Genome-Wide DNA Methylation Profiling Reveals Epigenetic Adaptation of Stickleback to Marine and Freshwater Conditions

Artem V. Artemov,*[,1] Nikolai S. Mugue,[2] Sergey M. Rastorguev,[3] Svetlana Zhenilo,[1] Alexander M. Mazur,[1] Svetlana V. Tsygankova,[3] Eugenia S. Boulygina,[3] Daria Kaplun,[1] Artem V. Nedoluzhko,[3] Yulia A. Medvedeva,[1,4] and Egor B. Prokhortchouk*[,1]

[1]Department of Vertebrate Genomics and Epigenomics, Institute of Bioengineering, Research Center of Biotechnology RAS, Moscow, Russia

[2]Russian Federal Research Institute of Fisheries and Oceanography, Moscow, Russia

[3]National Research Center 'Kurchatov Institute', Moscow, Russia

[4]Department of Computational Biology, Vavilov Institute of General Genetics, RAS, Moscow, Russia

***Corresponding authors:** E-mails: artem.v.artemov@gmail.com; prokhortchouk@gmail.com.

Associate editor: Connie Mulligan

Abstract

The three-spined stickleback (*Gasterosteus aculeatus*) represents a convenient model to study microevolution—adaptation to a freshwater environment. Although genetic adaptations to freshwater environments are well-studied, epigenetic adaptations have attracted little attention. In this work, we investigated the role of DNA methylation in the adaptation of the marine stickleback population to freshwater conditions. DNA methylation profiling was performed in marine and freshwater populations of sticklebacks, as well as in marine sticklebacks placed into a freshwater environment and freshwater sticklebacks placed into seawater. We showed that the DNA methylation profile after placing a marine stickleback into fresh water partially converged to that of a freshwater stickleback. For six genes including ATP4A ion pump and NELL1, believed to be involved in skeletal ossification, we demonstrated similar changes in DNA methylation in both evolutionary and short-term adaptation. This suggested that an immediate epigenetic response to freshwater conditions can be maintained in freshwater population. Interestingly, we observed enhanced epigenetic plasticity in freshwater sticklebacks that may serve as a compensatory regulatory mechanism for the lack of genetic variation in the freshwater population. For the first time, we demonstrated that genes encoding ion channels KCND3, CACNA1FB, and ATP4A were differentially methylated between the marine and the freshwater populations. Other genes encoding ion channels were previously reported to be under selection in freshwater populations. Nevertheless, the genes that harbor genetic and epigenetic changes were not the same, suggesting that epigenetic adaptation is a complementary mechanism to selection of genetic variants favorable for freshwater environment.

Key words: **epigenetics, DNA methylation, stickleback, adaptation, evolution, water salinity, marine, freshwater, fish, phenotypic variation, ion pumps**

Article

Introduction

The three-spined stickleback (*Gasterosteus aculeatus*) is an important model in the study of adaptive evolution. The species can survive in environments with a large variety of salinity, such as river estuaries and the brackish waters of the Baltic Sea (McCairns and Bernatchez 2010; Guo et al. 2015; Konijnendijk et al. 2015). Successful colonization of different environmental conditions involves rapid adaptation to factors such as a sharp change in temperature, salinity, other predators, and parasites (Barrett et al. 2011; Lescak et al. 2015).

Isolated freshwater populations of sticklebacks are believed to originate by separation from the marine population. Many independent acts of river and lake colonization by marine sticklebacks have been reported (Jones et al. 2012), making it possible to study a wide variety of evolutionary trajectories of adaptation to freshwater. Parallel adaptation between independently formed freshwater populations to a new habitat occurs by increasing the frequency of certain freshwater alleles that preexist at low frequency in the marine population (Hohenlohe et al. 2012; Jones et al. 2012). Adaptation to heterogeneous environmental conditions, such as the Baltic Sea, and to the freshwater lakes and creeks have similar genomic mechanisms (Roesti et al. 2014; Guo et al. 2015). They operate within genomic loci that contain mainly regulatory sequences rather than protein-coding regions (Jones et al. 2012). The plasticity of gene expression in response to changing environmental conditions is likely to help the three-spined stickleback colonize a wide range of habitats (McCairns and Bernatchez 2010; Morris et al. 2014).

High salt conditions represent a major challenge for living organisms because they cause DNA double-strand breaks and cell senescence (Dmitrieva and Burg 2007; Dmitrieva et al. 2011).

It has been reported that marine invertebrates have numerous DNA breaks due to high salt conditions (Dmitrieva et al. 2006). It is therefore of fundamental interest to explore the mechanisms of how organisms adapt to different osmotic conditions.

Several studies revealed genetic aspects of *Gasterosteus aculeatus* freshwater adaptation. Terekhanova and colleagues (Terekhanova et al. 2014) defined divergence islands as regions with a significant shift of allele frequency between marine and freshwater populations, which are presumably under selection in freshwater populations. Some divergence islands harbored genes that are believed to be associated with freshwater adaptation, such as ion transporters. Notably, some of the divergence islands lacked genes, suggesting their regulatory potential as enhancers. Reduction of the pelvic apparatus characteristic to many freshwater populations was shown to be caused by the loss of a tissue-specific enhancer of the Pitx1 gene (Chan et al. 2010). Among the genomic loci, associated with freshwater adaptation reported in (Jones et al. 2012), noncoding regulatory changes appeared to be predominant.

Even though genetic components are important to adaptation to osmotic conditions, various organisms can switch between sea and freshwater habitats within one generation, suggesting that epigenetic mechanisms, such as histone modifications, regulation by microRNA (Rastorguev et al. 2016), and DNA methylation (Varriale 2014) might be involved in adaptation. There are several examples when variations in phenotypic traits are accompanied by minimum genetic diversity, but a significant level of epigenetic diversity. Noteworthy examples include bats (Liu et al. 2012, 2015) and plants (Gao et al. 2010; Yi et al. 2010). A study in Arctic charr (*Salvelinus alpinus*) revealed that seawater exposure induced changes in DNA methylation and peroxynitrite formation in gills (Norman et al. 2014). In a study aimed to understand the functionality of miRNA in stickleback adaptation to freshwater environments, ten miRNAs were found in divergence islands and two miRNAs contained SNPs with shifted allele frequencies between the freshwater and the marine populations (Rastorguev et al. 2016). DNA methylation was shown to be associated with phenotypic variability between complete and low lateral plate morphs in the freshwater stickleback population (Smith et al. 2015). Discovered differentially methylated regions (DMRs, mostly intergenic) were associated with genes having potentially adaptive functions, including cardiovascular development, growth, and neuromuscular development. However, the study was focused only on freshwater fish and epigenetic adaptation to water salinity has not been studied yet.

High levels of genetic variation within a population may impose a high probability of adaptation of the population to a new environment. However, isolation of a small population inevitably leads to decreased genetic diversity. Yet, to provide a substrate for natural selection in a new environment, a population should demonstrate a certain amount of diversity. Such diversity may be achieved by epigenetic variations, since genetically inherited propensity to phenotypic variability, even with no change in the mean phenotype, substantially increases fitness (Feinberg and Irizarry 2010). The idea of an evolutionary benefit of increased epigenetic variability is formulated as the epigenetic plasticity hypothesis.

In this work, we performed whole-genome DNA methylation profiling of marine and freshwater sticklebacks, as well as sticklebacks moved for 4 days from a marine environment to a freshwater environment and vice versa. We investigated the role of DNA methylation in the short-term (4 days of exposure to a foreign environment) and long-term (differences between marine and freshwater populations) adaptation to changed salinity. We compared the differences in DNA methylation between marine and freshwater populations with genetic adaptations of a freshwater population studied in (Terekhanova et al. 2014). To confirm that changes in DNA methylation have a functional effect, we also investigated the expression of the genes associated with differentially methylated regions in the same sample groups.

Results

Changes in Water Salinity Have an Impact on DNA Methylation in Sticklebacks

To explore the role of DNA methylation in adaptation to freshwater conditions, we studied the following sample groups: i) marine sticklebacks kept in marine water, their natural habitat (M@M); ii) freshwater sticklebacks kept in freshwater (F@F); iii) marine sticklebacks incubated in fresh water for 4 days (M@F); iv) freshwater sticklebacks incubated in marine water for 4 days (F@M), (fig. 1).

DNA methylation was profiled in gills because this organ comes into direct contact with the surrounding water and is likely to be highly affected by osmotic stress. Three comparisons were performed between the studied groups of individuals to discover differential methylation between marine and freshwater populations of sticklebacks in water with natural salinity (M@M vs. F@F), methylation changes induced by placing a marine stickleback into a freshwater environment (M@M vs. M@F) and methylation changes induced by placing a freshwater stickleback into a marine environment (F@F vs. F@M). We found 61, 245 and 26 DMRs for the three listed comparisons (M@M vs. F@F, M@M vs. M@F, F@F vs. F@M, respectively), (Supplementary fig. S1A, Supplementary Material online). Two DMRs were validated with bisulfite treatment followed by Sanger sequencing (Supplementary fig. S2, Supplementary Material online). The number of DMRs discovered in each comparison did not reflect the difference between groups: MM and MF groups were the most similar but had the highest amount of DMRs if compared against each other (Supplementary fig. S3, Supplementary Material online). Instead, more DMRs were discovered between groups having low within-group variance.

Most of the DMRs we discovered were located within annotated genes or within 1 kb from gene boundaries (Supplementary fig. S4, Supplementary Material online), even though genes and gene flanks occupy no more than a quarter of the genome. Nevertheless, the observed significant

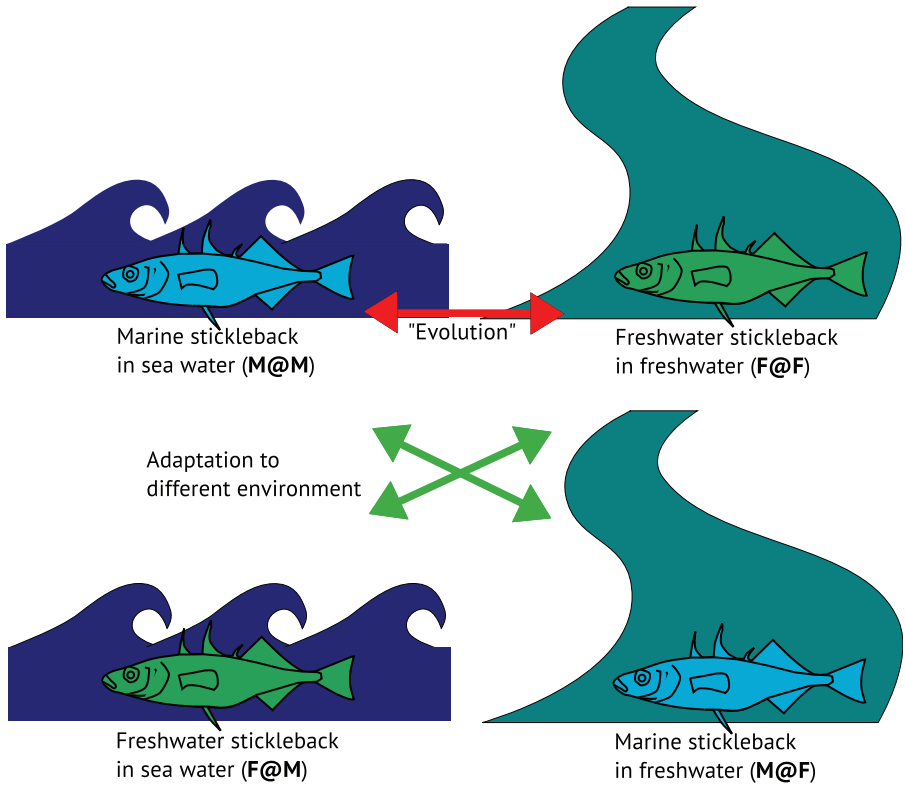

FIG. 1. Experimental design. DNA methylation was profiled in four conditions: marine stickleback in marine and freshwater environment, freshwater stickleback in freshwater and marine environment. Differentially methylated regions were found between marine and freshwater sticklebacks in their respective natural habitats. These methylation changes were compared with the immediate changes caused by placing marine sticklebacks into a freshwater environment. Additionally, a similar comparison was made to find DNA methylation changes after placing freshwater sticklebacks into a sea environment.

enrichment of DMRs within genes (p(Binomial) between $2*10^{-30}$ and 10^{-5}) can be explained by increased CpG density in the genic regions and in the regions profiled by RRBS rather than by the functional role of DMRs. We associated the DMRs with genes (see Materials and Methods). We called a gene differentially methylated (DMG) if it was linked to at least one DMR. We detected 40, 151 and 16 DMGs for M@M vs. F@F, M@M vs. M@F, and F@F vs. F@M comparisons, respectively (Supplementary fig. S1*B*, Supplementary Material online).

We also validated functional changes in expression of the genes associated with DMRs by RNA-seq. Out of 11 DMGs that were differentially expressed between marine and freshwater populations (M@M vs. F@F), eight genes had expression changes in line with the direction of change in DNA methylation: seven genes were hypomethylated and upregulated, one gene was hypermethylated and downregulated. This result indicated that, in general, DNA methylation was associated with repressed transcription.

Some of the observed DMRs were located in gene bodies which can be explained by increased GC content in exons and the bias of RRBS towards GC- and CpG-rich sequences. DMRs within gene bodies can occur at intergenic enhancers or at alternative promoters—in this case we would expect an increase of gene expression following hypomethylation of a DMR. Alternatively, DNA methylation can be changed because of altered transcription elongation in gene body—in this case increased expression is expected to cause increase in DNA methylation. Therefore, we observe negative correlation between DNA methylation and expression only in some of the detected DMGs.

Ion Channels, Membrane Proteins and Regulatory Genes Are Differentially Methylated between Marine and Freshwater Sticklebacks

To investigate long term evolutionary adaptation, we compared M@M to F@F (MMFF). A list of DMGs for MMFF consisted of 40 genes (Supplementary fig. S1*B*, Supplementary Material online). Gene category enrichment analysis revealed that the genes differentially methylated in the freshwater stickleback (F@F) population, as compared with the marine (M@M) population, were significantly enriched for ion channels and transmembrane proteins (fig. 2*A*). In fact, all the categories discovered were associated with membrane-bound proteins. The discovered ion channels included CACNA1FB (ENSGACG00000000800)—calcium channel, voltage-dependent, L type, alpha 1F subunit, H+/K+ exchanging ATPase ATP4A (ENSGACG00000008911) and potassium voltage-gated channel KCND3 (ENSGACG00000000195), as well as gap junction protein GJA3 (ENSGACG00000001367). In figure 3*A*, we summarized the methylation changes in individual CpGs in the DMR associated with one of these ion channels, CACNA1FB. A coordinated decrease in methylation was observed in each CpG within the DMR in the freshwater population (FF) compared

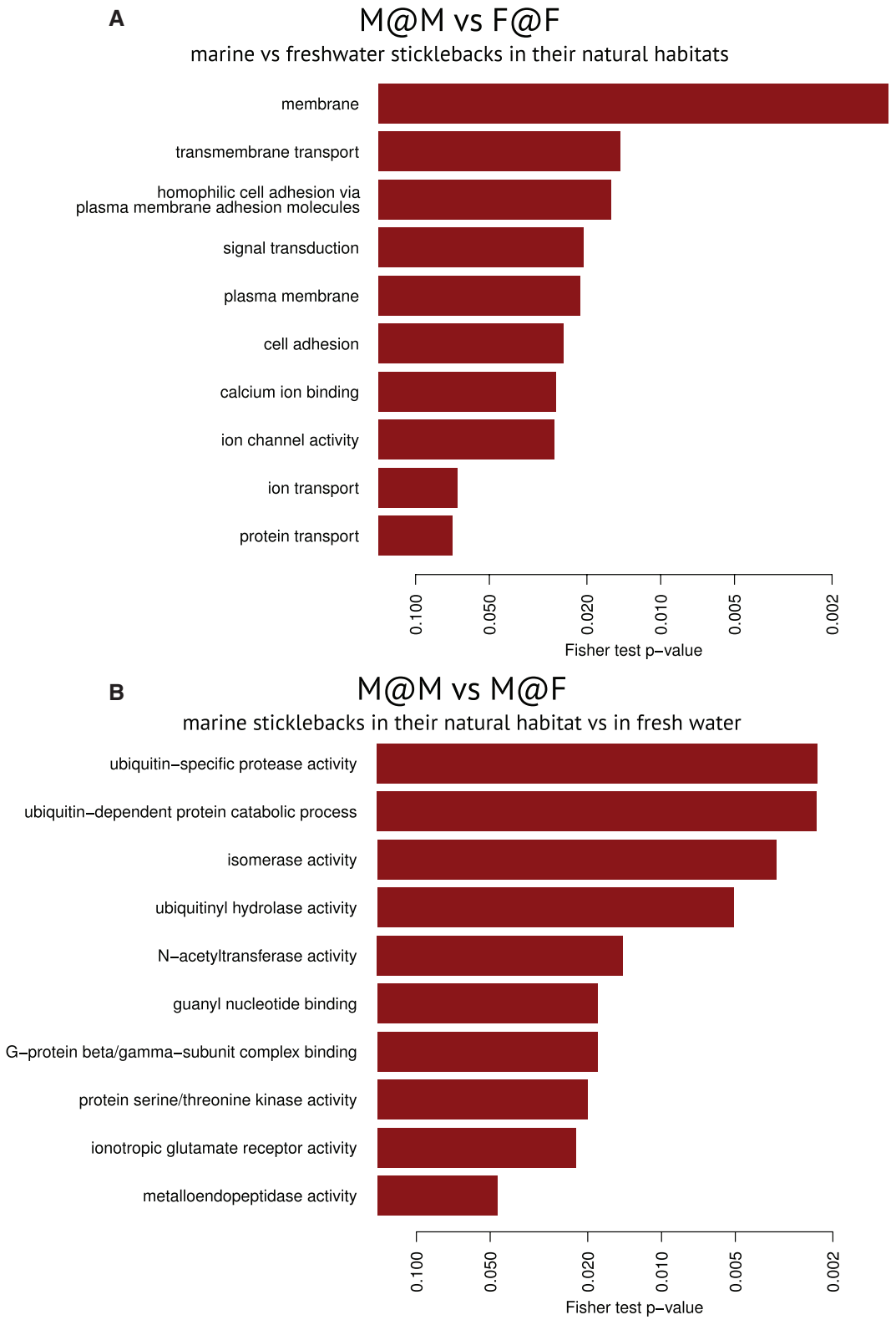

Fig. 2. GO categories enrichment plot for genes associated with DMRs between (*A*) marine and freshwater populations; (*B*) marine population in its native environment and marine sticklebacks placed into fresh water. *X*-axis shows enrichment *P* value according to Fisher exact test (logarithmic scale). Only categories with more than one differentially methylated gene were considered.

with the marine population (MM). Moreover, we showed that expression of CACNA1FB was significantly higher in the freshwater population than the marine population (fig. 3*B*, *P* value of differential expression between MM and PP groups $3.9*10^{-10}$, FDR $7.9*10^{-8}$). The direction of changes in DNA methylation and gene expression were in line with a common concept that DNA methylation in gene promoters was negatively correlated with gene expression. Similar figures

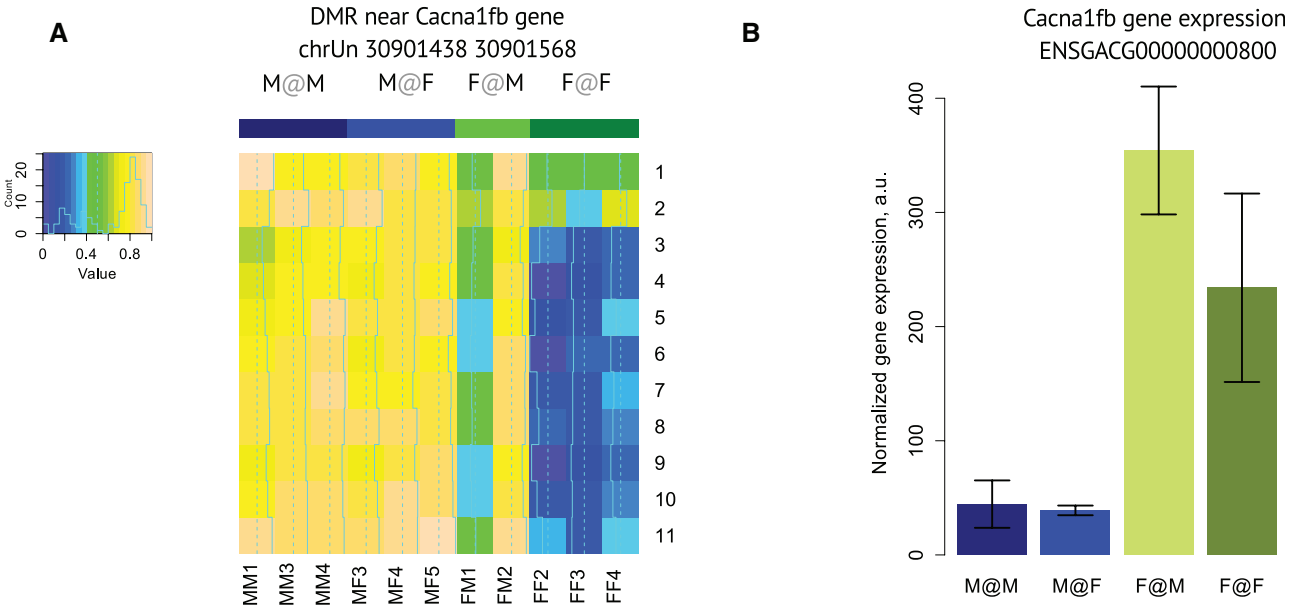

Fig. 3. DNA methylation status and gene expression of CACNA1FB (ENSGACG00000000800) gene in the profiled samples. Figure (*A*) shows methylation values of individual CpGs within a DMR found near the gene. Only positions covered at least ten times in each sample were considered. Figure (*B*) shows normalized expression counts for the gene derived from RNA-seq data. DESeq test of differential expression between MM and PP groups yielded a significant *P* value of $3.9*10^{-10}$, even after multiple testing correction (FDR = $7.9*10^{-8}$).

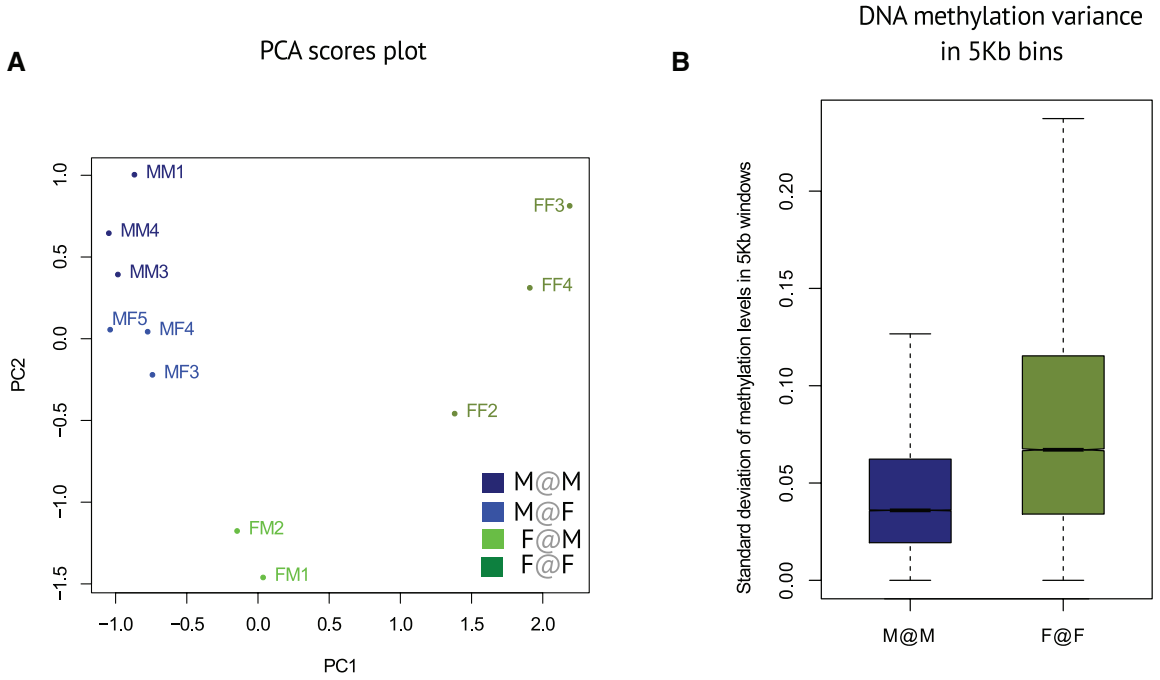

Fig. 4. (*A*) PCA plot showing all studied samples in four experimental groups. First principal component explained 42% of total variance, first two principal components explained together 59% of total variance. (*B*) Variance in DNA methylation levels in 5 kb genomic bins between individuals from stickleback populations. Significantly higher variance was observed in the freshwater population compared with the marine population, which could compensate for the lack of genetic variance in the freshwater population.

for other DMGs are provided in Supplementary Text S2, Supplementary Material online.

It is widely accepted that in evolution, marine sticklebacks colonized freshwater environments (and not the other way around), (Jones et al. 2012; Terekhanova et al. 2014). Therefore, we wanted to determine if long-term evolutionary adaptation to freshwater involved the same mechanisms as the short-term adaptation that could be observed in marine stickleback put into fresh water (M@F). As expected, correlation of methylation profiles of the studied MM, FF, and MF samples suggested that MF samples were more similar to their original marine population (MM) than to the freshwater population (Supplementary fig. S3, Supplementary Material online). Similar clustering was observed for RNA-seq data

(Supplementary fig. S5, Supplementary Material online). However, six genes were associated with significant methylation changes both between freshwater and marine populations (F@F|M@M) and in marine sticklebacks placed into freshwater compared with marine populations (M@F|M@M) (Supplementary fig. S1*B*, Supplementary Material online). Among them, we discovered an ion pump, alpha polypeptide of H+/K+ exchanging ATPase ATP4A (ENSGACG00000008911). Other genes were likely to have regulatory function: sirtuin SIRT2, a deacetylase involved in many cellular processes including histone deacetylation (ENSGACG00000005747), protein kinase C-binding protein NELL1 (ENSGACG00000017098), ubiquitin specific peptidase USP20 (ENSGACG00000015949), methylthioadenosine phosphorylase MTAP (ENSGACG00000017558), and M-phase phosphoprotein MPHOSPH9 (ENSGACG00000015639). Interestingly, NELL1 is believed to control skeletal ossification (James et al. 2015) and therefore may partially contribute to the well-known difference in the phenotypes of armor plates between marine and freshwater sticklebacks (Bell 2001).

DNA Methylation After a Short-Term Adaptation to Freshwater Partially Mimics Evolutionary Adaptation

As we have shown, for some genes, including the ion pump ATP4A, DNA methylation change induced by environmental stress mimicked that between marine and freshwater populations, but it was still unclear how universal this effect could be. First, we explored the overlap between the sets of DMGs from each of the three comparisons: M@M vs. F@F, M@M vs. M@F, and F@F vs. F@M (Supplementary fig. S1*B*, Supplementary Material online). The overlaps of 6 DMGs shared between the M@M vs. F@F and M@M vs. M@F comparisons and 3 DMGs shared between the M@M vs. F@F and F@F vs. F@M comparisons seemed relatively small, yet significant considering total number of genes not associated with DMRs (P values of $4*10^{-19}$ and 0.018, respectively; see Materials and Methods for the statistical procedure).

To figure out how similar immediate adaptations were to interpopulation differences on the genome scale, we performed a PCA analysis of DNA methylation levels in 5 kb genomic windows in all studied samples (fig. 4*A*). The samples fall into four groups according to their origin. A possible interpretation of the PCA plot could associate PC1 with adaptation to freshwater conditions (PC1 discriminates best between marine and freshwater populations in their natural habitats and F@M samples fall in between), whereas PC2 might reflect a general stress response after placing a fish into a different environment. Enrichment of the genes differentially methylated after placing fish into a different environment, for example, M@M vs. M@F, within gene categories related to signal transduction, particularly, ubiquitination, G-protein signaling, and kinases (fig. 2*B*), supported the idea that PC2, which separated fish in their native and foreign environments, was related to stress response. Taken together, our results suggested that some of the prospective epigenetic adaptations to decreased water salinity emerged both in the freshwater population and in marine sticklebacks "*de novo*" placed into freshwater. In other words, some of the immediate epigenetic adaptations to freshwater could be maintained in populations over time.

The Interplay between Genetic and Epigenetic Adaptation

To find out if DNA methylation can play a role in epigenetic activation or suppression of genes or regulatory regions that were under selection in the freshwater population (divergence islands, DI, [Terekhanova et al. 2014]), we searched for DMRs within DIs. We discovered that DI XXI-1 (chrXXI 5757849 7491073) overlapped with 3 DMRs from the M@F-M@M comparison, and no other overlaps were found (Supplementary fig. S6, Supplementary Material online).

As DNA methylation was known to be a major factor influencing mutagenesis of cytosines (C to T transitions), we next studied if DMRs were enriched by rare polymorphisms with alternative alleles present only in freshwater populations and absent in the marine population. The DMRs overlapped with neither rare freshwater-specific polymorphisms, nor any other SNPs having significantly different allele frequencies between marine and freshwater populations. The obtained results are not surprising, as the studied set of SNPs was believed to contain mostly SNPs that were preexisting in the marine population and selected in freshwater populations, rather than *de novo* mutations happened in freshwater populations.

Variation in DNA Methylation Is Increased in the Freshwater Population

We hypothesized that for the freshwater stickleback population, which had passed a population bottleneck and thus had limited genetic variability, it would be beneficial to compensate for the lack of heterozygosity by increased epigenetic variance. We calculated the variance of DNA methylation in 5 kb consecutive bins throughout the whole genome in the original marine (M@M) and freshwater (F@F) populations. Indeed, the freshwater population (F@F) demonstrated significantly higher variance in DNA methylation than the marine population (M@M, Wilcoxon test P value $< 2.2*10^{-16}$), suggesting that DNA methylation may contribute to freshwater adaptation by increasing the phenotypic variance (fig. 4*B*, Supplementary fig. S7, Supplementary Material online). This would yield a wider window of possibilities for a population to adapt to environmental challenges.

Discussion

Adaptation of marine sticklebacks to freshwater conditions was considered as a model system to study underlying genetic events. Freshwater fish were collected from Mashinnoe lake, which was physically separated from the sea by the motion of tectonic plates around 700 years ago (Kolka and Korsakova 2005). From this time on, the salinity of the lake dropped to a freshwater level. Current understanding of adaptation includes a model where preexisting alleles in the seawater population have selective advantages that allow certain fish to survive in freshwater environments. There was a number of studies that resolved genetic mechanisms and pointed to the

alleles that change their frequency upon this adaptation (Jones et al. 2012; Terekhanova et al. 2014). However, little is known on how the stickleback epigenome responds to altered salinity. We aimed to detect changes in DNA methylation in fish living at different salt concentrations.

Similarly to previously discovered selection-driven genetic alterations between marine and freshwater populations, the epigenetic changes which we found in this study were associated with transmembrane proteins and particularly ion pumps. Differentially methylated genes included a calcium channel CACNA1FB, a potassium channel KCND3, an H+/K+ exchanging ATPase ATP4A and a gap junction protein GJA3. Strikingly, genetic studies of marine and freshwater sticklebacks discovered functionally similar genes, such as potassium channel KCNH4 (Jones et al. 2012) and an ion-exchanging ATPase ATP1A1 (Terekhanova et al. 2014). Nevertheless, the sets of genes that were affected genetically and epigenetically did not overlap. This fact supported the idea that genetic and epigenetic adaptations represent independent mechanisms, which are both necessary for environmental adaptation. Moreover, it suggested that the observed epigenetic differences did not arise solely *in cis* due to different genetic backgrounds of fish populations. However, our model does not exclude a possibility that DNA methylation is affected by genetic background *in trans*.

Remarkably, the differences in DNA methylation between the marine and the freshwater populations had a significant similarity to the changes in DNA methylation that were induced by short-term exposure of marine fish to freshwater environment. A significant overlap of the DMGs found in both of those two comparisons included an ion pump (ATP4A) and several regulatory genes. Differential methylation in one of these genes, NELL1, could potentially affect skeletal ossification and explain the difference in the phenotypes of armor plates between marine and freshwater sticklebacks (Bell 2001; James et al. 2015). Alteration of DNA methylation in the genes that were differentially methylated in both comparisons might represent an initial response to freshwater conditions, which cannot be substituted by other evolutionary changes and thus the difference remains between the populations. However, the current data are not sufficient to distinguish if the methylation pattern characteristic to marine or freshwater populations is inherited through meiosis or results from permanent exposure to freshwater or marine environment.

The genes that were differentially methylated after placing a marine stickleback into freshwater were enriched with stress response and signal transduction genes, which resembled the response of other organisms to osmotic stress. In particular, we observed differential methylation of the genes involved in ubiquitination and autophagy, such as ubiquitin-specific peptidase USP16, ubiquitin carboxyl-terminal esterase L3 UCHL3, ubiquitin conjugating enzyme E2 E1 UBE2E1, as well as mitogen-activated protein kinase kinase kinase 4 (MAP3K4) which is believed to be activated in response to stress, including osmotic stress. Previously, osmotic stress was shown to cause accumulation of polyubiquitinated proteins through activation of p38 MAPK kinases which in turn inhibited proteasome activity (Whitmarsh and Davis 2007; Lee et al. 2010). Hypermethylation of MAP3K4 kinase and hypomethylation of uchl3 observed in marine sticklebacks placed into freshwater could in principle reduce protein ubiquitination level and serve as an adaptation to osmotic stress.

DNA methylation can potentially contribute to gene inactivation in two ways: on one hand, *de novo* methylated regions mark genes with altered expression, on the other hand, methylated cytosines are targets for increased (up to 5-fold) cytosine-uracil-thymine mutability through deamination (Cooper and Krawczak 1989). On a population level, DNA methylation was also shown to greatly influence cytosine mutation rate in humans (Xia et al. 2012). Thus, changes in DNA methylation in germline cells presumably had a potential to become fixed in the genome upon sea to freshwater adaptation, which could theoretically lead to increased abundance of C > T transitions between sea (C) and freshwater (T) sticklebacks among C's that are *de novo* methylated in freshwater fish. However, we found no overlap between polymorphisms that have differential allele frequencies in marine and freshwater sticklebacks and DMRs in any of the three comparisons. This observation does not disprove the hypothesis that DNA methylation could facilitate genetic silencing of genes by increasing the mutation rate of methylated CpGs in their regulatory regions, as the studied polymorphisms reflected selection rather than mutagenesis. Due to low sample sizes, current studies are unable to fetch polymorphisms with low minor allele frequencies. Therefore, most mutational effects are likely to be lost.

In Feinberg and Irizarry (2010), a new concept of variably methylated regions (VMRs) was defined as regions in which DNA methylation varies stochastically across the population, even within the same tissue and even in isogenic mice. In a conventional model, marine sticklebacks represent a genetic reservoir that gives rise to a freshwater population with minimal genetic diversity, but with a capability to live both at high and low salt concentrations. We formulated a hypothesis that a population challenged by minimal genetic diversity might adapt to a new environment and maintain a certain level of phenotypic variation by having a high level of epigenetic variability. Hence, we expected to see higher variance of DNA methylation in the freshwater population compared with the marine population. Indeed, figure 4*B* confirms this statement. In an attempt to find a mechanism driving higher epigenetic variance in the freshwater population, we checked whether genomic loci that were reported by Terekhanova et al. (2014) and served as landmarks of the genetic component of adaptation included genes that were responsible for chromatin architecture or DNA methylation. However, we found no chromatin regulators among the genes within the regions under selection. Genetics that could underlie higher epigenetic plasticity in freshwater sticklebacks still remains unclear.

Materials and Methods

Sample Collection

Ten freshwater three-spined sticklebacks were collected in the Mashinnoye Lake, Republic of Karelia, Russia, in

June–July 2014. The same number of marine three-spined sticklebacks were collected in the Kandalaksha Gulf Coast (15 km from the railway station Poyakonda, Republic of Karelia, Russia; water salinity of 26 promille) during the same time frame. To ensure equal physiological states of the fish, only males in breeding dress were collected.

To explore the immediate epigenetic response of a stickleback to changes in osmotic conditions of the environment, five marine fish were kept for 4 days in fresh water (M@F) and, vice versa, five freshwater fish were kept for 4 days in marine water (F@M, water salinity of 26 promille). We chose the exposure time to let the epigenome adapt to marine or freshwater conditions. It has been shown that for a cell line it takes up to 24 h to establish epigenome changes after exposure to high salt conditions (Dmitrieva et al. 2011). We hypothesized that it could take longer for the organism tissue to adapt, therefore we chose 4 days as a period of exposure to a different salinity conditions. It should be noted that control samples (M@M, F@F) were kept in the water with native salinity for the same time period to equal captivity stress influence.

Library Preparation, RRBS Sequencing and Data Processing

Control fish were kept in their native habitats (marine samples in marine water (M@M), freshwater samples in fresh water (F@F)) to equal captivity stress influence. After 4 days of exposure, DNA was isolated from gills (an organ that is directly exposed to water is likely to be highly affected by osmotic stress) using a standard method of DNA extraction from animal tissue (phenol-chloroform). The concentration of DNA in the samples was measured using a Nanodrop spectrophotometer (Supplementary table S2, Supplementary Material online).

Three individual fish from each of the four experimental conditions were taken for bisulfite sequencing (Supplementary table S2, Supplementary Material online). DNA methylation was profiled by reduced representation bisulfite sequencing (RRBS). Two micrograms of genomic DNA from three-spined stickleback samples of *G. aculeatus* of (Supplementary table S2, Supplementary Material online) and 4 ng of lambda phage DNA were digested using 60 U MspI (Fermentas, USA) in 50 µl at 37°C for 18–24 h, followed by QIAquick purification (Qiagen, Germany). The end of the digested DNA was repaired, and an adenine was added to the 3′ end of the DNA fragments according to the Illumina standard end repair and add_A protocol (Illumina, USA). Preannealed forked Illumina adaptors containing 5′-methylcytosine instead of cytosine were ligated to both ends of DNA fragments using standard Illumina adaptor ligation protocol (Illumina, USA). Ligated fragments were then separated by 2% agarose gel (Sigma-Aldrich, USA). Fragments between 170 and 350 bp, (includes adaptor length), were selected and cut from the gel. DNA from gel slices were purified using the Qiagen Gel Extraction Kit (Qiagen, Germany). The sodium bisulfite treatment and subsequent clean-up of size selected DNA was performed with the EZ DNA Methylation™ Kit (ZymoResearch, USA) according to the manufacturer's instructions. The bisulfite-treated DNA fragments were amplified using PCR and the following reaction: 5 µl of eluted DNA, 1 µl of NEB PE PCR two primers (1.0 and 2.0), and 45 µl Platinum PCR Supermix (Invitrogen, USA). The amplification conditions were as follows: 5 min at 95 °C, 30 s at 98 °C then 15 (10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C), followed by 5 min at 72 °C. The PCR reaction was purified by MinElute PCR Purification Kit (Qiagen), and final reduced representation bisulfite library was eluted in 15 µl EB buffer. The concentration of the final library was measured using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The library was sequenced on Illumina 2500 platform according to standard Illumina cluster generation and sequencing protocols. 100-bp single-end reads were generated.

Reads were mapped to gasAcu1 reference genome (augmented with the genomic sequence of lambda phage) with Bismark software (Krueger and Andrews 2011). As the spiked-in phage DNA was unmethylated, the methylation level of cytosines in lambda phage DNA profiled in the experiment served as a measure of how incomplete the bisulfite conversion was. The observed methylation levels of 0.30–0.65% for all sample except FM3 suggested that bisulfite conversion was 99.35–99.70% complete (Supplementary table S3, Supplementary Material online, "RRBS" sheet). However, for FM3 sample bisulfite conversion appeared to be incomplete (41%). This sample from the FM experimental group was excluded from the analysis after quality control. Individual samples from each experimental groups were treated as biological replicates. With bsseq R package (Hansen et al. 2012), we found individual differentially methylated CpG's and differentially methylated regions (DMRs). Default parameters for bsseq were used: minimal coverage per CpG 10, *t*-statistics cutoff 4.6, minimal mean methylation difference between groups 0.1, minimal number of CpGs per DMR 3. Individual DMRs were visualized to inspect methylation frequencies and read coverage per individual CpGs in individual samples. RRBS produced data for each CpG dinucleotide, thus giving information with single nucleotide resolution. To enhance the statistical power of the analysis, instead of looking for individual CpG dinucleotides which altered methylation level, we searched for differentially methylated regions (DMRs) that contain at least three CpG and demonstrate at least 10% difference in methylation. ENSEMBL gene annotation (version 81) was used to find genes localized no further than 1 kb from differentially methylated sites.

To estimate the false discovery rate (FDR) of the DMR discovery procedure, we applied a permutation approach. In every iteration, we randomly selected two groups of three samples each and ran the DMR discovery procedure, which would find differentially methylated regions between those two groups. We excluded the cases in which any of the two permuted groups was equal to one of the experimental groups (e.g., all samples in a permuted group came from the MM group). The procedure was repeated 100 times. This procedure yielded a median of 20 DMRs per iteration between permuted groups of samples, which can be used as an expected number of false discoveries. Considering that samples from all experimental groups were used for permutations, we compared the expected number of false

discoveries to the average number of real DMRs discovered in the MMFF and MMMF comparisons. We only considered the comparisons in which a group of three samples was compared with a group of three samples (i.e., MMFF and MMMF) as a lower number of samples in an experimental group (e.g., two samples in the FM group) can affect the number of the discovered DMRs. Based on these considerations, the FDR can be estimated on the level of 0.07 (20 expected false discoveries vs. 153 real hits).

For DMR validation, genomic DNA from gills was purified using a standard procedure of DNA extraction from animal tissue (phenol-chloroform). Bisulfite conversion was performed by EZ DNA methylation Kit (Zymo Research) according to the manufacturer's instructions. After bisulfite conversion, selected regions were amplified by the following pairs of primers:

TTGGATCCGTTTGGTATGTAATTATATTTGGT and

TTGAATTCACTACTACGAAATAAATAACACCC;

TTGGATCCTGATTGAATATGTAGTATGTAGTT and

TTGAATTCAACTCAACTACAACTCAATA.

PCR products were cloned to the T-vector. For each region ten clones were sequenced by Sanger and analyzed. The two profiled regions appeared to be differentially methylated in Sanger sequencing (Fisher's exact test P values were $<2.2*10^{-16}$ and $2.3*10^{-9}$; Supplementary fig. S2, Supplementary Material online). We also used DNA methylation values profiled by Sanger sequencing to validate precision of RRBS profiling for individual CpGs. Overall correlation of methylation rate profiled by RRBS and Sanger for individual CpGs in corresponding sample groups was 0.900 (P value $= 2*10^{-9}$).

To describe how the DMRs were localized with respect to genes and gene flanks, we calculated the number of DMRs within genes, the number of DMRs located no further than 1 kb from a gene and the number of DMRs in intergenic regions. We also estimated the total genomic length of genic regions, gene flanks, and intergenic region. A background probability of a DMR to appear within a gene or a gene flank under a null hypothesis of no enrichment was estimated as a fraction of the total length of genic regions and gene flanks in the genome. We performed a binomial test to estimate if the DMRs were enriched within genes and gene flanks.

We annotated the DMRs with respect to known stickleback genes. In particular, we searched for all differentially methylated regions (DMR) within 1kb upstream and 1kb downstream of all annotated genes. We considered a gene to be differentially methylated (DMG) if it had a DMR no further than 1 kb from it (Supplementary table S1, Supplementary Material online). Additionally, we provide average DNA methylation within the 2 kb regions surrounding a TSS for each transcript (Supplementary table S4, Supplementary Material online).

DMGs were further used for gene category enrichment analysis. Annotation of genes related to GO categories was taken from ENSEMBL database. Gene set enrichment analysis was done as follows: for each category having at least one gene in a gene set, we performed Exact Fisher's test to check if the category is enriched in a given gene set.

To estimate the significance of overlap between two sets of DMRs (e.g., M@M vs. F@F and M@M vs. M@F) under a strict assumption that most of DMRs were associated with gene promoters, the following test was performed. For each gene, we asked if a DMR from each of the two sets was associated with the gene. Next, we performed Fisher exact test to check if presence of a DMR in one set was associated with presence of a DMR in the other set.

RNA Preparation and RNA Sequencing

For RNA-seq, we used the same sample collection and treatment procedure as described for bisulfite sequencing. Four fish from each of the four experimental groups were taken for transcriptome analysis. Gills were isolated and fixed with IntactRNA® reagent (Evrogen).

Total RNA was extracted from the samples with Trisol reagent according to the manufacturer's instructions (Invitrogen). Quality was checked with the BioAnalyser and RNA 6000 Nano Kit (Agilent). PolyA RNA was purified with Dynabeads® mRNA Purification Kit (Ambion). An Illumina library was made from polyA RNA with NEBNext® mRNA Library Prep Reagent Set (NEB) according to the manual. Paired-end sequencing was performed on HiSeq1500 with 2 $\times$ 75 bp read length. Approximately 25 million reads were generated for each sample.

Reads were mapped to gasAcu1 genome with tophat2 software (version 2.1.0) (Kim et al. 2013). Number of RNA-seq reads, number of unmapped reads and mapping efficiency is summarized in Supplementary table S3, Supplementary Material online ("RNA-seq" sheet). Gene models of nonoverlapping exonic fragments were taken from ENSEMBL 54 database. For each exonic fragment, total coverage by mapped reads in each sample was calculated with bedtools multicov tool (version 2.17.0). Total gene coverage was calculated as a sum of coverages of all nonoverlapping exonic fragments of a gene. Differential expression analysis was performed by applying default read count normalization and performing per-gene negative binomial tests, implemented in edgeR R package, with default parameters (Robinson et al. 2010). For each gene, the package provided both P values and FDRs (P values after Benjamini-Hochberg multiple testing corrections). Raw read counts and normalized read counts for each gene in each sample, as well as differential expression P values and fold-changes for MMFF, MMMF, and FFFM comparisons are provided in Supplementary table S5, Supplementary Material online.

Exploratory Statistical Analyses

Principal component analysis (PCA) was performed with a standard prcomp function in R. The genome was split into 5 kb windows, a total amount of methylation-supporting reads and total coverage was calculated for each window, thus yielding a robust estimate of DNA methylation level. Methylation values for each bins were taken as input variables, whereas each point in a multidimensional space represented a stickleback individual.

To estimate variation of DNA methylation, we calculated DNA methylation levels in 5 kb windows the same way as for PCA analysis. Standard deviation of DNA methylation levels was calculated for each genomic window within each stickleback population. Distributions of obtained standard deviations were compared between populations, and each population contained the same amount of individuals profiled.

To confirm that the observed difference in variance was not caused by different number of reads in samples, we performed read sampling. From each sample, we randomly selected the number of reads equal to the minimal number of reads among all samples and repeated read mapping, methylation calling and variance analysis as described above.

Data Availability

RRBS data was deposited to NCBI GEO under GSE82310 study accession code.

Ethics Statement

This work was approved by ethical committee of Research Center of Biotechnology RAS, Moscow, Russia

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to acknowledge Nadezhda Terekhanova for providing preprocessed allele frequencies data in sea and freshwater populations from (Terekhanova et al. 2014) and for meaningful discussion. We would like to thank Cailin Kelly for editing the manuscript. We thank Prof Konstantin Skryabin for comprehensive assistance with this project. This work was supported by Russian Science Foundation (RSF) grant #14-24-00175.

References

Barrett RDH, Paccard A, Healy TM, Bergek S, Schulte PM, Schluter D, Rogers SM. 2011. Rapid evolution of cold tolerance in stickleback. *Proc Biol Sci*. 278:233–238.

Bell MA. 2001. Lateral plate evolution in the threespine stickleback: getting nowhere fast. *Genetica* 112–113:445–461.

Chan YF, Marks ME, Jones FC, Villarreal G Jr, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327:302–305.

Cooper DN, Krawczak M. 1989. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum. Genet* 83:181–188.

Dmitrieva NI, Burg MB. 2007. High NaCl promotes cellular senescence. *Cell Cycle* 6:3108–3113.

Dmitrieva NI, Cui K, Kitchaev DA, Zhao K, Burg MB. 2011. DNA double-strand breaks induced by high NaCl occur predominantly in gene deserts. *Proc Natl Acad Sci U S A*. 108:20796–20801.

Dmitrieva NI, Ferraris JD, Norenburg JL, Burg MB. 2006. The saltiness of the sea breaks DNA in marine invertebrates: possible implications for animal evolution. *Cell Cycle* 5:1320–1323.

Feinberg AP, Irizarry RA. 2010. Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A*. 107(Suppl 1):1757–1764.

Gao L, Geng Y, Li B, Chen J, Yang J. 2010. Genome-wide DNA methylation alterations of Alternanthera philoxeroides in natural and manipulated habitats: implications for epigenetic regulation of rapid responses to environmental fluctuation and phenotypic variation. *Plant Cell Environ* 33:1820–1827.

Guo B, DeFaveri J, Sotelo G, Nair A, Merilä J. 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol* 13:19.

Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* 13:R83.

Hohenlohe PA, Bassham S, Currey M, Cresko WA. 2012. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos Trans R Soc Lond B Biol Sci*. 367:395–408.

James AW, Shen J, Zhang X, Asatrian G, Goyal R, Kwak JH, Jiang L, Bengs B, Culiat CT, Turner AS, et al. 2015. NELL-1 in the treatment of osteoporotic bone loss. *Nat Commun*. 6:7362.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14:R36.

Kolka VV, Korsakova OP. 2005. Application of geological methods for dating of stone labyrinths on the White Sea coast. *Proc MSTU* 15:349–356.

Konijnendijk N, Shikano T, Daneels D, Volckaert FAM, Raeymaekers JAM. 2015. Signatures of selection in the three-spined stickleback along a small-scale brackish water - freshwater transition zone. *Ecol Evol*. 5:4174–4186.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571–1572.

Lee S-H, Park Y, Yoon SK, Yoon J-B. 2010. Osmotic stress inhibits proteasome by p38 MAPK-dependent phosphorylation. *J Biol Chem*. 285:41280–41289.

Lescak EA, Bassham SL, Catchen J, Gelmond O, Sherbick ML, von Hippel FA, Cresko WA. 2015. Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proc Natl Acad Sci U S A*. 112:E7204–E7212.

Liu S, Sun K, Jiang T, Feng J. 2015. Natural epigenetic variation in bats and its role in evolution. *J Exp Biol*. 218:100–106.

Liu S, Sun K, Jiang T, Ho JP, Liu B, Feng J. 2012. Natural epigenetic variation in the female great roundleaf bat (*Hipposideros armiger*) populations. *Mol Genet Genomics* 287:643–650.

McCairns RJS, Bernatchez L. 2010. Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution* 64:1029–1047.

Morris MRJ, Richard R, Leder EH, Barrett RDH, Aubin-Horth N, Rogers SM. 2014. Gene expression plasticity evolves in response to colonization of freshwater lakes in threespine stickleback. *Mol Ecol*. 23:3226–3240.

Norman JD, Ferguson MM, Danzmann RG. 2014. An integrated transcriptomic and comparative genomic analysis of differential gene expression in Arctic charr (*Salvelinus alpinus*) following seawater exposure. *J Exp Biol*. 217:4029–4042.

Rastorguev SM, Nedoluzhko AV, Sharko FS, Boulygina ES, Sokolov AS, Gruzdeva NM, Skryabin KG, Prokhortchouk EB. 2016. Identification of novel microRNA genes in freshwater and marine ecotypes of the three-spined stickleback (Gasterosteus aculeatus). *Mol Ecol Resour*. *Available from* http://dx.doi.org/10.1111/1755-0998.12545.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Roesti M, Marius R, Walter S. 2014. Natural selection: it's a many-small world after all. *Curr Biol*. 24:R959–R962.

Smith G, Smith C, Kenny JG, Chaudhuri RR, Ritchie MG. 2015. Genome-wide DNA methylation patterns in wild samples of two morphotypes of threespine stickleback (Gasterosteus aculeatus). *Mol Biol Evol*. 32:888–895.

Terekhanova NV, Logacheva MD, Penin AA, Neretina TV, Barmintseva AE, Bazykin GA, Kondrashov AS, Mugue NS. 2014. Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genet*. 10:e1004696.

Varriale A. 2014. DNA methylation, epigenetics, and evolution in vertebrates: facts and challenges. *Int J Evol Biol*. 2014:475981.

Whitmarsh AJ, Davis RJ. 2007. Role of mitogen-activated protein kinase kinase 4 in cancer. *Oncogene* 26:3172–3184.

Xia J, Han L, Zhao Z. 2012. Investigating the relationship of DNA methylation with mutation rate and allele frequency in the human genome. *BMC Genomics* 13(Suppl 8):S7.

Yi C, Zhang S, Liu X, Bui HTN, Hong Y. 2010. Does epigenetic polymorphism contribute to phenotypic variances in Jatropha curcas L.? *BMC Plant Biol*. 10:259.