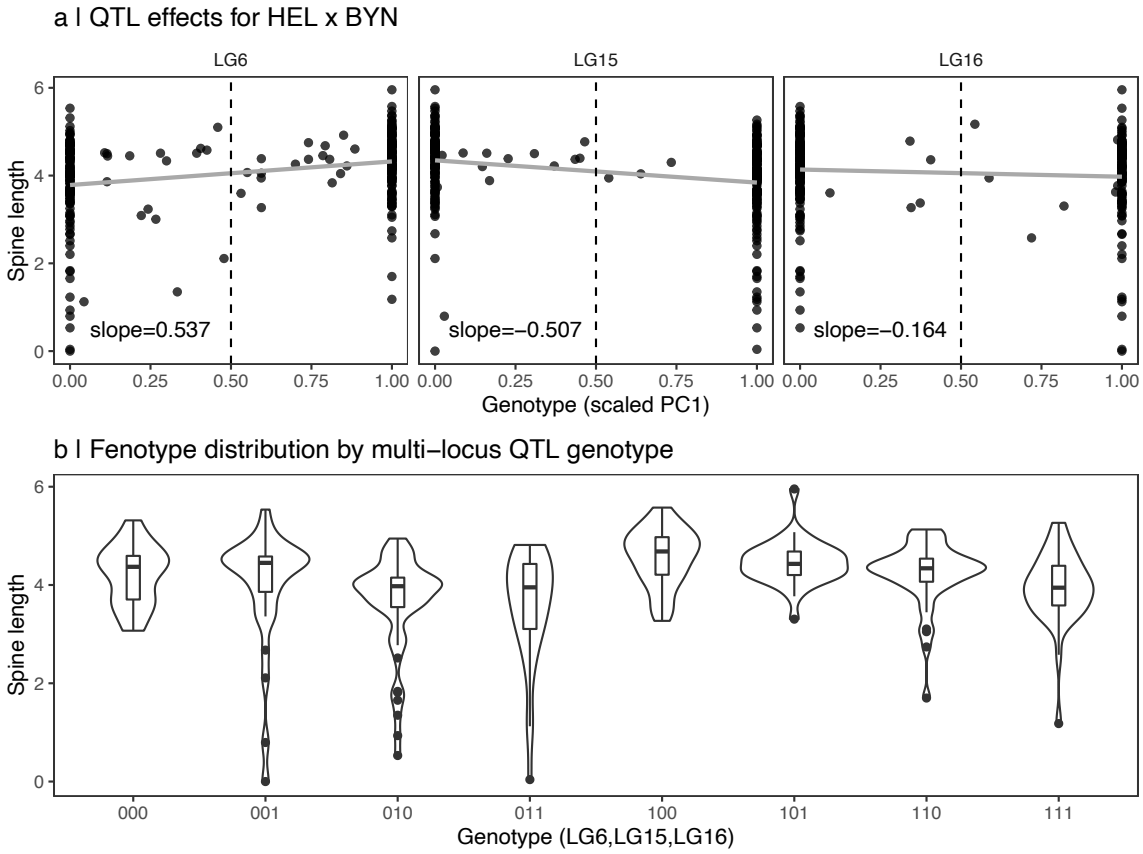

Figure 3 | Epistatic interactions in pelvic spine development for HEL x BYN cross. Shown are (a) individual genotype effects where the genotype is given by the first PC (scaled between 0 and 1) from the cluster of SNPs, that was the most significant for spine length on LG6, LG15 and LG16 (see Fig. 2), 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. In (b) distribution of spine lengths for all multi-locus genotypes from (a). The multi-locus genotype was based 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 “Allel 1”. The first digit for genotypes in (b) represent thus the genotype of the QTL on LG6, followed by LG15 and LG16, respectively.

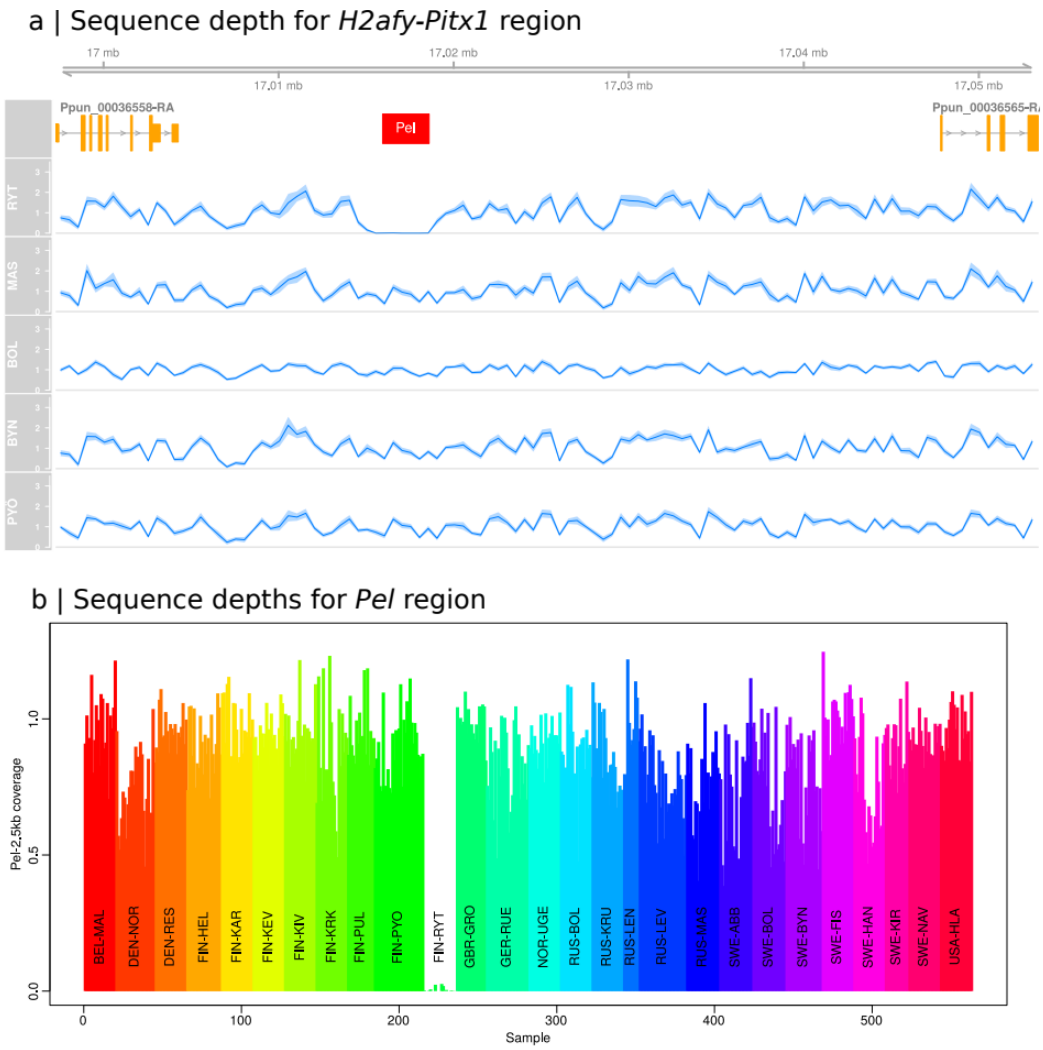

Figure 4 | Sequence depths for *H2afy-Pitx1* region. (a) Relative sequencing depth across *H2afy-Pitx1* intergenic region for 10 individuals from five different populations. Blue line and shading indicate population average and 95% confidence intervals, respectively. The 3.5kb deletion in RYT, seen as 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.

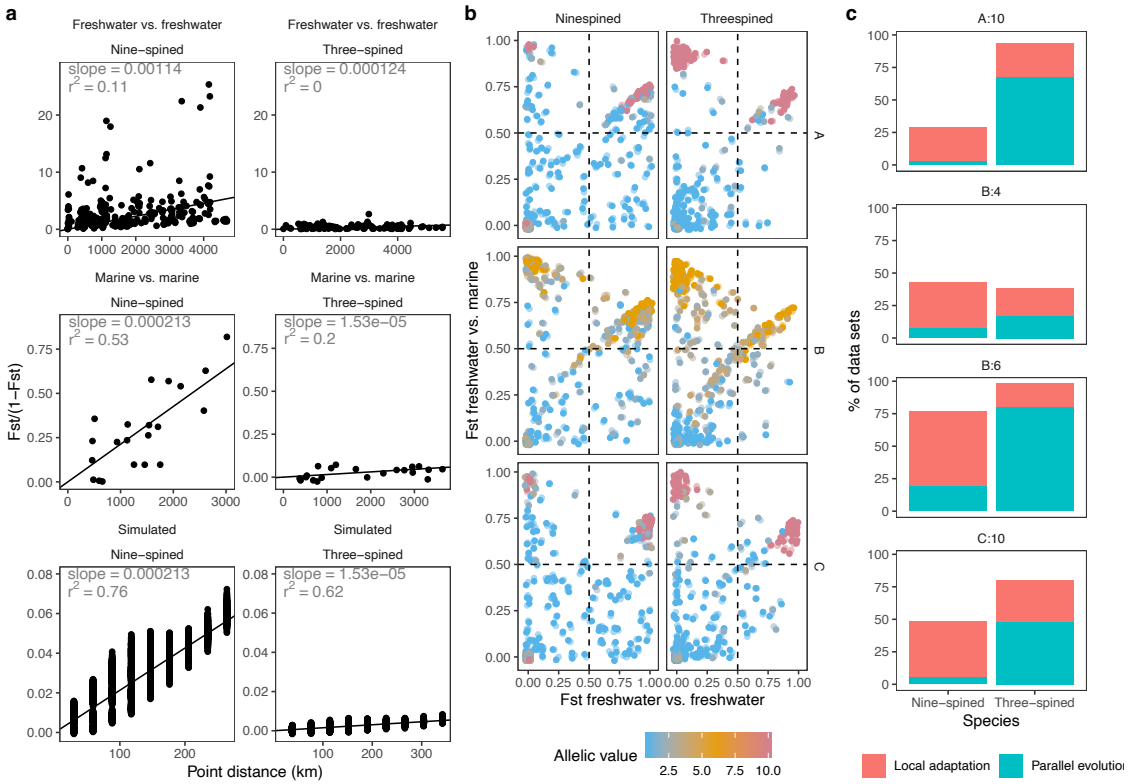

Figure 5 | Simulation results. (a) Shows linearised F_{ST} against geographic distance (IBD) for empirical and simulated data, with slope and squared Pearson's product moment correlation coefficient indicated. Geographic distance for simulated data (IBD in the sea) is scaled to match the slope for the IBD-plot in the sea in empirical data. (b) shows freswater-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. 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.