Identification of Local- and Habitat-Dependent Selection: Scanning Functionally Important Genes in Nine-Spined Sticklebacks (*Pungitius pungitius*)

Takahito Shikano,* Jetty Ramadevi,† and Juha Merilä

Ecological Genetics Research Unit, Department of Biosciences, University of Helsinki, Helsinki, Finland

*Present address: School of Life Sciences, University of Hyderabad, Hyderabad 500 046, India.

**Corresponding author:** E-mail: takahito.shikano@helsinki.fi.

Associate editor: Koichiro Tamura

Abstract

Understanding the selective forces promoting adaptive population divergence is a central issue in evolutionary biology. The role of environmental salinity in driving adaptation and evolution in aquatic organisms is still poorly understood. We investigated the relative impacts of habitat type (cf. saltwater vs. freshwater) and geographic area in shaping adaptive population divergence, as well as genes responsible for adaptation to different salinities in nine-spined sticklebacks (*Pungitius pungitius*). To this end, we employed a hitchhiking mapping approach with 111 microsatellite loci and one insertion/deletion locus including 63 loci situated within or close to genes with important physiological functions such as osmoregulation, growth, and thermal response. Using three pairs of marine and freshwater populations from different geographic areas, we identified several loci showing consistent evidence of being under directional selection in different outlier tests. Analyses of molecular variance at the loci under selection indicated that geographic area rather than habitat type has been acting as a central force in shaping adaptive population divergence. Nevertheless, both outlier tests and a spatial analysis method indicated that two loci (growth hormone receptor 2 and DEAD box polypeptide 56) are involved in adaptation to different habitats, implying that environmental salinity has been affecting them as a selective force. These loci are promising candidates for further investigations focusing on the molecular mechanisms of adaptation to marine and freshwater environments.

Key words: adaptation, genome scan, hitchhiking mapping, natural selection, osmoregulation, *Pungitius*.

Introduction

Understanding the selective forces promoting adaptive population divergence is a central issue in evolutionary biology. Divergent selection stemming from abiotic and ecological factors is a key process in adaptive differentiation and speciation (Schluter 2001; Rundle and Nosil 2005; Funk et al. 2006; Schluter 2009). Accordingly, specific and predictable environmental conditions are expected to create similar patterns of natural selection. In aquatic organisms, the ability to tolerate hyperosmotic and hypoosmotic environments is a critical factor determining their habitat distributions (Bayly 1972; Evans 1984). As such, in most teleosts, dispersal between marine and freshwater habitats is prevented by salinity gradients even in the absence of physical barriers. However, some euryhaline species, such as Gasterosteiformes, Anguilliformes, and Cyprinodontiformes, have penetrated into both habitat types (Bayly 1972; Evans 1984). In the past decades, significant advances have been made in understanding the functional and molecular systems underlying their physiological acclimation to hyperosmotic and hypoosmotic environments (Marshall 2002; Hirose et al. 2003; Evans et al. 2005; Grosell 2006). Yet, very little is known about the genetic mechanisms involved in adaptation to marine and freshwater environments and the impact of habitat type on adaptive population differentiation (but see, e.g., Lee 2002; Colosimo et al. 2005; Fuller 2008; Purcell et al. 2008). Hence, the role of environmental salinity in adaptation and evolution of a species remains largely an unresolved issue in fish.

Current genomic approaches offer a powerful means of identifying the genomic targets of natural selection (Nielsen 2005; Storz 2005; Stinchcombe and Hoekstra 2008; Nosil et al. 2009). Genome scan approaches, such as quantitative trait locus and hitchhiking mapping, provide the opportunity to investigate the genetic basis of adaptation at the genome-wide scale. Hitchhiking mapping is a population genomic approach to identify genomic regions influenced by natural selection (Beaumont and Nichols 1996; Black et al. 2001; Schlötterer 2002; Beaumont and Balding 2004; Beaumont 2005). The great advantage of this approach is that it can be performed using molecular markers alone in the absence of prior knowledge about phenotypic traits under selection. Therefore, this approach is especially suitable for organisms in which breeding experiments are difficult to conduct due to long generation times or other logistic constraints. In addition, it allows for identifying signatures of selection on any traits including the physiological characteristics, such as sensitivity to environmental factors that are difficult to quantify in the wild or even under controlled conditions. In the recent years, an increasing number of studies have conducted hitchhiking mapping in a variety
of organisms (reviewed in Holderegger et al. 2008; Stinchcombe and Hoekstra 2008; Nosil et al. 2009).

The power of hitchhiking mapping relies largely on the density of molecular markers in the genome of a species because signatures of selection decay with increasing chromosomal distances from the selected sites (Wiehe 1998; Storz 2005; Stinchcombe and Hoekstra 2008; Nosil et al. 2009). Empirical studies have shown that a hitchhiking effect is generally observed in a very small genomic region around the selected gene (e.g., 20–300 kb; Wootton et al. 2002; Nash et al. 2005; Olsen et al. 2006; Mäkinen, Shikano, et al. 2008; Teschke et al. 2008). Accordingly, thousands of genetic markers would be required to scan a whole genome, although the utility of polymorphic markers within expressed sequence tags (ESTs) can increase the probability to detect signatures of selection (Vigouroux et al. 2002; Vasemägi et al. 2005; Yatabe et al. 2007; Namroud et al. 2008). Due to practical reasons, hitchhiking mapping in nonmodel organisms has typically been conducted with relatively small numbers of markers randomly derived from the genome and ESTs (Holderegger et al. 2008; Nosil et al. 2009). Consequently, several regions of the genome involved in adaptation can be easily missed in such scans (Excoffier et al. 2009). Moreover, even if signatures of selection are detected, the undefined link between phenotype and genotype makes it difficult to understand what kinds of selection pressures have been acting on these genomic regions.

Despite the limitations of hitchhiking mapping, its effectiveness can be enhanced by incorporating other approaches. Recent advances in the physiological and molecular understanding of genes provide better predictions of genes potentially involved in adaptation and phenotypic variation (Aerts et al. 2006; Zhu and Zhao 2007; Piertney and Webster 2010). From the standpoint of a traditional candidate gene approach, targeting functionally important genes—rather than random genes and genomic regions—is a straightforward way to scan key genes in the genome. At the same time, a priori knowledge of the biological functions of genes provides obviously better insights into the nature of selection pressures underlying the identified selection rather than a posteriori analyses of gene homology and annotation, which are typically performed for random EST markers. Further inference of environmental pressures can be gained by investigating associations between allele frequencies and environmental variables (Joost et al. 2007). These complementary approaches provide a better understanding of the genomic targets of selection and the environmental factors defining their underlying selective pressures.

The nine-spined stickleback (Pungitius pungitius) is an euryhaline teleost with a broader salinity tolerance than primary freshwater teleosts (Heuts 1943). In the postglacial epoch, nine-spined sticklebacks are widespread in the northern hemisphere and inhabit a wide variety of environments (Wootton 1976; Paepke 2001). Although their occurrence in marine environments is limited (McPhail 1963; Takata et al. 1987; Paepke 2001), they occupy brackish water and saltwater environments along coastal areas of the Arctic Ocean, the White Sea, and the Baltic Sea (Paepke 2001). In Fennoscandia, they are common in both freshwater and coastal habitats. Such penetrations into divergent environments may lead to natural selection on genes responsible for osmoregulatory functions. In addition, heterogeneous local environmental conditions can facilitate natural selection on genes underlying life-history, physiological, morphological, and behavioral traits independently of the impact of environmental salinity. The distribution patterns of Fennoscandian nine-spined sticklebacks allow for multiple comparisons between coastal and freshwater populations in different geographic areas, thus offering an excellent system to assess genomic targets of local- and habitat-dependent selection.

In this study, we investigated the relative impacts of habitat type and geographic area in shaping adaptive population divergence, as well as genes responsible for adaptation to environments differing in salinity in nine-spined sticklebacks. To this end, we employed a hitchhiking mapping approach with 111 microsatellite loci and one insertion/deletion locus including 63 loci situated within or close to genes with known physiological functions in fish such as osmoregulation, growth, and thermal response. To identify local- and habitat-dependent selection, we used three pairs of marine and freshwater populations from different geographic areas of Fennoscandia. In addition to outlier analyses, a spatial analysis method (SAM; Joost et al. 2007) was conducted to determine whether loci subject to selection are associated with environmental salinity or not. We also compared the results with those from an earlier study (Mäkinen, Cano, and Merila 2008) of three-spined sticklebacks (Gasterosteus aculeatus) for 32 loci shared by both studies to see if directional selection has been acting on the same loci in these ecologically and morphologically similar species inhabiting the same area.

Materials and Methods

Study Populations
Six populations consisting of three pairs of marine and freshwater populations were used in this study (fig. 1). These population pairs were selected from different geographic areas: northern (RU: RU-LEV—66°18’N, 33°24’E and RU-BOL—66°18’N, 33°24’E), central (SE: SE-BOL—63°40’N, 20°13’E and SE-KRO—63°41’N, 20°24’E), and southern Fennoscandia (FI: FI-HIL—60°12’N, 25°11’E and FI-MAT—60°12’N, 24°51’E). Marine populations were sampled from the White Sea (RU-LEV) and the Baltic Sea (SE-BOL and FI-HIL), and freshwater populations were from a pond (RU-BOL), a lake (SE-KRO), and a river (FI-MAT). Different freshwater habitat types were used to identify selection acting between marine and freshwater habitats independently of freshwater habitat types. The pond and lake populations were physically isolated from the marine populations. Although the river is connected to the sea, a significant $F_{ST}$ value (0.05) between the river and closest marine populations—as well as lower genetic variation in the river population as compared with the marine
population—suggests that migration and gene flow between them are limited (see Results). Fish were collected with seine nets or minnow traps during 2003–2008. For each population, 24 individuals were used for the analyses, yielding a total sample size of 144 comprising 72 individuals from freshwater and 72 marine samples.

Marker Genotyping
Total DNA was extracted from fin clips stored in 70–95% ethanol using the phenol–chloroform method (Taggart et al. 1992) following proteinase K digestion. Allelic variation was assessed for 111 microsatellite loci and one insertion/deletion locus (supplementary tables S1 and S2, Supplementary Material online). Of the 112 markers, 63 were developed for particular genes responsible for significant functions in fish such as osmoregulation, growth, thermal response, and immune function (gene-based markers; Shikano, Ramadevi, et al. 2010; supplementary table S1, Supplementary Material online). Six microsatellite markers for functionally important genes were developed in this study based on the three-spined stickleback genome sequences (http://www.ensembl.org/Gasterosteus_aculeatus/Info/Index; supplementary table S1, Supplementary Material online). The remaining 21 and 28 markers were derived from three-spined stickleback ESTs (random EST markers) and genomic libraries (random genomic markers), respectively (Largiadèr et al. 1999; Peichel et al. 2001; Heckel et al. 2002; Colosimo et al. 2004). Each forward primer was labeled with a fluorescent dye (FAM, HEX, or TET), and the 5’-end of each reverse primer was modified with a GTTT-tail (Brownstein et al. 1996). These loci were arranged in multiplex polymerase chain reaction (PCR) panels with nonoverlapping size ranges in each dye. PCRs were carried out using the Qiagen Multiplex PCR Kit (Qiagen) in 10 µl reaction volumes containing 1× Qiagen Multiplex PCR Master Mix, 0.5× Q-Solution, 2 pmol of each primer, and 10–20 ng of template DNA. The reactions were performed by the following cycle: an initial activation step at 95 °C for 15 min, followed by 30 s at 94 °C, 90 s at 53 or 55 °C, and 60 s at 72 °C for 30 cycles with a final extension at 60 °C for 5 min (supplementary table S1, Supplementary Material online; Shikano, Ramadevi, et al. 2010). PCR products were resolved with a MegaBACE 1000 automated sequencer (Amersham Biosciences), and their sizes were determined with ET-ROX 550 size standard (Amersham Biosciences). Alleles were scored using FRAGMENT PROFILER 1.2 (Amersham Biosciences) with visual inspection and manual corrections of alleles.

Genetic Variation and Differentiation
Locus- and population-specific gene diversities (H_L; Nei 1987) were calculated using FSTAT 2.9.3 (Goudet 2001). Allelic richness was estimated using HP-RARE 1.0 (Kalinowski 2005) with a rarefaction sample size of 24 individuals. To detect possible deviations from Hardy–Weinberg equilibrium, within-population and locus-specific FIS were estimated with 10 000 permutations. Significant deviations were evaluated with and without sequential Bonferroni corrections (Rice 1989). The degree of population differentiation was quantified using FST (θ; Weir and Cockerham 1984) with FSTAT 2.9.3. Standard errors of FST were obtained by jackknifing over loci with 1000 permutations.

Because the inclusion of severely bottlenecked populations in outlier tests can increase the number of false positives (Teshima et al. 2006; Wiehe et al. 2007; Foll and Gaggiotti 2008), we evaluated the possibility of recent population bottlenecks using the approaches of Cornuet and Luikart (1996) and Luikart et al. (1998), as implemented in BOTTLENECK (Piry et al. 1999). We used the two-phase model (Di Rienzo et al. 1994) with 90% stepwise mutations. In the test of Cornuet and Luikart (1996), the significance of heterozygote excess was assessed by the Wilcoxon signed-rank test with 1000 iterations.

Detection of Selection Signatures
As a standard approach, we conducted global outlier tests using all populations to identify outlier loci representative of the whole data set. Because a more specific aim of our study was to assess the relative importance of geographic and habitat factors in adaptive divergence, we attempted to identify local- and habitat-dependent selection by investigating the patterns of selection among populations. Several studies have shown that outlier tests with different combinations of populations in a data set are effective to identify populations under selection and thereby potential environmental factors involved in selection (Campbell and Bernatchez 2004; Vasemägi et al. 2005; Bonin et al. 2006; Oetjen and Reusch 2007; Mäkinen, Cano, and Merila 2008; Paris et al. 2010). Accordingly, we performed outlier tests using sets
of populations with the same habitat type (i.e., three marine or three freshwater populations) from different areas to identify loci under selection independent of habitat type. In addition, pairwise comparisons of populations from different habitat types within respective areas were used to investigate habitat-dependent selection without geographic effects. In this analysis, we considered potential habitat-dependent selection if a particular locus was apparent as an outlier in multiple areas. Because outlier tests determine putatively neutral and selected loci based on genetic parameters in a set of loci and populations used for the analyses, some loci that are not identified as outliers under the global tests can appear as outliers in analyses with subsets of populations in the absence of populations exhibiting strong selection signatures for particular loci. We considered outliers identified only in the analyses with subsets of populations as weakly selected loci compared with those identified in the global tests.

Loci influenced by directional selection are expected to exhibit lower intrapopulation variability and larger interpopulation differentiation than neutral loci (Lewontin and Krakauer 1973; Beaumont and Nichols 1996; Schlötterer 2002; Beaumont and Balding 2004). Accordingly, signatures of directional selection were investigated based on the patterns of heterozygosity (Kauer et al. 2003), $F_{ST}$ (Vitalis et al. 2001; Foll and Gaggiotti 2008), and both heterozygosity and $F_{ST}$ (Beaumont and Nichols 1996). These conceptually different approaches were employed to reduce the number of false positives.

To screen reductions in heterozygosity, $\ln RH$ tests were performed in a pairwise fashion following Kauer et al. (2003). We estimated $\ln RH$ among populations rather than categorical (e.g., habitat type or geographic) groups consisting of individuals from different populations in order to avoid assumed increases in expected heterozygosity due to violation of Hardy–Weinberg equilibrium. For loci monomorphic in a data set, we assumed that one additional allele differed from the others following Kauer et al. (2003). After standardization of $\ln RH$ estimates with the mean of 0 and a standard deviation of 1, 95% and 99% of the loci are expected to have values of $-1.96$ to $1.96$ and $-2.58$ to $2.58$, respectively (Kauer et al. 2003). Loci with $\ln RH$ values outside these boundaries were considered as significant at the respective levels. In comparisons among three or more populations, $\ln RH$ estimates for a selected locus are expected to be significant in multiple pairwise comparisons that include a particular population under selection. Although it is difficult to define rigorous boundaries that eliminate false positives, because population pairs are not independent, we screened loci exhibiting significant $\ln RH$ values for at least three and two combinations in the comparisons among all six populations and three populations of the same habitat type, respectively. After this screening, the results were compared with those of other outlier tests so that false positives could be further minimized (see below).

To detect increased population differentiation, we employed the hierarchical Bayesian method of Foll and Gaggiotti (2008), which has been modified based on the approach proposed by Beaumont and Balding (2004). The method of Foll and Gaggiotti (2008) estimates population-specific $F_{ST}$ coefficients accounting for different intensities of genetic drift in the different populations. The analysis was conducted using BAYESCAN (http://www.leca.ujf-grenoble.fr/logiciels.htm) with 10 pilot runs of 5000 iterations, followed by additional 150 000 iterations with a burn-in of 50 000 iterations. Outliers were determined based on the 90%, 95%, and 99% posterior probabilities, which are the approximate indications of strong, very strong, and decisive selection, respectively, according to the instructions for BAYESCAN. Sample sizes for pairwise comparisons impeded convergence of Markov chain Monte Carlo simulations in the Bayesian method (Foll and Gaggiotti 2008; Paris et al. 2010); consequently, outlier tests for population pairs with different habitat types were conducted with the approach of Vitalis et al. (2001), which investigates outliers in a pairwise fashion based on population-specific $F$ statistics. The coalescent simulations were performed with DETSEL 1.0 (Vitalis et al. 2003). Null distributions were generated using the following parameters: population size before the split $N_0 = 500$; mutation rate $\mu = 0.0001$ and 0.00001; ancestral population size $N_s = 500$, 1000, and 10 000; time since bottleneck $T_0 = 50$, 100, and 1000; and time since population split $t = 100$. Outliers were determined based on an empirical $P$ value for each locus at the 95% and 99% levels using the 2D arrays of $50 \times 50$ square cells (Vitalis et al. 2001).

To evaluate $F_{ST}$ and heterozygosity simultaneously, we adopted a coalescent approach developed by Beaumont and Nichols (1996) as implemented in LOSITAN (Antao et al. 2008). This method identifies outliers based on the joint distributions of $F_{ST}$ and expected heterozygosity under an island model of migration (Beaumont and Nichols 1996). After removing a candidate subset of selected loci (outside the 95% confidence interval) determined by an initial run with 100 000 simulations, the distribution of neutral $F_{ST}$ was computed using putatively neutral loci with 100 000 simulations and a bisection approximation algorithm. Outliers were determined by comparing observed distributions with the neutral expectations at the 95% and 99% confidence levels.

There are both advantages and disadvantages for empirical and model-based tests (Storz 2005). Outliers detected by multiple methods are likely to be indicative of truly adaptive genomic regions because different approaches use different algorithms and assumptions (Vasemägi et al. 2005; Storz 2005; Kane and Rieseberg 2007; Paris et al. 2010; Hansen et al. 2010). Therefore, we minimized the number of potential false positives by comparing the results of these conceptually different tests rather than applying a Bonferroni correction with respective tests following previous studies (e.g., Vasemägi et al. 2005; Bonin et al. 2006; Kane and Rieseberg 2007; Paris et al. 2010; Willing et al. 2010). We considered the loci indicated to be under directional selection by two or three tests as directionally selected loci, and loci
indicated by a single test as probable false positives. As a complementary analysis for the global outlier tests with all populations, we used the hierarchical method of Excoffier et al. (2009), which has been modified based on the approach of Beaumont and Nichols (1996). We conducted the analysis under the hierarchical island model by pooling two populations from the White Sea area (RU-LEV and RU-BOL) into one group and four populations from the Baltic Sea area (SE-BOL, SE-KRO, FI-HEL, and FI-MAT) into another, as per the genetic population structure assessed by putatively neutral loci (see Results). The analysis was carried out under the infinite allele model assuming the presence of 10 groups of 100 demes with 50,000 simulations using ARLEQUIN 3.5 (Excoffier and Lischer 2010).

In addition to directional selection, balancing selection was investigated using the whole data set with the methods of Foll and Gaggiotti (2008), Beaumont and Nichols (1996), and Excoffier et al. (2009). We considered the loci identified by two or more approaches as loci subject to balancing selection. However, although the concept of balancing selection is well established (Charlesworth 2006), there are still methodological limitations for the identification of balancing selection in hitchhiking mapping (Akey et al. 2002, Beaumont and Nichols 2008; Hansen et al. 2010). Therefore, the analysis of balancing selection was aimed to distinguish putative loci subject to balancing selection from neutral loci.

Divergence at Neutral and Selected Loci
To examine the patterns of population differentiation at the loci identified to be under directional selection in the global comparisons, $F_{ST}$ distances were estimated for the respective loci as well as the putatively neutral loci using POPULATIONS 1.2 (Langella 2002). Neighbor joining trees were constructed from the distance matrices.

We investigated the relative contribution of habitat type and local area in adaptive population divergence with analyses of molecular variance (AMOVAs; Excoffier et al. 1992). The analysis was conducted separately for the four groups of loci classified by the outlier tests: putatively neutral, false positives, directional, and balancing selection loci. To evaluate the contribution of habitat type and local area, populations were grouped according to their habitat type (marine or freshwater), local area (area 1: RU, SE, or FI) and sea area (area 2: White Sea or Baltic Sea). We expected that if these contributions are significant, the among-group variance components in directionally selected loci should exceed those observed in neutral loci. The analysis was performed using ARLEQUIN 3.5 (Excoffier and Lischer 2010) with 1000 permutations.

Identification of Loci Associated with Environmental Salinity
In addition to outlier analyses, we used the approach of Joost et al. (2007) to investigate whether outlier loci were associated with differences in environmental salinity. This approach identifies potentially effective selection pressures from the environment based on univariate logistic regressions between allele frequencies and environmental variables. The analysis was performed using SAM (Joost et al. 2008). Salinity was set to 0 ppt in the freshwater populations, 25 in the RU-LEV, 4.5 in the SE-BOL, and 6.5 in the FI-HEL, based on data reported in the National Oceanographic Data Center (http://www.nodc.noaa.gov/) and the Baltic Sea Portal (http://www.itameriportaali.fi/en_GB/). The annual fluctuations of salinity in the marine areas are slight—in the range of approximately 10% (http://www.nodc.noaa.gov/; http://www.itameriportaali.fi/en_GB/). Both likelihood ratio and Wald tests, as well as Bonferroni corrections, were applied to determine significant associations (Joost et al. 2008).

Results
Genetic Variation and Differentiation
A total of 1410 alleles were observed in six populations across the 112 loci, with an average of 12.6 alleles per locus (range: 2–67; supplementary table S3, Supplementary Material online). The average expected heterozygosity ranged from 0.007 to 0.943 among the loci (supplementary table S3, Supplementary Material online). Heterozygosity did not differ among the populations (0.536–0.580; analysis of variance [ANOVA], $P = 0.640$), whereas allelic richness varied from 4.8 to 7.1 (ANOVA, $P < 0.01$; supplementary table S3, Supplementary Material online). Although deviations from Hardy–Weinberg equilibrium were detected in 58 of 640 comparisons before correcting for multiple tests ($P < 0.05$), significance levels in most ($n = 42$) of the 58 comparisons were relatively low ($P = 0.01–0.05$). After sequential Bonferroni corrections, no evidence for Hardy–Weinberg departures was obtained at any locus in any of the populations, except for the locus GS1 (supplementary table S3, Supplementary Material online). Significant departures at this locus in multiple populations suggested the possible presence of null alleles. Nonetheless, this locus was not included in the analyses of adaptive and neutral population differentiation based on the results of outlier tests (see below). The bottleneck tests of Cornuet and Luikart (1996) and Luikart et al. (1998) did not provide any evidence for recent population bottlenecks in any of the populations. Although the average level of genetic differentiation among the six populations was low ($F_{ST} = 0.073$), the locus-specific $F_{ST}$ estimates varied considerably, ranging from $-0.008$ to 0.288 (supplementary table S4, Supplementary Material online). The freshwater populations exhibited a slightly higher average $F_{ST}$ (0.097) than the marine populations (0.070; supplementary table S4, Supplementary Material online). The $F_{ST}$ value between pairs of marine and freshwater populations varied from 0.017 to 0.045 among the areas (supplementary table S4, Supplementary Material online).

Detection of Outlier Loci
Under global outlier tests using all six populations, signatures of directional selection were detected at 14, 13, and
17 loci with ln RH, BAYESCAN, and LOSITAN, respectively (fig. 2). In total, 27 loci were suggested to be under directional selection by at least one of the outlier tests applied (table 1 and supplementary table S5, Supplementary Material online). Of these, 11 were identified by two or three tests and considered as strong candidate loci influenced by directional selection (table 1). These 11 loci included four gene-based markers (FGF6aPP, FGF8, UDPGT, and GR1) for fibroblast growth factor isoforms 6 and 8 (FGF6 and FGF8), UDP glucuronosyltransferase 1 (UGT), and glucocorticoid receptor isoform 1 (GR1) genes (supplementary table S2, Supplementary Material online). The results of a complementary analysis with ARLEQUIN were not fully congruent with those of LOSITAN. Signatures of directional selection were detected at 10 of the 17 loci indicated by LOSITAN (table 1 and supplementary
Of the 15 loci, 12 exhibited low $F_{ST}$ values (0.010–0.046) associated with high heterozygosity (mean across populations, 0.764–0.943; supplementary table S3, Supplementary Material online). The low $F_{ST}$ values at these 12 loci might be artifacts due to high mutation rates, as suggested in previous studies (Beaumont 2008; Excoffier et al. 2009; Hansen et al. 2010). Nevertheless, heterozygosity at the other three loci (Ppgm19, Ppgm35, and HSC70PP) was lower (0.387, 0.463, and 0.154, respectively) than that observed at the aforementioned 12 loci. Therefore, low $F_{ST}$ values at these three loci (−0.008, 0.002, and −0.001, respectively) are unlikely due to higher mutation rates. These three loci were gene-based markers for ferritin H subunit I, heat shock protein 70 kDa beta, and heat shock cognate 70 kDa genes (supplementary table S2, Supplementary Material online).

In outlier analyses with a set of marine populations, one gene-based (GR1) and three random loci (Stn46, Stn19, and Stn96) were indicated to be under directional selection by two or three outlier tests (table 1). In a set of freshwater populations, directional selection was identified at three gene-based loci (FGF8, UDPGT, and FGF6aPP) and two random loci (Stn79 and Stn195; table 1). All the nine loci identified in these comparisons were indicated to be under directional selection in the global comparisons. On the other hand, outlier tests with pairs of marine and freshwater populations in each area detected five outliers in RU area, four in SE, and seven in FI (table 1). Of the 14 outliers detected in this analysis, seven (FGF8, UDPGT, Ppgm9, Ppgm25, Ppgm30, MYHb, and PVALBa) were found in gene-based markers for FGF8, UGT, V-type H+–ATPase subunit A (ATP6V1Aa), growth hormone receptor 2 (GHR2), heat shock protein 25 kDa (HSP25), myosin heavy chain (MYHb), and parvalbumin (PVALBa) genes (table 1 and supplementary table S2, Supplementary Material online). Of the 14 outliers, 9 were not identified as directionally selected loci in the global comparisons (table 1). Hence, we considered these nine loci as weakly selected loci relative to the other five loci identified in the global comparisons. Yet, signatures of directional selection for Ppgm25 and Stn198 were detected in two independent areas (i.e., RU and FI; table 1). Whereas Stn198 was a random genomic marker, Ppgm25 was a gene-based marker for GHR2.

In the outlier analyses with all and subsets of populations, signatures of directional selection were identified at 20 loci by two or three tests (directionally selected loci; table 1) and at 29 loci by a single test (potentially false positives). Of the 14 outliers, 9 were not identified as directionally selected loci in the global comparisons (table 1). Hence, we considered these nine loci as weakly selected loci relative to the other five loci identified in the global comparisons. Yet, signatures of directional selection for Ppgm25 and Stn198 were detected in two independent areas (i.e., RU and FI; table 1). Whereas Stn198 was a random genomic marker, Ppgm25 was a gene-based marker for GHR2.

### Table 1. Loci Indicated to be Under Directional Selection by Outlier Tests in Global Comparisons, in Three Populations with the Same Habitat Type from Different Areas, and in Population Pairs from Different Habitat Types Within Each Area.

<table>
<thead>
<tr>
<th>Comparison Locus</th>
<th>Global All</th>
<th>Among Local Areas with Same Habitat Type</th>
<th>Between Habitat Types within a Local Area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RH/BS/LS/AR</td>
<td>Marine</td>
<td>Freshwater</td>
</tr>
<tr>
<td>FGF8</td>
<td>5/3/1/1/1</td>
<td>-/-/-</td>
<td>2/-/-/-</td>
</tr>
<tr>
<td>UDPGT</td>
<td>5/3/1/1/1</td>
<td>-/-/-</td>
<td>2/-/-/-</td>
</tr>
<tr>
<td>Stn46</td>
<td>3/1/1/1/1</td>
<td>2/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Stn79</td>
<td>4/1/1/1/1</td>
<td>-/-/-/-</td>
<td>2/-/-/-</td>
</tr>
<tr>
<td>GGF6aPP</td>
<td>8/3/1/1/1</td>
<td>2/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Stn19</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Stn198</td>
<td>4/1/1/1/1</td>
<td>2/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Global and habitat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAest3</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Global</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Stn108</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Habitat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ppgm9</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Ppgm25</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Ppgm30</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>MYHb</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>PVALBa</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>GAest7</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Stn102</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Stn130</td>
<td>5/-/-/-/-/-</td>
<td>2/-/-/-</td>
<td>-/-/-/-</td>
</tr>
<tr>
<td>Stn198</td>
<td>-/-/-/-/-/-</td>
<td>-/-/-/-</td>
<td>-/-/-/-</td>
</tr>
</tbody>
</table>

**Note:** RH, ln RH; BS, BAYESCAN; LS, LOSITAN; AR, ARLEQUIN; DS, DETSEL. *P < 0.10 (only for BAYESCAN), **P < 0.05, and ***P < 0.01. Significant results obtained in two or three tests (AR was not applied) were indicated by gray shadings. In RH results of global and local area comparisons indicate the numbers of significant pairwise combinations (only ≥3 and ≥2 were shown, respectively).
positives; supplementary table S5, Supplementary Material online). Excluding further 15 loci suggested to be under balancing selection, the remaining 53 loci were considered as being neutral (supplementary table S2, Supplementary Material online). No difference was found in an incidence of directional or balancing selection among marker types (i.e., gene-based, random EST, and random genomic markers; G-test, \( P = 0.259 \) or \( P = 0.330 \), respectively).

### Patterns of Adaptive Divergence

The average \( F_{ST} \) among the six populations at the 53 neutral loci was 0.059. The level of genetic differentiation was higher between White Sea (RU) and Baltic Sea (SE and FI) areas (0.091) compared with that within each area (White Sea: 0.005; Baltic Sea: 0.022; fig. 3). As expected, the loci indicated to be under directional selection in the global comparisons exhibited a higher degree of differentiation than the neutral loci (fig. 3). Some directionally selected loci exhibited a large divergence between the sea areas, and the heterozygosity tended to be lower in the White Sea area than in the Baltic Sea area (i.e., GR1, FGF6aPP, Stn19, Stn96, and Stn195; fig. 3). However, differentiation patterns at other loci were variable (fig. 3). For example, a relatively large divergence was found between RU-BOL and other populations at FGF8 and UDPGT, between SE-KRO and others at Stn79, and between RU-LEV and others at GAest3. In the case of Stn46, FI-MAT showed a large differentiation from other populations of the Baltic Sea areas. Heterozygosity at UDPGT tended to be lower in RU-BOL than other populations, whereas RU-BOL or SE-KRO showed higher heterozygosity at FGF8 or Stn79 than other populations, respectively (fig. 3).

In the 53 neutral loci, a significant proportion of the total genetic variance resided among the local areas (cf. area 1; 5.4%, \( P < 0.001 \)) or between the White Sea and Baltic Sea areas (cf. area 2; 7.5%, \( P < 0.001 \)), but the contribution of habitat type (cf. marine vs. freshwater) was not significant (\( P = 1.00 \); fig. 4). The patterns in the 29 potentially false-positive loci were similar to those of the neutral loci (fig. 4). However, in the 20 directionally selected loci, a larger proportion of genetic variance resided among the local areas (12.2%, \( P < 0.001 \)) and between the sea areas (17.7%, \( P < 0.001 \)) as compared with the neutral loci (fig. 4). In contrast, the contribution of habitat type was not significant (\( P = 0.99 \); fig. 4). When the analysis was restricted to the 11 directionally selected loci identified under the global comparisons, the contributions of the local and sea areas increased (area 1: 18.5%, \( P < 0.001 \); area 2: 26.1%, \( P < 0.001 \)), although no significant contribution of habitat type (\( P = 0.99 \)) was observed. In the 15 loci under balancing selection, a very low proportion of genetic variance was identified in the local or sea area (1.2% or 1.4%, \( P < 0.001 \), respectively), and the contribution of habitat type was not significant (\( P = 0.96 \); fig. 4).

### Association between Environmental Salinity and Loci

The SAM identified significant associations between environmental salinity and alleles at seven loci (table 2). Of these, Ppgm25 and Stn198 were indicated to be under directional selection in the outlier analyses (table 1). Signatures of selection for these two loci, Ppgm25 and Stn198, were identified in comparisons between habitat types in both RU and FI areas (table 1).

### Discussion

Targeting a large number of functionally important genes, hitchhiking mapping revealed several footprints of directional selection in the genome of nine-spined sticklebacks. An incidence of directional selection identified in this study was higher (9.1%) than that observed in most previous studies (1–10%; see below). In general, our results demonstrated a stronger impact of geographic area than habitat type on adaptive population differentiation. In particular, a large divergence between the White Sea and Baltic Sea areas was found at some of the loci under directional selection. Nevertheless, two loci were indicated to be under directional selection in different habitats in multiple areas, and SAM identified an association between environmental salinity and allele frequencies at these loci. In what follows, we will discuss these and related issues in light of the results obtained.

### Identification of Directional Selection

In the global comparisons, incidence of signatures of directional selection varied from 8.9% to 15.2% of loci depending on the method applied. In particular, a large number of directionally selected loci were detected by the approach of Beaumont and Nichols (1996), which has frequently been used in previous studies (reviewed in Holderegger et al. 2008). Excoffier et al. (2009) pointed out that much lower numbers of outlier loci were identified under a hierarchical island model than nonhierarchical models. In line with this, several outlier loci identified by the approach of Beaumont and Nichols (1996) were not detected by that of Excoffier et al. (2009). In contrast, most of the outliers detected by both methods were identified to be under directional selection in other methods employed. By taking into account conceptually different approaches, we identified 11 loci consistently showing evidence of being under directional selection in Fennoscandian nine-spined sticklebacks. Hitchhiking mapping in various organisms has typically found footprints of directional selection for 1–10% of the loci examined (reviewed in Holderegger et al. 2008; Stinchcombe and Hoekstra 2008; Nosil et al. 2009), although different studies vary in the numbers and designs of populations, the outlier tests applied, and their criteria for determining outliers. Even after eliminating potential false positives, an incidence of directional selection (9.1%) identified in our study is relatively high compared with that observed in previous studies (e.g., Campbell and Bernatchez 2004; Vasemägi et al. 2005; Bonin et al. 2006; Kane and Rieseberg 2007; Mäkinen, Cano, and Merilä 2008; Namroud et al. 2008; Nielsen et al. 2009).

Outlier tests rely on the assumption that selection affects particular regions of the genome whereas neutral
demographic processes affect the whole genome in a more or less homogenous manner (Cavalli-Sforza 1966). A substantial challenge is to clearly distinguish selection from neutral effects. Despite continuous improvements in the methodology, there are still several problems with outlier tests. For example, one serious concern is that the inclusion of severely bottlenecked populations can lead to a high rate of false-positive signatures of selection (Teshima et al. 2006; Wiehe et al. 2007; Foll and Gaggiotti 2008). In our study populations, there was no difference in the level of heterozygosity, but allelic richness tended to be lower in the freshwater populations than in the marine populations, as typically observed in fish (Ward et al. 1994; DeWoody and Avise 2000). However, bottleneck tests provided no evidence of recent population bottlenecks. Therefore, it is unlikely that our outlier tests were unduly influenced by the effect of severe population bottlenecks. Another crucial issue is that outlier tests tend to underestimate demographic effects, which can increase the number of false positives (Excoffier and Ray 2008; Hofer et al. 2009; Excoffier et al. 2009). In our data set, varying algorithms and assumptions of different methods resulted in different proportions of outliers. Nonetheless, the number of false positives can be minimized by comparing results from different approaches (e.g., Vasevății et al. 2005; Bonin et al. 2006; Kane and Rieseberg 2007; Paris et al. 2010), by using a suitable demographic model (Excoffier et al. 2009), and by evaluating replicated comparisons of particular types of population pairs (Nosil et al. 2009; see also below).
Patterns of Adaptive Divergence

AMOVAs revealed strong geographic impacts on the patterns of population differentiation at the 11 candidate loci under directional selection. In particular, the overall contribution of sea area was stronger than that of local area. In contrast, the effect of habitat type was not significant. Hence, our results suggest that geographic area rather than habitat type has been acting as a central force in shaping adaptive population divergence in Fennoscandian nine-spined sticklebacks. Based on the patterns of genetic differentiation, five loci (GR1, FGF6aPP, Stn19, Stn96, and Stn195) exhibited a large divergence between the White and Baltic Sea areas. Because a locus affected by directional selection is expected to be less diverse than that evolving under neutrality (Schlöterer 2002; Kauer et al. 2003), lower variability at the five loci in the populations of the White Sea area than those of the Baltic Sea area implies that natural selection has been acting in the White Sea area. In line with this view, signatures of directional selection were not detected at any of these loci in outlier analyses with population pairs within local areas. Yet, the results of outlier tests with subsets of marine or freshwater populations were not fully congruent with the patterns of genetic divergence and variation. For instance, although it seems likely that the same natural selection has been acting on most populations in different areas and habitat types, it might have been targeted by natural selection in local areas.

Four candidate loci were identified with gene-based markers developed for FGF6, FGF8, GR1, and UGT. The primary biological functions of FGF6 and FGF8 are known to be involved in development in fish as well as in other taxa (Rescan 2005; Jovelin et al. 2010). Based on the expression patterns, FGF6 is supposed to be involved in the development of muscle (Rescan 1998). FGF8 has been shown to have a direct function in induction and differentiation of heart tissue (Reifers et al. 2000; Jovelin et al. 2007). GR1 is known to mediate the actions of glucocorticoids, such as stress response, osmoregulation, and immune function (Mommsen et al. 1999; Greenwood et al. 2003; Evans et al. 2005; Stolte et al. 2008). In teleosts, UGT is presumed to play an important role in the removal of toxins in the liver (Leaver et al. 2007). Such biological functions and traits might have been targeted by natural selection in local areas. Nevertheless, these candidate genes should be further investigated to determine if they are the real targets of natural selection. However, on the basis of their important biological functions, comparative analyses of gene expression in different populations, as well as association mapping with divergent phenotypes, should be informative for this purpose.

![AMOVAs grouping by habitat type (habitat), local area (area 1), and sea area (area 2). Black, gray, and white shadings represent the among-group, among-population, and within-population variance components, respectively. (A) Neutral loci, (B) loci under directional selection, (C) false-positive loci, (D) loci under balancing selection.](image-url)
Locis under Habitat-Dependent Selection

Variations in environmental salinity can have direct and unavoidable impact on susceptible cells, tissues, and organ systems in fish because their respective plasma ion concentrations differ from those of the external environment, creating internal ionic imbalances across variably permeable membranes that can disrupt crucial physiological processes (Marshall and Groell 2006). As such, we predicted that divergent selection has been acting between marine and freshwater habitats at the inception of this study. However, directional selection for the 11 candidate loci identified in the global comparisons was not obviously dependent on a particular habitat type. Yet, outlier analyses with population pairs from different habitat types revealed two loci showing signatures of directional selection in the multiple areas with the exception of the Swedish area where salinity concentration in marine habitat is lower than in the other areas. Although extensive multiple outlier testing is a general problem to increase the number of false positives (Beaumont 2008), the detection of selection signatures across multiple population pairs is unlikely to arise via nonselective factors such as genetic drift, mutation rate variation, or type I error (Luikart et al. 2003; Campbell and Bernatchez 2004; Storz 2005; Vasmägi et al. 2005; Bonin et al. 2006; Nosil et al. 2009). The fact that signatures of selection were detected in the same multiple areas strongly suggests that these two loci are involved in adaptation to different habitat types. Moreover, the SAM revealed that allele frequencies of both loci are associated with environmental salinity. These results implicate that environmental salinity has been affecting these loci as a selective force, although the selection pressure might not be as strong as that for the 11 loci identified in the global comparisons.

One of the two candidate loci was identified with a gene-based marker located within GHR2. The pleiotropic effects of growth hormone result from its interaction with the growth hormone receptor (i.e., GHR2; Jiao et al. 2006). Several studies have shown that growth hormone is involved in somatic growth and osmoregulation in fish (reviewed in Björnsson 1997; McCormick 2001). As for osmoregulatory function, growth hormone is known to enhance hyperosmotic ability by stimulating the proliferation and differentiation of chloride cells and the expression of Na⁺, K⁺-ATPase in gills (Sakamoto et al. 1993; Madsen et al. 1995; McCormick 2001). Nilsen et al. (2008) showed that plasma growth hormone and gill GHR2 expression levels increase in response to hyperosmotic environments in Atlantic salmon (Salmo salar). In addition, they found that both growth hormone and GHR2 levels are higher in anadromous fish than in landlocked fish. Given its biological function, GHR2 is a promising candidate gene that may be responsible for genetic adaptation to different saline environments in nine-spined sticklebacks. The other candidate locus was identified with a random marker located within the DEAD box polypeptide 56 gene. Although this gene is involved in early development in zebrafish (Danio rerio; Amsterdam et al. 2004), its biological function is not well known in other fish species. Because this candidate locus was identified with a random marker, further investigation around this region will be undertaken to determine a candidate gene.

Our hitchhiking mapping approach employed genetic markers developed for functionally important genes, including a large number of genes responsible for osmoregulatory function (Shikano, Ramadevi, et al. 2010). These genes are known to change their expression levels in response to varying environmental osmotic conditions (e.g., Tipsmark et al. 2004; Scott et al. 2005; Tse et al. 2006; Tomy et al. 2009). An increasing number of studies have shown that the level of gene expression is highly heritable under both cis- and trans-regulation (Petretto et al. 2006; Dixon et al. 2007; Stranger et al. 2007). Larsen et al. (2008) demonstrated that several genes are differently expressed between Atlantic and Baltic Sea populations of European flounders (Platichthys flesus), suggesting adaptive differentiation of gene expression patterns in response to salinity gradients. Hitchhiking mapping identifies the genomic regions of any types of variation (e.g., amino acid substitution or transcriptional regulation) affected by natural selection, but it cannot detect selection on the expression levels of the genes in question if they are trans-regulated. In our study, no signatures of selection for most of the osmoregulatory genes investigated (see also Shikano, Ramadevi, et al. 2010) suggest that directional selection has not been acting on these genes. However, it should be noted that this does not imply the absence of selection on the expression levels of these genes because they can be trans-regulated.

Nine-Spined Versus Three-Spined Sticklebacks

Because of their similar ecological and morphological characteristics, three-spined and nine-spined sticklebacks provide good models to study genetic mechanisms underlying convergent evolution (Shapiro et al. 2006, 2009). Mäkinen, Cano, and Merilä (2008) conducted hitchhiking mapping with 105 markers in European, mainly Fennoscandian, three-spined sticklebacks. They identified signatures of directional selection for two novel genomic regions and one previously known reference gene (ectodysplasin: EDA) in marine and freshwater populations. In particular, a strong signature of directional selection was found for EDA, which is responsible for the reduction of armor plate number in this species (Colosimo et al. 2005). In nine-spined sticklebacks, Shapiro et al. (2009) demonstrated that the major loci controlling variation in armor plate number are located in different genomic regions than EDA (Shapiro et al. 2009). In our study, a signature of directional selection was not identified for EDA. Therefore, although the role of EDA in adaptation and evolution is crucial in three-spined sticklebacks (Colosimo et al. 2005; Barrett et al. 2008; Kitano et al. 2008), its impact in nine-spined sticklebacks might not be as strong as in three-spined sticklebacks.

Of the 112 markers used in our study, one gene-based (EDA) and 31 random markers were used in the study of Mäkinen, Cano, and Merilä (2008). Although they did not
find any signatures of directional selection for these random loci in three-spined sticklebacks, six loci (Stn46, Stn79, Stn19, Stn96, Stn195, and GAest3) were indicated to be subject to directional selection in our study. Therefore, directional selection might have been acting on several different genes in these stickleback species in Fennoscandia. Although both species have likely colonized northern Europe after the last glacial maximum, molecular data suggest that their colonization patterns were largely different (Mäkinen et al. 2006; Mäkinen and Merilä 2008; Shikano, Shimada, et al. 2010). For instance, northern Europe was colonized by ancestral marine three-spined sticklebacks, whereas nine-spined sticklebacks may have colonized this region via freshwater. Because colonization of newly formed habitats is followed by natural selection and thereby adaptive genetic changes (Hewitt 1996, 2000), different colonization histories for nine-spined and three-spined sticklebacks could have led to differential selection regimes in Fennoscandia. Alternatively, different genes might be responsible for responding to the same selection regimes in these species. Their contemporary sympatric distributions in Fennoscandia suggest that both species are subject to similar environmental factors, such as osmotic and temperature stresses. Thus, investigating footprints of natural selection on large numbers of functionally important genes in sympatric populations of these species may provide insights into the relative roles of colonization history and contemporary environmental conditions in shaping adaptive population divergence, as well as genetic mechanisms underlying their evolutionary convergence.

Conclusions
The impact of environmental salinity on adaptation and evolution has been little explored for many aquatic organisms. We addressed this issue by integrating functional information of genes into a population genomics approach. Our results demonstrated that adaptive population divergence has been strongly shaped by the impact of geographic area and less by habitat type in nine-spined sticklebacks. Nevertheless, two loci, including growth hormone receptor 2, are potentially responsible for adaptation to different habitat types, implying that environmental salinity has been affecting them as a selective force. These loci are promising candidates for further investigating the molecular mechanisms of adaptation to marine and freshwater environments.

Supplementary Material
Supplementary tables S1–S5 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
We thank Göran Englund, Abigel Gonda, Gábor Herczeg, Tuomas Leinonen, John Loehr, Yukinori Shimada, and people at the White Sea Biological Station for help in obtaining samples. Thanks are also due to Scott McCairns for his comments. This work was supported by the Academy of Finland, University of Helsinki and the Japan Society for the Promotion of Science.

References


Hofer T, Ray N, Wegmann D. 2009. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. Ann Hum Genet. 73:95–108.


