

RESEARCH ARTICLE

Open Access

# Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks

Baocheng Guo<sup>1\*</sup>, Jacquelin DeFaveri<sup>1</sup>, Graciela Sotelo<sup>1,2</sup>, Abhilash Nair<sup>1,3</sup> and Juha Merilä<sup>1</sup>

## Abstract

**Background:** The degree of genetic differentiation among populations experiencing high levels of gene flow is expected to be low for neutral genomic sites, but substantial divergence can occur in sites subject to directional selection. Studies of highly mobile marine fish populations provide an opportunity to investigate this kind of heterogeneous genomic differentiation, but most studies to this effect have focused on a relatively low number of genetic markers and/or few populations. Hence, the patterns and extent of genomic divergence in high-gene-flow marine fish populations remain poorly understood.

**Results:** We here investigated genome-wide patterns of genetic variability and differentiation in ten marine populations of three-spined stickleback (*Gasterosteus aculeatus*) distributed across a steep salinity and temperature gradient in the Baltic Sea, by utilizing >30,000 single nucleotide polymorphisms obtained with a pooled RAD-seq approach. We found that genetic diversity and differentiation varied widely across the genome, and identified numerous fairly narrow genomic regions exhibiting signatures of both divergent and balancing selection. Evidence was uncovered for substantial genetic differentiation associated with both salinity and temperature gradients, and many candidate genes associated with local adaptation in the Baltic Sea were identified.

**Conclusions:** The patterns of genetic diversity and differentiation, as well as candidate genes associated with adaptation, in Baltic Sea sticklebacks were similar to those observed in earlier comparisons between marine and freshwater populations, suggesting that similar processes may be driving adaptation to brackish and freshwater environments. Taken together, our results provide strong evidence for heterogenic genomic divergence driven by local adaptation in the face of gene flow along an environmental gradient in the post-glacially formed Baltic Sea.

**Keywords:** *Gasterosteus aculeatus*, RAD-sequencing, SNP, population differentiation, local adaptation, Baltic Sea

## Background

While local adaptation is likely to be of commonplace occurrence, demonstrating its occurrence can be difficult and take substantial research efforts [1-3]. Traditionally, studies of local adaptation have been built upon quantitative genetic approaches that make use of common garden experiments and statistical genetics methods to infer genetic differentiation in phenotypic traits (e.g., [4-6]). Quantitative genetic methods have also been increasingly combined with population genetic tools to infer local adaptation (e.g., [7-10]). More recently, advances in genomic

technologies have made it possible to identify candidate genomic regions underlying local adaptation (e.g., [11-14]). Among such approaches are genome scan or outlier detection methods (e.g., [15-17]), which allow inferences about adaptive differentiation to be made without the application of common garden experiments.

Outlier detection methods have become particularly popular in identifying population structuring and adaptive differentiation in marine fishes, which generally show very low levels of genetic differentiation in neutral marker genes [18-22] and in which common garden experiments are often logistically demanding, if not impossible, to conduct (but see [23-27]). However, with few notable exceptions (e.g., [28-32]), genome scan studies of marine fishes have typically been limited to tens – or

\* Correspondence: baochengguo@gmail.com

<sup>1</sup>Department of Biosciences, Ecological Genetics Research Unit, University of Helsinki, PO Box 65, Helsinki FI-00014, Finland

Full list of author information is available at the end of the article

in rare cases hundreds – of loci (e.g., [33–38]). As such, genome-wide patterns of diversity and divergence cannot thoroughly be explored, especially when the markers employed are anonymous. On the other hand, when high-throughput approaches have been used to screen thousands of loci across the genome, only a small number of populations have been under focus – generally those with obvious differentiation [29,39–41]. Hence, high-throughput population genomic studies aimed at detecting adaptive differentiation in marine fishes are rare, especially those employing a comprehensive sampling scheme. The latter point is particularly relevant in the context of seascape genetics, which aims to integrate environmental features with population genetic data to assess their impact on the genetic structure of marine populations [42,43]. In such approaches, sampling of multiple populations across environmental gradients becomes critical for inferences about genotype–environment associations.

There has been considerable interest in studying local adaptation and genetic differentiation in three-spined sticklebacks (*Gasterosteus aculeatus*; e.g., [25,44–54] and reviewed in [55]). However, most of these studies – particularly those using high-throughput methods [13,56–65] – have focused on marine–freshwater or lake–stream differentiation, with less focus on differentiation within the ancestral marine environment (but see [25,33,47,48,56,58,66]). Nevertheless, the studies thus far – irrespective of the approach used (common garden: [25];  $Q_{ST}$ – $F_{ST}$ : [48]; population genetics or genomics: [33,47,51,53,58]) – suggest that there is substantial population structuring and local adaptation also in marine three-spined sticklebacks (but see [56,60]). This is most apparent in the thoroughly studied Baltic Sea seascape, which is characterized by steep salinity and temperature gradients. Using microsatellite markers, DeFaveri et al. [47] uncovered evidence for heterogeneous genomic differentiation and adaptive population structuring in three-spined sticklebacks across the Baltic Sea, suggesting that environmental heterogeneity can generate disruptive selection despite the considerable gene flow in this highly connected marine environment. In fact, the unique ecosystem of the Baltic Sea has attracted the attention of many evolutionary and population genetics studies that have also sought to understand local adaptation and genetic structuring of Baltic Sea organisms (e.g., [35,36,67,68]; reviewed in [69]). However, as yet, studies based on genome-wide characterizations of variability with high-throughput approaches and comprehensive sampling of Baltic Sea populations are still missing (but see [29,31]). Hence, the spatial scale of genetic structuring at the genome-wide level cannot truly be defined in any Baltic Sea organism.

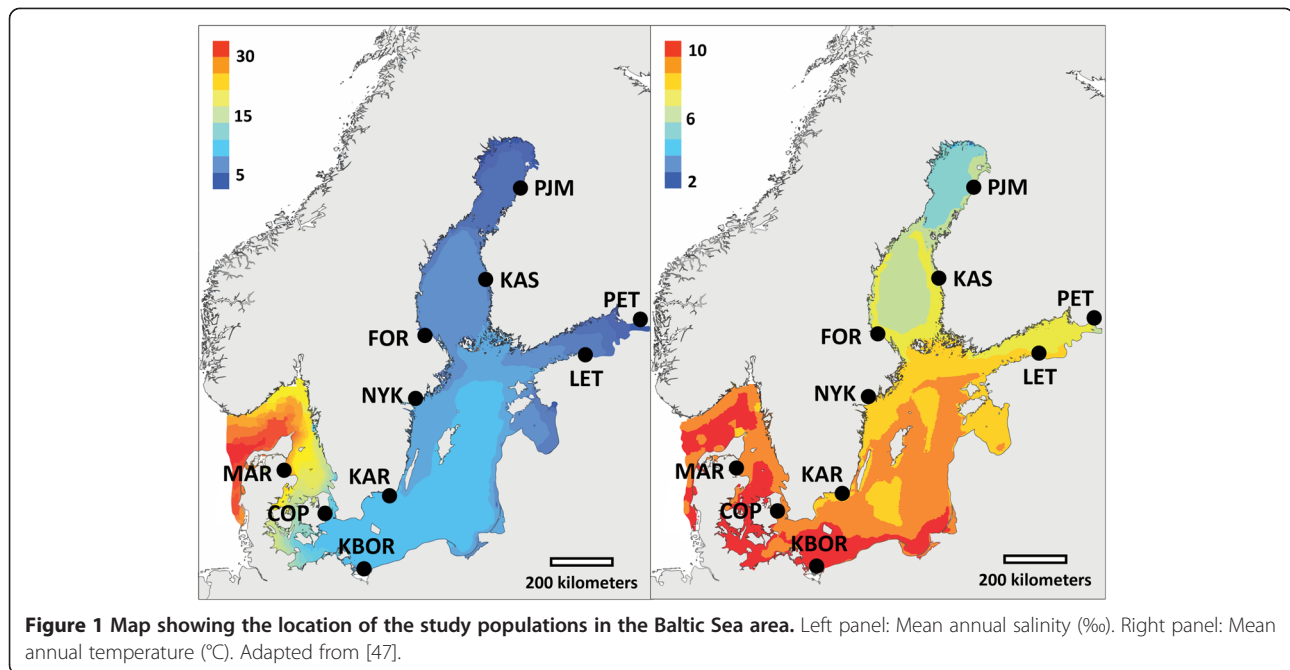
The main aim of this study was to investigate genome-wide patterns of genetic variability and differentiation in

marine three-spined sticklebacks across the Baltic Sea – a relatively young sea area with steep environmental gradients, subject to many earlier low-throughput studies in local adaptation (reviewed in [25,69]) and genetic differentiation (reviewed in [70,71]). In particular, we were interested in characterizing genomic variation across study sites that are connected by gene flow, and identifying genomic regions showing footprints of directional (and balancing) selection in association with key environmental parameters (viz. temperature and salinity). In addition, we were interested in knowing if the detected outliers correspond to those identified to be under selection in earlier stickleback studies, not only in this particular system [47] but also in other more divergent population pairs (e.g., [13,57,58,60,61,63–65]). To this end, we utilized a pooled restriction site associated DNA sequencing (RAD-seq) approach [72] to generate polymorphism data of >30,000 single nucleotide sites across the genome of 10 three-spined stickleback populations in the Baltic Sea (Figure 1 and Table 1), and subjected the data to various outlier analyses, including BAYENV [73], which tests for associations between outliers and environmental parameters.

## Results

### Restriction site associated DNA sequencing dataset and SNPs

The three-spined stickleback genome used as a reference included 317,852 *PstI* restriction sites (Additional file 1: Table S1). The number of *PstI* restriction sites on a given chromosome was significantly correlated with chromosome length ( $r_s > 0.98$ ,  $P < 1.71 \times 10^{-16}$ ). The quality-filtered RAD-seq dataset used for alignment contained approximately 35.3 million reads, and the number of reads ranged between 2.4 and 4.4 million within each population (Additional file 1: Table S1). In total, 12.3 million reads were aligned to the reference genome, and the number of aligned reads ranged from 0.7 to 1.8 million within each population (Additional file 1: Table S1). The number of RAD-seq reads aligned on a given chromosome was significantly and positively correlated with chromosome length ( $r_s > 0.51$ ,  $P < 0.01$ ), both within and across populations. The number of SNPs within the aligned RAD-seq reads varied from 13,738 to 34,676 depending on the population, and in total 143,560 SNPs were identified across all populations (Additional file 1: Table S2). The number of SNPs identified on each chromosome was significantly and positively correlated with the number of reads aligned on the chromosome ( $r_s > 0.95$ ,  $P < 2.2 \times 10^{-16}$ ) and with chromosome length ( $r_s > 0.54$ ,  $P < 0.01$ ) within each population, as well across all the populations. An example using the population COP is shown in Additional file 2: Figure S1: the number of *PstI* restriction sites, mapped reads and number of SNPs on each chromosome are each significantly positively correlated



with chromosome length. Taken together, this suggests that the loci used in the downstream analyses are not a biased sample across chromosomes in respect to chromosome length.

#### Genome-wide genetic variation

The expected heterozygosity for all SNPs (Table 2) across all populations was 0.29721 and ranged from 0.21416 to 0.25750 within each population. The genome-wide average nucleotide diversity (Tajima's  $\pi$ ) was 0.00358 (standard error (SE) = 0.00003) across all populations and ranged from 0.00284 to 0.00332 within each population (Table 2). Average  $\pi$  values of each chromosome within each population and across all populations are listed in Additional file 1: Table S3. The highest average  $\pi$  value within each population and across all populations was detected in

chromosome XIX, which is the sex chromosome. Although the average nucleotide diversity in a given chromosome was significantly and positively correlated across populations ( $r_s > 0.56$ ,  $P < 0.01$ ), there was clear genomic heterogeneity in the levels of nucleotide diversity in different chromosomes. For example, the average  $\pi$  values for chromosomes III, VI, VII, VIII, IX, X, XI, XIV, XV, XX and XXI were larger than their genome-wide average  $\pi$  values in some populations, but smaller in other populations. The genome-wide average  $\theta_w$  value was 0.00771 (SE = 0.00005) across all populations, and ranged from 0.00317 to 0.00403 within each population (Table 2).

The genome-wide distribution patterns for  $\pi$  and  $\theta_w$  were very similar to each other across all populations (Figure 2), as well as within each population (Additional file 3: Figure S2), in respect to mean values of  $\pi$  and  $\theta_w$

**Table 1** Sampling information of populations used in this study

Location	Basin/region	Code	Coordinates	Average annual salinity (ppt)	Average annual temperature (°C)	$N^a$
Copenhagen	Kattegat	COP	55°42'05"N, 12°35'55"E	10.76	9.85	36
Forsmark	Bothnian Sea	FOR	60°24'09"N, 18°11'05"E	4.73	7.82	36
Karlskrona	Baltic Proper	KAR	56°10'13"N, 15°24'34"E	7.55	9.43	36
Kaskinen	Bothnian Sea	KAS	62°23'02"N, 25°54'52"E	3.15	7.51	36
Karsibor	Karsibor	KBOR	53°52'14"N, 14°17'03"E	7	9.69	36
Letipea	Gulf of Finland	LET	59°32'58"N, 26°35'49"E	4.63	7.80	36
Mariager Fjord	Kattegat	MAR	56°38'58"N, 09°57'05"E	26.82	9.88	36
Nyköping	Baltic Proper	NYK	58°39'04"N, 17°06'02"E	6.58	8.20	36
Petergofa	Gulf of Finland	PET	60°03'14"N, 29°58'16"E	0	7.42	36
Pyhäjoki	Bothnian Bay	PJM	64°28'42"N, 24°13'13"E	3.44	6.09	36

<sup>a</sup> $N$  is the number of individuals sequenced.

**Table 2 Summary statistics of RAD data, including and estimates of basic population genetic parameters in each population**

Population	Number of raw reads	Number of mapped reads	Number of SNPs <sup>a</sup>	Average heterozygosity <sup>b</sup>	Tajima's $\pi$	Watterson's $\theta$	Proportion of SNPs <sup>c</sup>	Number of private SNPs <sup>d</sup>
COP	3,751,563	1,220,231	22,850	0.23404	0.00303	0.00367	0.61313	64
FOR	3,548,488	1,167,788	21,196	0.22072	0.00301	0.00378	0.61854	51
KAR	3,490,058	1,765,802	34,676	0.25750	0.00275	0.00317	0.59648	71
KAS	3,875,084	1,665,176	29,022	0.22843	0.00286	0.00347	0.65246	107
KBOR	3,513,975	1,449,015	25,849	0.24355	0.00284	0.00336	0.60546	84
LET	3,452,354	789,771	15,746	0.23002	0.00332	0.00408	0.55612	21
MAR	3,419,749	1,532,462	29,957	0.23845	0.00300	0.00361	0.61900	118
NYK	2,418,564	696,118	13,738	0.23560	0.00332	0.00403	0.52402	8
PET	4,379,234	1,221,772	21,503	0.21416	0.00307	0.00391	0.62023	59
PJM	3,454,537	803,583	14,183	0.22936	0.00311	0.00380	0.55930	35

<sup>a</sup>SNPs identified using PoPoolation, used for average estimation of heterozygosity, Tajima's  $\pi$ , and Watterson's  $\theta$ .

<sup>b</sup>Average heterozygosity is the sum of  $[2 \times p \times (1 - p)]$  for all SNPs with the total number of all SNPs identified in each population, where  $p$  is the frequency of the most common allele.

<sup>c</sup>Proportion of SNPs that have read support for both alleles in each population in the total 30,871 SNPs identified using PoPoolation2.

<sup>d</sup>Number of SNPs that are unique to each population in the total 30,871 SNPs identified using PoPoolation2.

for each chromosome ( $r_s > 0.88$ ,  $P < 2.65 \times 10^{-8}$ ). However, the range of  $\pi$  and  $\theta_W$  values varied widely across the genome across all populations and also within each population. For instance,  $\pi$  ranged from 0.00001 to 0.03038 and  $\theta_W$  from 0.00017 to 0.03015 with the 100-kb sliding window (Figure 2). However, the genomic regions with high or low diversity were not consistent among populations (Additional file 3: Figure S2), suggesting genetic differentiation among populations.

To test for deviations from neutrality across all populations, Tajima's  $D$  was estimated for 261 regions across the genome across all populations, of which 255 were negative (suggestive of positive or purifying selection) and six were positive (suggesting balancing selection; Figure 2). This observation of most genomic regions having negative Tajima's  $D$  was found also within each population (Additional file 3: Figure S2). These patterns suggest an excess of low frequency variants.

### Genome-wide population differentiation

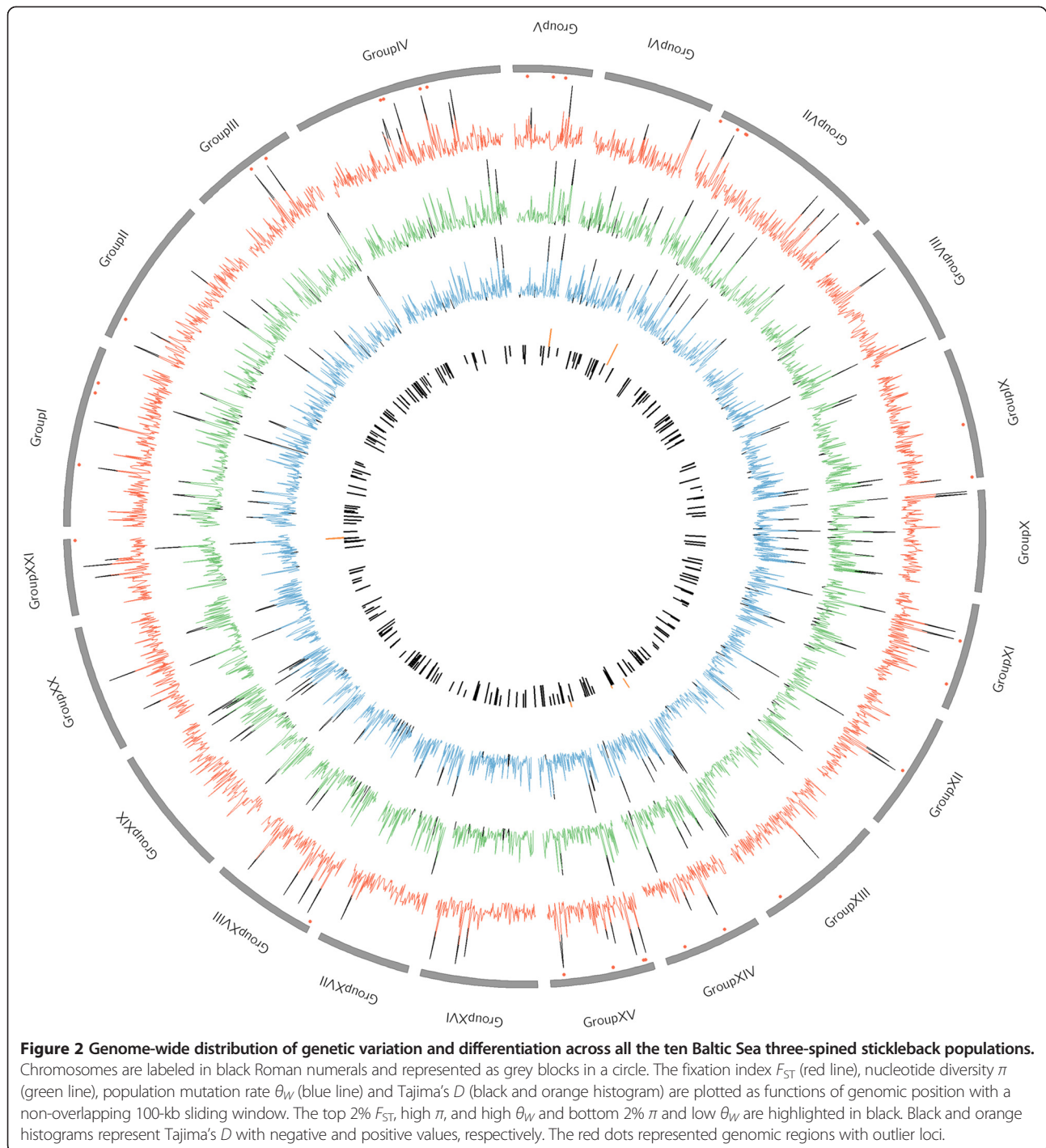
With a minor allele count of 4 across all the 10 populations and coverage between 10 and 500 within each population, 30,871 SNPs were identified by PoPoolation2. Pairwise  $F_{ST}$  values for the 30,871 SNPs estimated with a non-overlapping 100-kb sliding window across the genome yielded an overall average pairwise  $F_{ST}$  estimate of 0.02825 (SE = 0.00035) across all populations (range = 0.00178 to 0.27074; median = 0.02451). Comparison of the genome-wide profile of genetic differentiation (Figure 2) and diversity (Figure 2, Additional file 3: Figure S2) revealed certain general patterns. First, multiple genomic regions with high genetic diversity displayed low genetic differentiation (Figure 2), suggesting a role for balancing selection in maintaining high genetic diversity within and among populations. Inspection of the Tajima's  $D$  estimates

gave additional evidence for presence of balancing selection in genomic regions with elevated diversity:  $D$ -values were positive in genomic regions with high genetic diversity but low genetic differentiation (Figure 2). Second, numerous genomic regions on all chromosomes (except XIX) showing low genetic diversity exhibited a high degree of genetic differentiation (Figure 2), suggesting a varying degree of directional selection among populations.

### Candidate genes associated with adaptation

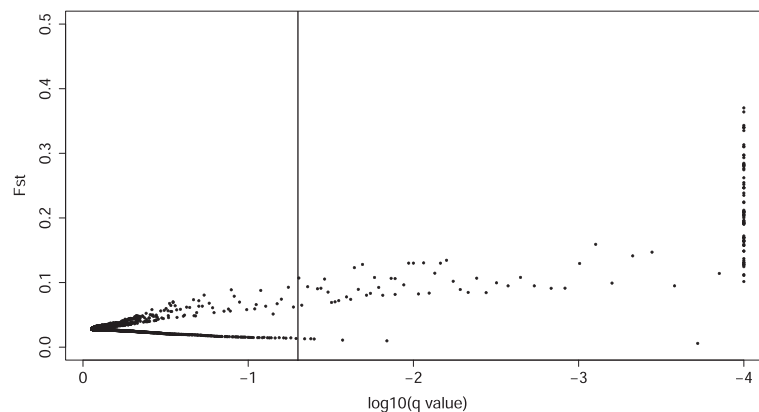
In total, a subset of 9,404 SNPs located in 1,879 genes were identified across all populations (see Methods for criteria), and were used for detecting selection footprints. Using an empirical outlier detection approach, 530 (5.64%) SNPs were found at least once in the top 0.5% tails and 112 SNPs fell within the top 0.5% tails of at least 5 pairwise  $F_{ST}$  comparisons and as such, were identified as potential SNPs under selection. Using the BayeScan approach, 136 SNPs were identified as outliers (130 directionally selected and 6 under balancing selection; Figure 3) at the false discovery rate (FDR) threshold of 0.05. In total, 94 SNPs were identified as outliers by both approaches, all of which were under directional selection and located in genomic regions of high genetic differentiation (Figure 2). Of these 94 outlier loci, 74 (79%) were located within 26 genes (Additional file 1: Table S4).

The BAYENV analysis based on the subset of 9,404 SNPs revealed that both environmental parameters (salinity and temperature) were associated with numerous SNPs across the genome (Figure 4). With the criterion of  $\log_{10}$  Bayes factor (BF) greater than 1.5 [74] as evidence for an association between environmental parameter and allelic distribution, 432 SNPs were correlated with variation in salinity (259 of which were located in 204 genes; Additional file 1: Table S4), and 413 SNPs were correlated



with temperature variation (243 of which were located in 179 genes; Figure 4 and Additional file 1: Table S4). Moreover, 161 SNPs were significantly associated with variation in both salinity and temperature. When considering the correlation with salinity variation, 89 SNPs had  $\log_{10}$  (BF) > 5 and 70 occurred in genomic regions with a high degree of genetic differentiation but low genetic diversity. The 89 loci were located in 39 genes; the SNP

within the gene *CPEB4* (ENSGACG00000018422) on chromosome IV had the highest  $\log_{10}$  (BF) (Figure 4a). When considering correlation with annual temperature variation, 73 loci had  $\log_{10}$  (BF) > 5 and 49 occurred in genomic regions with a high degree of genetic differentiation but low genetic diversity. The 73 loci were located in 35 genes; the SNP within the gene *SMAP1* (ENSGACG00000018297) on chromosome IX had the



**Figure 3** Global outlier detection among the subset of 9,404 SNPs in 10 three-spined stickleback populations from the Baltic Sea. The vertical line represents a false discovery threshold of 0.05.

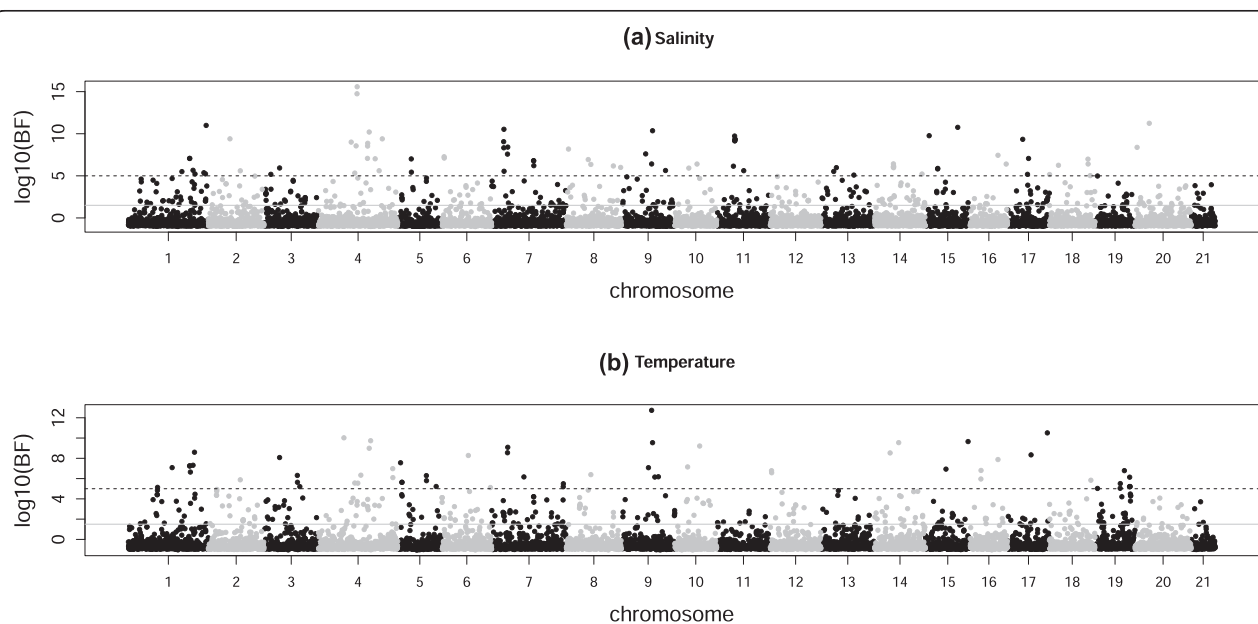
highest  $\log_{10}$  (BF) (Figure 4b). Of the 432 SNPs significantly associated with annual salinity variation, 45 were identified as outliers by BayeScan. Of the 413 SNPs significantly associated with annual temperature variation, 56 were identified as outliers by BayeScan. Of the 161 SNPs significantly associated with both salinity and temperature variation, 26 were identified as outliers by BayeScan; 14 of these were located in 11 genes.

A total of 297 genes included SNPs that were identified either as outliers or as having environmental correlations in the BAYENV analysis (Additional file 1: Table S4). These candidate genes showed a broad range of gene ontology (GO) annotations, and significant enrichment in several

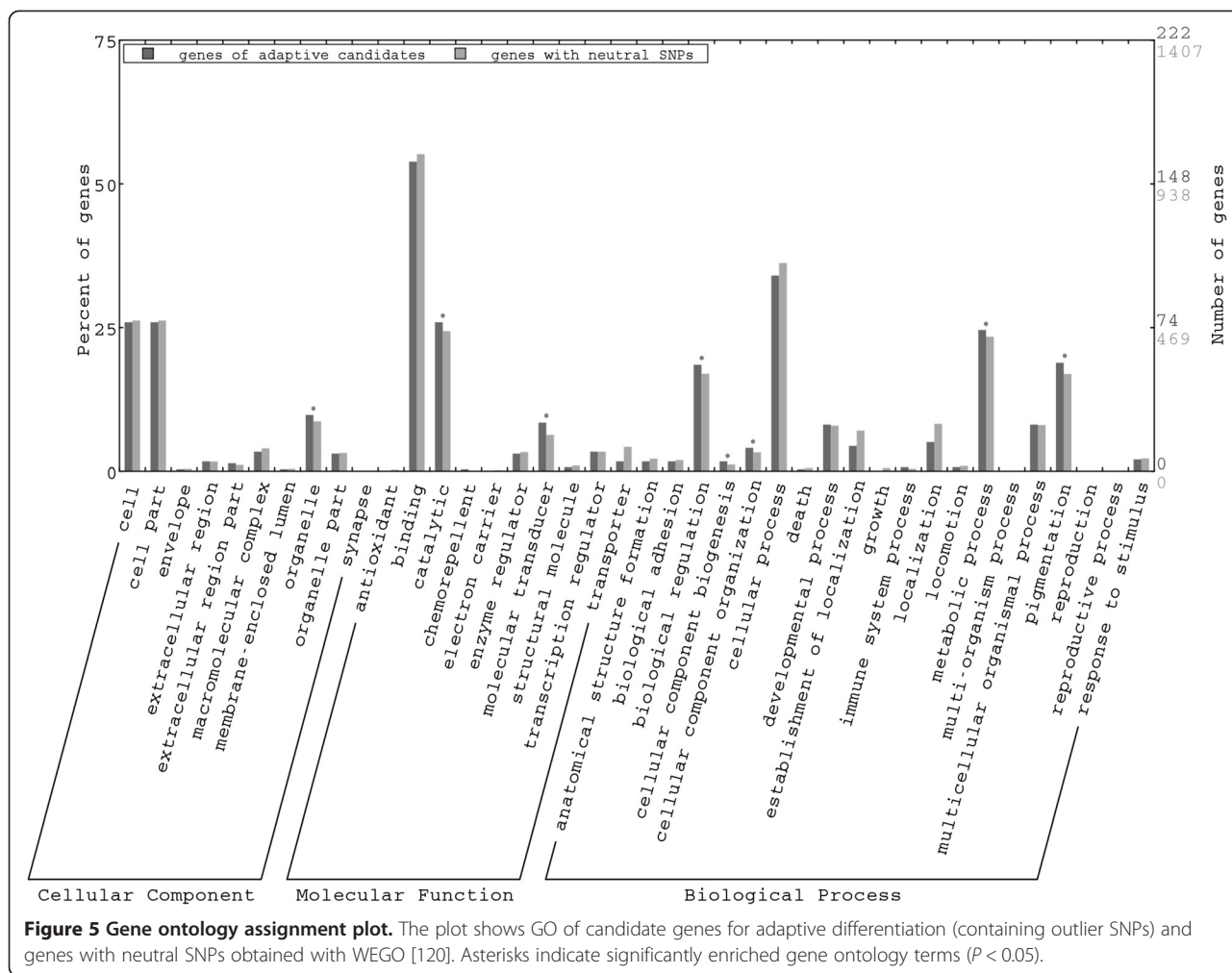
functional categories (metabolic process, catalytic activity, organelle, pigmentation and signal transduction) when compared to the genes harboring neutral SNPs ( $P < 0.05$ , Figure 5).

#### Population structure

We first took the overall average pairwise  $F_{ST}$  values among populations calculated over the subset of the 9,404 SNPs to look for population structuring across the Baltic Sea. The average pairwise  $F_{ST}$  values ranged from 0.00864 to 0.01548 (Table 3). The principal coordinates analysis (PCoA) plot revealed a geographically ordered pattern: the populations from the Danish Straits (MAR



**Figure 4** Manhattan plot of genetic differentiation associated with environmental parameters. It shows the SNP allele frequency variation associated with variation in (a) annual salinity and (b) annual temperature across different chromosomes. Grey solid lines mark lower thresholds of  $\log_{10}$  (BF) = 1.5 and black dashed lines mark higher thresholds of  $\log_{10}$  (BF) = 5.



and COP) formed one distinct group and clustered close to the southern Baltic site KBOR along PCO 1; populations from the Gulf of Bothnia (KAS and PJM) and one from the south-west Baltic Sea (KAR) formed one cluster along PCO 2, and the populations from the Gulf of Finland (LET and PET) and Baltic Proper (FOR and NYK) clustered together (Figure 6a). The population structure portrayed by the neighbor-joining tree was very similar to that seen in the PCO plot of overall average pairwise  $F_{ST}$ , but showed that populations from the Gulf of Finland (LET and PET) formed a distinct group (Figure 6b). This pattern of population structuring is consistent with that recovered by an earlier microsatellite study by DeFaveri et al. [47]. Accordingly, there was a significant correlation between pairwise genetic distances as measured by  $F_{ST}$  estimated from 40 microsatellites and the 9,404 SNPs ( $r = 0.46, P = 0.022$ , Mantel's test). However, genetic diversity across the populations as estimated from microsatellite and SNP heterozygosity (or  $\pi$ ) was uncorrelated ( $r_s < 0.25, P > 0.25$ ). A signal

of isolation by distance was detected ( $r = 0.41, P = 0.004$ , Mantel's test).

### Discussion

Amongst the most salient findings of this study was the observation that although the average levels of genetic differentiation among Baltic Sea three-spined stickleback populations were low by all standards, numerous genomic regions displayed a high degree of population differentiation. Specifically, by utilizing stringent quality and filtering criteria, we identified numerous SNPs likely to have diverged due to directional selection, often apparently in response to either salinity- and/or temperature-mediated selection. Moreover, we also detected several genomic regions likely to be under balancing selection (i.e. regions showing high diversity and low divergence), including genomic regions harboring genes important for immune function. Interestingly, the patterns of genome-wide genetic variation and differentiation, as well as several candidate genes for local adaptation

**Table 3 Average pairwise  $F_{ST}$  among ten Baltic Sea three-spined stickleback populations based on the subset of 9,404 SNPs**

	COP	FOR	KAR	KAS	KBOR	LET	MAR	NYK	PET	PJM
COP	–	–	–	–	–	–	–	–	–	–
FOR	0.00902	–	–	–	–	–	–	–	–	–
KAR	0.01139	0.01097	–	–	–	–	–	–	–	–
KAS	0.01064	0.00864	0.01011	–	–	–	–	–	–	–
KBOR	0.01060	0.01093	0.01142	0.01089	–	–	–	–	–	–
LET	0.01280	0.01112	0.01389	0.01251	0.01470	–	–	–	–	–
MAR	0.01018	0.01240	0.01291	0.01278	0.01233	0.01491	–	–	–	–
NYK	0.01182	0.01027	0.01347	0.01230	0.01365	0.01102	0.01391	–	–	–
PET	0.01176	0.00938	0.01340	0.01059	0.01279	0.01169	0.01439	0.01194	–	–
PJM	0.01210	0.00992	0.01104	0.01032	0.01307	0.01234	0.01548	0.01142	0.01193	–

detected in this study, were similar to those observed in the marine–freshwater divergence of three-spined sticklebacks on various geographic scales [13,57,60,61,65], despite that the samples used in our study were derived from physically connected marine populations across the Baltic Sea. In the following, we will discuss these findings and relate our results to those of earlier studies of three-spined sticklebacks and other marine organisms. We will also discuss the implications of our findings for our understanding of local adaptation and genetic biodiversity in the Baltic Sea environment.

#### Genome-wide heterogeneous differentiation in Baltic Sea three-spined stickleback

The patterns of genetic diversity and differentiation varied widely across the ten populations in the current study, indicating heterogeneous genomic divergence and divergent selection in Baltic Sea three-spined sticklebacks. These results are fully concurrent with those of an earlier study of Baltic Sea three-spined sticklebacks utilizing a much more limited number of microsatellite markers [47]. Although this kind of heterogeneous divergence has also been demonstrated in other studies of marine fish populations [75], to date very few such studies have employed high-throughput population genomic approaches – nor have they had access to reference genomes – to characterize the genomic architecture of adaptation (but see [28–31,58]). In this regard, we were able to provide a more refined picture of the genome-wide distribution of regions of differentiation in this open system of populations likely to experience multidirectional gene flow. Specifically, outlier loci/regions of divergence were uncovered on every chromosome, indicating that directional selection is acting across the entire genome, rather than being restricted to a few chromosomes, as was the case for the Atlantic cod [28]. Furthermore, many divergent genomic regions showed

associations with salinity and temperature, supporting the interpretation that much of this differentiation could be driven by spatially varying selection pressures. As the use of pooled samples does not allow for investigations of linkage disequilibrium, we cannot ascertain the degree or distance of linkage between these regions of divergence. However, these isolated genomic regions usually span less than 100 kb in length, and the 94 outlier loci we identified were actually located in 38 different 100-kb sliding windows. Hence it is likely that individual loci within these regions are tightly linked, creating islands of divergence (*cf.* [76]) in the midst of the otherwise low neutral baseline divergence. This pattern of divergence hitchhiking is consistent with theoretical [77] and empirical [76,78–82] studies of populations experiencing ongoing gene flow (see also [83]). Moreover, the evidence for heterogeneous genomic divergence at the genome-wide level from this study aligns with the results of our earlier study that found a clear signal of isolation by adaptation, suggesting adaptive divergence has been reducing gene flow at a genome-wide scale [47]. As such, our results support the view (*e.g.*, [28,47,84]) that selection is able to overcome the homogenizing effect of gene flow even in high-gene-flow marine environments.

Several regions of reduced divergence were also uncovered in this study. However, only a few of these regions gave evidence for balancing selection – far less than those of directional selection. In this respect, our results are in stark contrast with those of an early genome scan study of this species, which found evidence for the predominant role of balancing, rather than directional selection, on expressed sequence tag and quantitative trait locus associated microsatellite loci [85,86]. These contrasting results may be explained because highly mutable – and hence also polymorphic – microsatellites may generate spurious signals of balancing selection [85,87], whereas SNPs with lower mutation rates are less likely to generate such biases (*cf.* [88]).



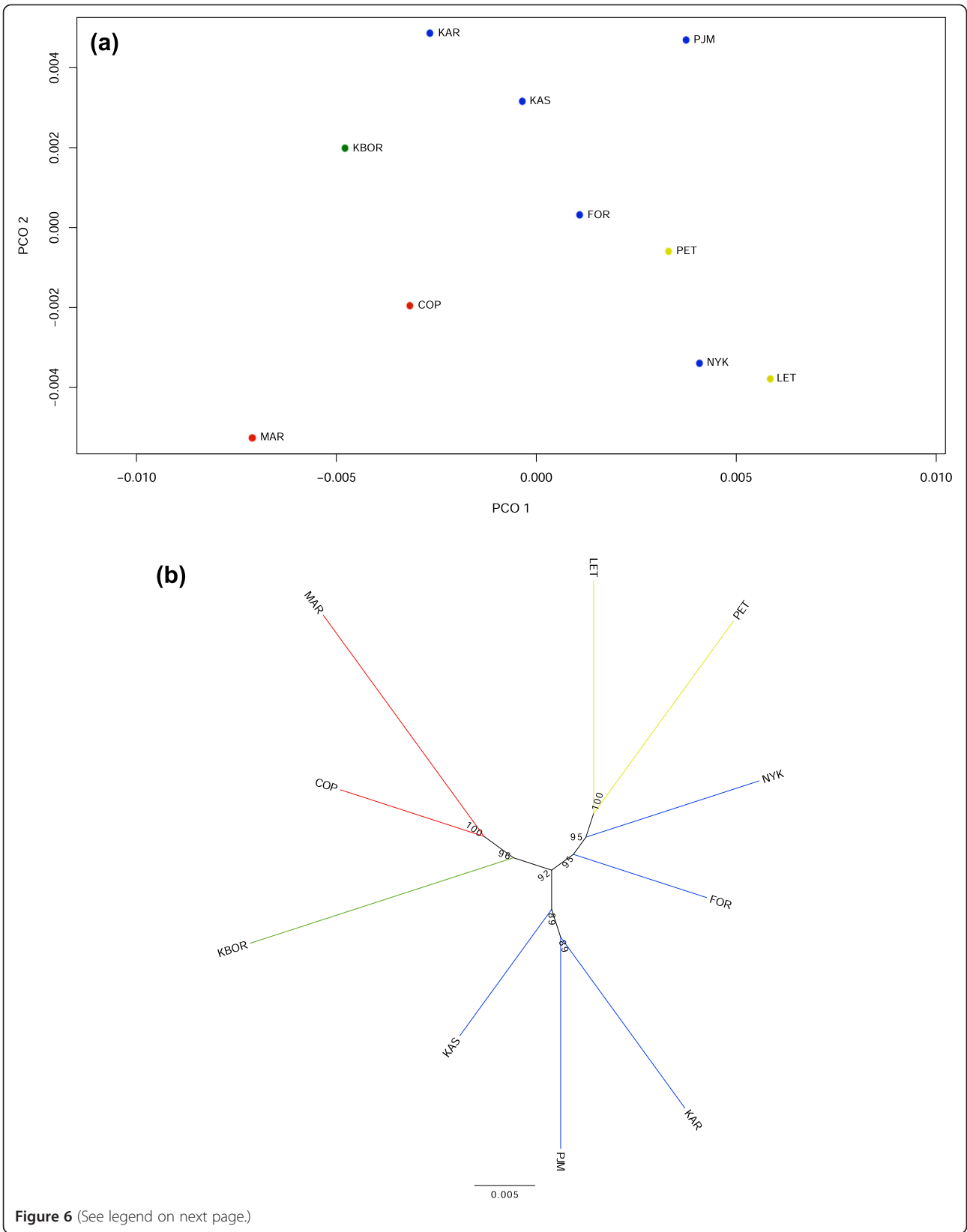


Figure 6 (See legend on next page.)

(See figure on previous page.)

**Figure 6 Genetic relationships among the 10 three-spined stickleback populations from the Baltic Sea.** (a) Principal coordinates analysis (PCoA) plot of the overall average pairwise  $F_{ST}$  values of the subset of 9,404 SNPs among the 10 three-spined stickleback populations. (b) Neighbor-joining tree of the same populations based on  $F_{ST}$  values of each of the subset of 9,404 SNPs. Numbers on tree nodes represent bootstrap values of 1,000 replicates. Populations from different geographic regions are marked in color: red, North Sea; green, southern Baltic Sea; yellow, Gulf of Finland; blue, Gulf of Bothnia and Baltic proper.

### Candidate genes for adaptation

Our results show that three-spined sticklebacks in the Baltic Sea exhibit similar patterns of genetic differentiation and diversity as seen in earlier comparisons of marine–freshwater populations across their global distribution. For example, the distribution of diversity was similar to earlier studies that also reported elevated and lowered levels of genetic diversity at the ends and centers, respectively, of chromosomes I, III, XIII, XVIII and XX [58,60]. Not surprisingly, these regions of increased diversity also exhibited a lowered degree of divergence and a higher incidence of balancing selection compared to other parts of the genome. For instance, as in other stickleback studies [58,60], evidence for balancing selection was detected for the 3' end of chromosome III. This genomic region harbors several candidate genes involved in defense against pathogens (ENSGACG00000017648, ENSGACG00000017778 and ENSGACG00000017779), inflammation pathways (ENSGACG00000017812, ENSGACG00000017834 and ENSGACG00000017927), as well as TRIM genes (ENSGACG00000014250, ENSGACG00000014251, ENSGACG00000014283 and ENSGACG00000014403), which are known targets of balancing selection in primates [89], suggesting the importance of this genomic region for immune responses [60]. There is growing evidence from various vertebrate studies to suggest that genetic diversity in genes related to immunity is elevated and under balancing selection presumably due to their importance for many biological functions, including immunity, mate selection and kin recognition (reviewed in [90,91]). Since pathogens are strong selective agents [92] and the diversity and prevalence of stickleback parasites in the Baltic Sea is known to be high [93,94], the observed footprints of balancing selection on immunity-related genes are understandable.

Several genomic regions of reduced diversity and increased divergence detected in this study are also consistent with those reported in earlier studies of differentiation among marine–freshwater three-spined sticklebacks [49,60,65], as well as sticklebacks in the Baltic Sea [47]. For example, genomic regions on chromosomes I, IV, XI and XXI have repeatedly been identified as divergent between marine and freshwater three-spined sticklebacks in North America [60] and Eurasia [57,65]. Chromosome IV is of particular interest, as the gene related to lateral plate number variation – *Ectodysplasin A*

(*eda*) [46] – is located within a genomic region of increased divergence on this chromosome (Figure 2). Our finding of elevated divergence in the genomic region containing the *eda* locus matches well with results of an earlier study that reported significant differentiation among Baltic Sea three-spined sticklebacks in both the number of lateral plates and the quantitative trait locus tightly linked to *eda* [48]. Thus, the results of our study provide matching evidence for ongoing selection on *eda* in Baltic Sea three-spined sticklebacks, and act as a proof-of-principle demonstration that the uncovered signatures of selection are likely to be real, rather than methodological artifacts or noise. The same is true for the gene *ahr1b* (*2 of 2*) (ENSGACG00000015615; Additional file 1: Table S4), which was identified as a candidate gene both in the current study as well as in our earlier genome scan of Baltic Sea sticklebacks [47]. Further evidence to substantiate this interpretation is provided by comparison to earlier targeted genome scan studies based on microsatellite markers in sticklebacks. These studies provided evidence for directional selection on 17 genomic regions harboring genes physiologically relevant for freshwater adaptation in a global survey of marine–freshwater populations [49], and for nine genomic regions in Baltic Sea populations [47]. The average pairwise  $F_{ST}$  for the genomic regions harboring these markers in this study was 0.033, which was slightly higher than the average  $F_{ST}$  across the whole genome (0.028), demonstrating that the regions harboring genes indicated to be under directional selection in earlier studies also show increased genomic divergence in the current study.

To set our results more firmly in the context of earlier targeted genome scan studies [33,47,49] and genome-wide sequencing studies [13,57,60,61,65], we compiled a list of candidate genes (i.e. genes containing outlier SNPs and/or SNPs associated with environmental variation) detected in our study (Additional file 1: Table S4) and compared these to those found in earlier studies. Of the 297 candidate genes identified here, 15 (5%) were also identified to be involved in marine–freshwater divergence of three-spined sticklebacks in earlier studies (Table 4), for example, genes *cpeb4* (cytoplasmic polyadenylation element binding protein 4, ENSGACG00000018422) and *pparaa* (peroxisome proliferator-activated receptor alpha a, ENSGACG00000018958) on chromosome IV [13,60,61]. An additional 22 were identified in a study investigating differential

**Table 4 Representative list of genes that have been identified as candidates for adaptation**

Ensembl gene ID	Gene name	Gene description	BayeScan	BAYENV		References
			Outlier <sup>a</sup>	Salinity <sup>b</sup>	Temperature <sup>b</sup>	
ENSGACG00000018422	<i>cpeb4</i>	Cytoplasmic polyadenylation element binding protein 4	Yes	**	*	[13,61,65]
ENSGACG00000018320	<i>flt4</i>	fms-related tyrosine kinase 4	–	**	**	[60,61]
ENSGACG00000007263	<i>pde4ba</i>	Phosphodiesterase 4B, cAMP-specific a	–	**	*	[13,60]
ENSGACG00000015515	<i>pde4ca</i>	Phosphodiesterase 4C, cAMP-specific a	–	*	–	[13]
ENSGACG00000001583	<i>pex5</i>	Peroxisomal biogenesis factor 5-Like	–	*	–	[13]
<b>ENSGACG00000018958</b>	<b><i>pparaa</i></b>	<b>Peroxisome proliferator-activated receptor alpha a</b>	<b>Yes</b>	<b>**</b>	<b>*</b>	[13,60,65,96]
ENSGACG00000008634	<i>stats1</i>	Signal transducer and activator of transcription 5.1	–	**	–	[13,60,61,65]
ENSGACG00000014429	<i>zgc:85722</i>	Family with sequence similarity 184, member A	–	*	**	[60,61]
<b>ENSGACG00000009393</b>	<b><i>FAM19A1</i></b>	<b>Family with sequence similarity 19 (chemokine (C-C motif)-like), member A1</b>	–	–	*	[57,95]
ENSGACG00000002723	<i>pi15a</i>	Peptidase inhibitor 15a	–	*	–	[60,65]
ENSGACG00000007629	<i>acer1</i>	Alkaline ceramidase 1	–	**	**	[60]
ENSGACG00000008897	<i>STAC2 (2 of 2)</i>	SH3 and cysteine rich domain 2	–	**	–	[60]
<b>ENSGACG00000019342</b>			–	–	*	[60,95]
ENSGACG00000020327	<i>aifm1</i>	Apoptosis-inducing factor, mitochondrion-associated 1	–	*	*	[60]
ENSGACG00000017985	<i>ctnna2</i>	Catenin (cadherin-associated protein), alpha 2	–	–	*	[65]
<b>ENSGACG00000000827</b>	<b><i>pygmb</i></b>	<b>Phosphorylase, glycogen (muscle) b</b>	–	*	*	[95]
<b>ENSGACG00000001963</b>	<b><i>enpep</i></b>	<b>Glutamyl aminopeptidase</b>	–	–	*	[95]
<b>ENSGACG00000002497</b>	<b><i>fam65a</i></b>	<b>Family with sequence similarity 65, member A</b>	–	–	*	[95]
<b>ENSGACG000000003512</b>	<b><i>crb2b</i></b>	<b>Crumbs homolog 2b</b>	–	*	–	[95]
<b>ENSGACG00000004737</b>	<b><i>rrbp1a</i></b>	<b>Ribosome binding protein 1 homolog a</b>	–	–	*	[95]
<b>ENSGACG000000005034</b>	<b><i>NAV1 (1 of 2)</i></b>		–	–	*	[95]
<b>ENSGACG000000006980</b>	<b><i>prkd3</i></b>	<b>Protein kinase D3</b>	–	*	–	[95]
<b>ENSGACG000000009748</b>	<b><i>swap70b</i></b>	<b>SWAP switching B-cell complex subunit 70b</b>	–	*	–	[95]
<b>ENSGACG00000011184</b>	<b><i>tcf7l1b</i></b>	<b>Transcription factor 7-like 1b</b>	–	*	–	[95]
<b>ENSGACG00000011691</b>			–	*	–	[95]
<b>ENSGACG00000012972</b>	<b><i>gorasp2</i></b>	<b>Golgi reassembly stacking protein 2</b>	–	*	*	[95]
<b>ENSGACG00000013300</b>	<b><i>si:ch211-241e1.3</i></b>	<b>si:ch211-241e1.3</b>	–	**	–	[95]
<b>ENSGACG00000014605</b>			–	*	–	[95]
<b>ENSGACG00000015419</b>	<b><i>cmtm4</i></b>	<b>CKLF-like MARVEL transmembrane domain containing 4</b>	–	–	*	[95]
<b>ENSGACG00000015537</b>	<b><i>cybb</i></b>	<b>Cytochrome b-245, beta polypeptide (chronic granulomatous disease)</b>	–	*	–	[95]
<b>ENSGACG00000015777</b>	<b><i>dmgdh</i></b>	<b>Dimethylglycine dehydrogenase</b>	–	*	*	[95]
<b>ENSGACG00000017100</b>	<b><i>LRP4</i></b>	<b>Low density lipoprotein receptor-related protein 4</b>	–	*	*	[95]
<b>ENSGACG00000017390</b>	<b><i>cntln</i></b>	<b>Centlein, centrosomal protein</b>	–	*	*	[95]
<b>ENSGACG00000019432</b>			<b>Yes</b>	<b>*</b>	<b>*</b>	[95]

**Table 4 Representative list of genes that have been identified as candidates for adaptation (Continued)**

ENSGACG00000019512	<i>npxe3 (7 of 8)</i>	<b>Neurexophilin and PC-esterase domain family, member 3</b>	–	**	*	[95]
ENSGACG00000019730	<i>si:dkeyp-34c12.2</i>	<b>si:dkeyp-34c12.2</b>	–	*	–	[95]
ENSGACG00000020156	<i>si:ch211-55a7.1</i>	<b>si:ch211-55a7.1</b>	–	–	*	[95]

These are for both Baltic Sea and marine–freshwater pairs of three-spined sticklebacks (regular font), as well as those differentially expressed in salinity and control treatments (bold font).

<sup>a</sup>Gene with SNP loci identified as outliers by BayeScan.

<sup>b</sup>Gene with SNP loci in which allelic distribution is significantly associated with variation in salinity and/or temperature by BAYENV. \*\*:  $\log_{10}(\text{BF}) > 5$ ;

\*:  $\log_{10}(\text{BF}) > 1.5$ .

expression of salinity-related genes among freshwater and seawater sticklebacks acclimated to different salinity treatments [95]. Interestingly, allelic variation in most of these candidate genes (e.g., *cpeb4*) was strongly associated with salinity variation, suggesting that environmental salinity has been the selective agent driving genetic differentiation in these loci among Baltic Sea three-spined populations. In addition to the genes listed in Table 4, different paralogs from the same gene families were identified to be under selection both in the Baltic Sea and other three-spined stickleback populations. For example, the gene *slc6a17 (2 of 2)* (solute carrier family 6, member 17; ENSGACG0000007913) was significantly associated with annual salinity variation in Baltic Sea three-spined sticklebacks, whereas its highly similar paralog *slc6a3 (2 of 2)* (solute carrier family 6, member 3; ENSGACG00000018983) has been identified as a candidate for marine–freshwater divergence [60]. This suggests that some of the candidate genes that contribute to repeated adaptation of three-spined sticklebacks to freshwater habitats may also be involved with local adaptation in the environmentally heterogeneous Baltic Sea environment. However, it is worth noting that many candidate genes – and also the general patterns of diversity and divergence – identified in marine–freshwater population pairs are also found among pairs of lake–stream sticklebacks [63,96]. Hence, this may instead indicate that similar constraints imposed by the architecture of the stickleback genome generate similar patterns between our and earlier marine–freshwater studies, rather than similar processes (i.e. salinity-mediated selection). Moreover, it is important to note that in spite of the various lines of evidence that selection is acting on specific genes, empirical demonstration of their functional role is necessary ultimately to validate the inference of selection on candidate variants.

#### Local adaptation to the margin: Baltic Sea

The Baltic Sea is a relatively young postglacial ecosystem, formed 6,500 to 9,800 years ago and characterized by steep environmental gradients in salinity, temperature and community composition [69,97]. Earlier reviews have continually drawn attention to the reduced genetic diversity of Baltic Sea organisms compared to populations in the surrounding seas [69,70]. Given that diversity is a prerequisite for adaptation, it may appear that

populations in the Baltic Sea may face challenges in adapting to the projected environmental changes, e.g., in salinity and temperature. However, the results of this study suggest the contrary. Earlier evidence for adaptive divergence among Baltic Sea sticklebacks as revealed by a limited number of microsatellite markers was here confirmed to be ubiquitous across the genome. It is likely that such adaptation has arisen from the use of standing genetic variation, since the young age of the Baltic Sea has not allowed much time for new mutations to accumulate. Indeed, the importance of standing genetic variation, as well as the general features of the genomic architecture in ancestral marine sticklebacks, have been demonstrated to play important roles in extensive and parallel genome-wide evolution [59]. However, this has mostly been demonstrated in divergent, isolated population pairs. Our results suggest that the same processes can also occur in the face of gene flow, possibly due to the genomic architecture, which may provide a mechanism for the rapid re-assembly and evolution of multi-locus genotypes in newly colonized freshwater habitats [59,98]. Similar evidence for adaptive divergence is also available from other Baltic Sea species, albeit the scale of sampling and/or marker numbers have often been modest (e.g., [29,31,35,36,38,67,68,70,99]). To this end, the results support the view that in spite of its young age and low species diversity (e.g., [69,97]), the genetic biodiversity stemming from local adaptation to the Baltic Sea seascape may be more widespread than is currently anticipated.

#### Methodological considerations

Finally, regarding the methodological considerations, we note that theoretical treatments have shown that sequencing of pooled DNA samples (pool-seq) can be more effective in SNP discovery and can provide more accurate allele frequency estimates than individual sequencing [100]. Nevertheless, pool-seq has its shortcomings: it compromises the ability to conduct certain types of analyses, and certain types of biases and artifacts are possible (e.g., [101–103]). First, information about associations among alleles in different loci is lost, as is the opportunity to estimate linkage disequilibrium. Second, differential amplification of individual samples can create biases in allele frequency estimates [101,103]. Likewise, the assessment of population

divergence will be complicated since the sample size (sequencing coverage) available for allele frequency estimation varies among loci. For example, when using PoPoolation tools [104], the accuracy of allele frequency estimation from pooled samples is highly dependent on the sequencing coverage, although the pipeline implements unbiased estimation by considering pool size and sequencing coverage [100]. However, we believe our inference is robust in respect to these potential caveats on the basis of the following. First, the distribution pattern of sequencing coverage for the SNPs we identified was very similar across populations (Additional file 4: Figure S3), suggesting little heterogeneity in sequencing coverage (and by inference, differential amplification of individual samples) across populations. Second, we found that the genome-wide patterns of population differentiation were stable when sequencing coverage varied (Additional file 5: Figure S4). Thus, at least for sequencing coverage, the results and inferences in this paper should be robust. This inference is further supported because the patterns of genomic variability and differentiation observed in this study align well with those seen in other RAD-seq based analyses (e.g., [40,105]), as well as those seen for microsatellite markers in the same set of populations (see above). If large biases in allele frequency estimates were present, such similarities would be unexpected. Comparison of the pool-seq allele frequency estimates with those generated from 30 SNPs genotyped at the individual level verified this conjecture: the correlation between allele frequency estimates across different loci ranged from  $r = 0.75$  to  $r = 0.95$  depending on the population (Additional file 1: Table S5). Likewise, the correlation between allele frequencies from individual and pooled samples across the 30 loci and all populations was very high ( $r = 0.88$ ,  $P < 2.2 \times 10^{-16}$ ).

## Conclusions

In summary, we discovered that genome-wide patterns of genetic diversity and differentiation among continuously distributed Baltic Sea three-spined stickleback populations – as assessed from polymorphisms in over 30,000 SNP loci – varied widely across the genome. As such, we identified numerous genomic regions exhibiting signatures of divergent – and to some extent also balancing – selection. We also uncovered strong evidence for substantial genetic differentiation associated with both salinity and temperature gradients, suggesting local adaptation in this high-gene-flow environment. The patterns of genome-wide genetic diversity and differentiation in Baltic Sea three-spined sticklebacks were similar to those observed in earlier studies of marine–freshwater divergence in three-spined sticklebacks, suggesting that the same genetic processes and loci may often underlie adaptation both to freshwater and brackish water environments. Hence, apart

from providing strong evidence for genome-wide but heterogeneous genomic divergence driven by local adaptation along an environmental gradient in the post-glacially formed Baltic Sea seascape, our results suggest similarities in genomic signatures of adaptation to freshwater and brackish water environments.

## Methods

The samples used in this study were collected in accordance with the national legislation of the countries concerned. As our samples were derived from wild collected fish, no approval by the Finnish National Animal Experiment Board was required.

## Samples and study sites

Adult three-spined sticklebacks were sampled during the early breeding season of 2009 from ten sites covering most of the Baltic Sea and its opening to the North Sea (Figure 1 and Table 1). The sampling was done with hand seines or minnow traps (mesh size 6 mm). Upon capture, the fish were over-anesthetized with tricaine methanesulfonate (Sigma-Aldrich Co., Saint Louis, United States of America) and stored in 96% ethanol (Altia Oyj, Helsinki, Finland). The study sites and samples are a subset of those used in [47], where more sites were included in analyses with microsatellite markers. The data for average annual salinity and temperature were derived from Table One in [47].

## DNA extraction and restriction site associated DNA sequencing library construction

Whole genomic DNA was extracted from 36 individuals per sampling location, using a NucleoSpin® Tissue kit (Macherey-Nagel, Düren, Germany) following the manufacturer's protocol. DNA was visualized on a 1% agarose gel to assess quality, and quantified with both a NanoDrop® ND-1000 spectrophotometer and Qubit® fluorometer. Each sample was diluted to 10 ng/μl, re-quantified, and pooled according to sampling location. Each pooled sample was then quantified with both the spectrophotometer and fluorometer and equalized to 10 ng/μl.

RAD library preparation was done by following the protocol detailed by Elshire et al. [106]. Briefly, each of the ten pooled samples was digested with 30 U *Pst*I (New England Biolabs® Inc., Frankfurt, Germany) in 20 μL volumes containing 1× NEBuffer 3 (New England Biolabs® Inc., Frankfurt, Germany) and 1× BSA (New England Biolabs® Inc., Frankfurt, Germany). Reactions were first incubated at 37°C for 2 h, then the temperature was increased to 74°C for 15 min and cooled to 4°C for 10 min. The restriction product was then added to 1× ligation buffer, T4-ligase (New England Biolabs® Inc., Frankfurt, Germany) and an adapter mix containing a common adapter and a barcode adapter unique to each sample. Barcode adapters were selected from the list of 96

sequences provided by Elshire et al. [106]. Then 50  $\mu$ L ligation reactions were first incubated at 22°C for 1 h, followed by 30 min at 65°C and cooled to 4°C for 10 min. Following purification with a Qiagen Qiaquick kit (QIAGEN, Stockach, Germany), 10  $\mu$ L of ligation from each population were pooled for library amplification. The library amplification reaction used 10  $\mu$ L of the pooled ligation product, Phire enzyme, 1 $\times$  reaction buffer, 1.5 mM MgCl<sub>2</sub>, 10 mM dNTP (New England Biolabs® Inc., Frankfurt, Germany), and 0.5  $\mu$ M primer mix (see [106] for primer sequences) in 50  $\mu$ L volumes. PCR was initiated at 72°C for 1 min, then raised to 95°C for an additional 30 s, followed by 18 cycles of 95°C for 30 s, 65°C for 30 s and 72°C for 20 s. A final extension at 72°C for 5 min concluded the reaction. Products were visualized on 2.5% MetaPhor low-melt agarose gel, and fragments of 250 to 350 bp were excised after running for 2 h at 80 V and cleaned with QIAquick® Gel Extraction Kit (QIAGEN, Stockach, Germany) according to the manufacturer's protocol.

#### Sequencing, data processing, and alignment

Barcoded RAD samples were sequenced on one lane of the Illumina HiSeq2000 platform with a 100-bp single-end strategy by BGI Hongkong Co, Limited. Using the FastX toolkit [107], all raw reads were end-trimmed to a length of 90 bp, and reads containing one or more bases with a Phred quality score below 10 or more than 5% of the positions below 20 were discarded.

Quality filtered reads from each sample were aligned to the three-spined stickleback genome (release-73, Ensembl) separately using BWA 0.7.5a [108]. The maximum edit distance was two and the maximum number of alignments for each read was one. The mapping results in SAM format were converted into BAM format using SAMtools 0.1.18 [109] and filtered for a minimum mapping quality of 20. BAM files were then converted into mpileup format using SAMtools 0.1.18 with a maximum of 1,000 reads at a given position per BAM file.

#### Estimation of genome-wide genetic variation and differentiation

To characterize genome-wide patterns of genetic variation and population differentiation, nucleotide diversity (Tajima's  $\pi$ ), population mutation rate (Watterson's theta,  $\theta_W$ ) and Tajima's  $D$  were estimated using PoPoolation 1.2.2 [104]. In addition, the fixation index ( $F_{ST}$ ) values for each pairwise comparison were estimated using PoPoolation2 [110], by implementing a number of stringent criteria to define genomic sites for analysis across the entire genome. Since the accuracy of allele frequency estimation in the sequencing of pooled individuals is highly dependent on sequence coverage, we used high sequence coverage and large sliding windows (see below), as they are expected to increase the accuracy of the above-

mentioned population genetic parameters by decreasing stochastic error [104]. To estimate  $\pi$  and  $\theta_W$ , all genomic sites subjected to analysis were required to have a minimum minor allele count of 2 and coverage between 10 and 500 for each population, as well as a minimum minor allele count of 4 and coverage between 20 and 1,000 across all the 10 populations. Since Tajima's  $D$  is sensitive to variation in coverage [104], it was only calculated for genomic sites with a coverage of 36 for each population and for alleles with a coverage of 72 across all the 10 populations.  $F_{ST}$  values for each pairwise comparison were estimated for genomic sites with a minimum minor allele count of 4 across all the 10 populations and coverage between 10 and 500 within each population. To make this study comparable to other population genomic studies of marine and/or freshwater three-spined sticklebacks [57,58,60,65], a non-overlapping 100-kb sliding window was used for estimating the above-mentioned population genetic parameters across the entire genome with a minimum base Phred quality of 20 for the analyzed genomic sites. Patterns of genomic variation as reflected in  $F_{ST}$ , Tajima's  $\pi$ ,  $\theta_W$  and Tajima's  $D$  were visualized using Circos [111].

#### Detection of selection footprints

To identify genes likely to be differentiated as a result of selection, a subset of SNPs were identified using PoPoolation2 with stringent criteria: a minimum minor allele count of 6, and coverage between 36 and 500 across all the 10 populations. Two independent methods were employed to identify selection. First, pairwise  $F_{ST}$  values for each of the subsets of SNPs were calculated between populations using PoPoolation2. SNPs falling into the upper 0.5% tails of at least 5 of the 45 pairwise  $F_{ST}$  comparisons were identified as potentially differentiated loci, following an empirical outlier detection approach [58,112]. Second, to verify whether this empirical approach is reliable, a simulated multi-locus dataset of the subset of SNPs was exported from PoPoolation2, and BayeScan 2.1 [113] was used for estimating the posterior probability that a given locus is affected by selection. Briefly, prior odds of 100 were used for identifying the top candidates of the selected loci and a total of 55,000 reversible-jump Markov chain Monte Carlo chains were run with a thinning interval of 10, following 20 pilot runs of 5,000 iterations each, and a burn-in length of 50,000. Loci were considered under selection with a FDR of 0.05. Only SNPs that were identified as outliers by both of the two above-mentioned approaches were considered as truly differentiated loci.

#### Detection of genetic differentiation associated with environmental parameters

To test for association between genetic differentiation and environmental parameters, a Bayesian approach as

implemented in BAYENV [73] was applied to the subset of SNPs identified by PoPoolation2. The Bayesian approach takes into account the effect of population structure, using a covariance matrix based on neutral markers to control for demographic effects when testing for correlations between environmental and genetic differentiation [73]. To do so, a neutral covariance matrix based on the neutral SNPs (as revealed by outlier tests; see below) was first estimated, and then two environmental parameters (viz. average annual salinity and average annual temperature; Table 1) were tested for association with genetic variation. Each environmental parameter was standardized by subtracting the mean and dividing by the standard deviation of the parameter across all sites. To verify that the results were not sensitive to stochastic errors, three independent runs with different random seeds were run.

#### SNP annotation and gene ontology analysis

The three-spined stickleback genome annotations were downloaded from Ensembl (release-73). BEDTools 2.17.0 [114] was used for annotation of the subset of SNPs identified by PoPoolation2 to characterize whether the SNPs were located within a gene. GO terms for the three-spined stickleback genes were retrieved with BioMart [115] from Ensembl. A GO enrichment analysis was conducted to test if certain gene classes were over- or underrepresented among genes harboring outlier loci compared to the genes harboring the remaining neutral SNPs, using GOSSIP [116].

#### Characterization of population structure

The population structure was characterized on the basis of the pairwise  $F_{ST}$  matrices among populations estimated using the subset of SNPs identified by PoPoolation2. To visualize the multi-locus patterns of population differentiation, a PCO plot was generated using the R package labdsv [117] based on average  $F_{ST}$  values. To examine further the patterns of population differentiation, a neighbor-joining tree based on  $F_{ST}$  values [118] was constructed using the simulated multi-locus dataset of the subset of SNPs identified by PoPoolation2 with 1,000 bootstrap replicates in Populations 1.2.32 [119] software.

To compare the patterns of genetic variability and differentiation in SNP markers with those in microsatellite markers, we retrieved data from 40 microsatellite loci genotyped for these same populations [47]. A simple correlation analysis was used to compare genetic variability (average heterozygosity) across populations, whereas the patterns of population differentiation (as reflected in pairwise  $F_{ST}$  estimates) were compared with a Mantel's test. Tests for isolation by distance were conducted with a Mantel's test using linearized  $F_{ST}$  values [ $F_{ST} / (1 - F_{ST})$ ] and log-transformed geographic distances separating sampling locations.

#### Allele frequency validation

To validate estimates of allelic frequency from the pool-seq data, we genotyped a subset of 30 SNPs from each of the individual fish used for the pooled DNA analyses using the iPLEX Gold® assay on the MassARRAY® platform (Sequenom) system. This genotyping was performed by the Technology Centre of the Institute for Molecular Medicine Finland at the University of Helsinki. Allele frequencies from this data were estimated with a custom Perl script and compared to estimates from pooled data as obtained using the procedures above.

#### Data accessibility statement

Sequences underlying this study have been deposited in NCBI's Sequence Read Archive and accession numbers are SRR1596320, SRR1596321, SRR1596322, SRR1596323, SRR1596324, SRR1596325, SRR1596326, SRR1596327, SRR1596328, and SRR1596329.

#### Additional files

**Additional file 1: Table S1.** Chromosome length (bp), number of *Pst*I restriction sites and RAD-seq read mapped on each chromosome in each population. **Table S2.** Number of SNPs identified on each chromosome in each population with PoPoolation. **Table S3.** Nucleotide diversity ( $\pi / \theta_{w}$ ) of each chromosome in each population based on polymorphic loci shown in Table S2. **Table S4.** Candidate genes associated with adaptation. Outlier: Gene included outlier loci. Salinity: Gene included SNPs for which allele frequency was associated with salinity variation of sampling site. Temperature: Gene included SNPs for which allele frequency was associated with annual temperature variation of sampling site. \*: Gene included SNPs had  $\log_{10}$  (Bayes factor) > 5. **Table S5.** Comparison of allele frequencies between individual pool-seq samples in ten different populations at 30 SNP loci.  $r$  is the correlation in allele frequencies across the loci in a given population.

**Additional file 2: Figure S1.** Correlations between number of *Pst*I restriction sites (a), mapped reads (b), and number of SNPs on each chromosome (c) against chromosome length in the population COP. Mbp, megabase pair.

**Additional file 3: Figure S2.** Genome-wide distribution of genetic variation in each of the ten study populations. Chromosomes are labeled in black Roman numerals and represented as grey and black blocks.

**Additional file 4: Figure S3.** Sequencing coverage for SNPs identified in each of the study populations.

**Additional file 5: Figure S4.** Genome-wide distribution of genetic differentiation across all of the 10 three-spined populations in the Baltic Sea with variable sequencing coverage. Red line: sequencing coverage between 10 and 500; green line: sequencing coverage between 10 and 360; blue line: sequencing coverage between 36 and 360.

#### Abbreviations

BF: Bayes factor; bp: base pair; BSA: bovine serum albumin; FDR: false discovery rate; GO: gene ontology; kb: kilobase; PCoA: principal coordinates analysis; PCR: polymerase chain reaction; RAD-seq: restriction site associated DNA sequencing; SNP: single nucleotide polymorphism; SE: standard error.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

BG, JD and JM conceived the project. JD and GS performed the laboratory experiments. BG, AN and JM analyzed the data. BG, JD and JM wrote the paper. All authors read and approved the final manuscript.

### Acknowledgements

We would like to thank Anders Adill, Janis Birzaks, Michael Hansen, Markku Kaukoranta, Lotta Kvarnemo, Tuomas Leinonen, Hannu Mäkinen and Jouko Pokela for help with collecting the samples, and Kirsi Kähkönen, Erica Leder, Hannu Mäkinen and Heidi Viitaniemi for help and advice with the laboratory work. We also thank Michael Hansen and an anonymous reviewer for thoughtful and constructive comments on an earlier version of this manuscript. We thank Heini Natri for preparing Figure 1, and the Technology Centre of the Institute for Molecular Medicine Finland at the University of Helsinki for SNP genotyping. The Finnish IT Center for Science Ltd, administered by the Finnish Ministry of Education and Culture, provided computing resource support. Our research was supported by the Academy of Finland (grants 118673, 134728, 213491 and 129662 to JM), the University of Helsinki (JM) and LUOVA Graduate School (JD). The research leading to these results has also received funding from the European Community's Seventh Framework Program (FP/2007-2013) under grant agreement no 217246 made with BONUS, the joint Baltic Sea research and development program (JM).

### Author details

<sup>1</sup>Department of Biosciences, Ecological Genetics Research Unit, University of Helsinki, PO Box 65, Helsinki FI-00014, Finland. <sup>2</sup>Current address: CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO - Laboratório Associado, Universidade do Porto, Vairão 4485-661, Portugal. <sup>3</sup>Current address: Department of Biosciences, Metapopulation Research Group, University of Helsinki, PO Box 65, Helsinki FI-00014, Finland.

Received: 10 November 2014 Accepted: 12 March 2015

Published online: 24 March 2015

### References

- Blanquart F, Gandon S, Nuismer SL. The effects of migration and drift on local adaptation to a heterogeneous environment. *J Evol Biol.* 2012;25:1351–63.
- Blanquart F, Kaltz O, Nuismer SL, Gandon S. A practical guide to measuring local adaptation. *Ecol Lett.* 2013;16:1195–205.
- Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. *Nat Rev Genet.* 2013;14:807–20.
- Hereford J. A quantitative survey of local adaptation and fitness trade-offs. *Am Nat.* 2009;173:579–88.
- Kawecki TJ, Ebert D. Conceptual issues in local adaptation. *Ecol Lett.* 2004;7:1225–41.
- Savolainen O, Pyhäjärvi T, Knurr T. Gene flow and local adaptation in trees. *Annu Rev Ecol Evol S.* 2007;38:595–619.
- Kremer A, Le Corre V. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity.* 2012;108:375–85.
- Leinonen T, McCairns RJS, O'Hara RB, Merilä J.  $Q_{ST}$ - $F_{ST}$  comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat Rev Genet.* 2013;14:179–90.
- Leinonen T, O'Hara RB, Cano JM, Merilä J. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *J Evol Biol.* 2008;21:1–17.
- McKay JK, Latta RG. Adaptive population divergence: markers, QTL and traits. *Trends Ecol Evol.* 2002;17:285–91.
- Fan SH, Elmer KR, Meyer A. Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and neotropical lineages. *Philos T R Soc B.* 2012;367:385–94.
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, et al. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science.* 2011;334:83–6.
- Jones FC, Grabherr MG, Chan YF, Russell P, Maucci E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature.* 2012;484:55–61.
- Pool JE, Hellmann I, Jensen JD, Nielsen R. Population genetic inference from genomic sequence variation. *Genome Res.* 2010;20:291–300.
- De Mita S, Thuillet AC, Gay L, Ahmadi N, Manel S, Ronfort J, et al. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol.* 2013;22:1383–99.
- Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of  $F_{ST}$  outlier tests. *Mol Ecol.* 2014;23:2178–92.
- Storz JF. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol.* 2005;14:671–88.
- Cano JM, Shikano T, Kuparinen A, Merilä J. Genetic differentiation, effective population size and gene flow in marine fishes: implications for stock management. *J Integr Field Biol.* 2008;5:1–10.
- DeWoody JA, Avise JC. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J Fish Biol.* 2000;56:461–73.
- Grosberg R, Cunningham CW. Genetic structure in the sea: from populations to communities. In: Bertness MD, Gaines SD, Hay ME, editors. *Marine Community Ecology.* Sunderland, Massachusetts: Sinauer Associates; 2001. p. 61–84.
- Gyllenstein U. The genetic-structure of fish – differences in the intraspecific distribution of biochemical genetic-variation between marine, anadromous, and fresh-water species. *J Fish Biol.* 1985;26:691–9.
- Ward RD, Woodward M, Skibinski DOF. A comparison of genetic diversity levels in marine, fresh-water, and anadromous fishes. *J Fish Biol.* 1994;44:213–32.
- Conover DO. Local adaptation in marine fishes: evidence and implications for stock enhancement. *B Mar Sci.* 1998;62:477–93.
- Conover DO, Clarke LM, Munch SB, Wagner GN. Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *J Fish Biol.* 2006;69:21–47.
- DeFaveri J, Merilä J. Local adaptation to salinity in the three-spined stickleback? *J Evol Biol.* 2014;27:290–302.
- Hice LA, Duffy TA, Munch SB, Conover DO. Spatial scale and divergent patterns of variation in adapted traits in the ocean. *Ecol Lett.* 2012;15:568–75.
- Marcil J, Swain DP, Hutchings JA. Countergradient variation in body shape between two populations of Atlantic cod (*Gadus morhua*). *P Roy Soc B-Biol Sci.* 2006;273:217–23.
- Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, et al. Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *P Roy Soc B-Biol Sci.* 2010;277:3725–34.
- Corander J, Majander KK, Cheng L, Merilä J. High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Mol Ecol.* 2013;22:2931–40.
- Hale MC, Thrower FP, Berntson EA, Miller MR, Nichols KM. Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. *G3-Genes Genom Genet.* 2013;3:1273–85.
- Lamichhane S, Barrio AM, Rafati N, Sundstrom G, Rubin CJ, Gilbert ER, et al. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *P Natl Acad Sci USA.* 2012;109:19345–50.
- Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Munch K, Jonsson B, et al. Genome-wide single-generation signatures of local selection in the panmictic European eel. *Mol Ecol.* 2014;23:2514–28.
- DeFaveri J, Shikano T, Shimada Y, Merilä J. High degree of genetic differentiation in marine three-spined sticklebacks (*Gasterosteus aculeatus*). *Mol Ecol.* 2013;22:4811–28.
- Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V. Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus* L.). *Heredity.* 2007;99:592–600.
- Limborg MT, Helyar SJ, de Bruyn M, Taylor MI, Nielsen EE, Ogdren R, et al. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Mol Ecol.* 2012;21:3686–703.
- Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, et al. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evol Biol.* 2009;9:276.
- Shikano T, Ramadevi J, Merilä J. Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Mol Biol Evol.* 2010;27:2775–89.
- Teacher AGF, Andre C, Jonsson PR, Merilä J. Oceanographic connectivity and environmental correlates of genetic structuring in Atlantic herring in the Baltic Sea. *Evol Appl.* 2013;6:549–67.
- Jackson AM, Semmens BX. Sadvoy de Mitcheson Y, Nemeth RS, Heppell SA, Bush PG, et al. Population structure and phylogeography in Nassau grouper (*Epinephelus striatus*), a mass-aggregating marine fish. *PLoS One.* 2014;9:e97508.
- Karlsen BO, Klingan K, Emblem A, Jorgensen TE, Jueterbock A, Furmanek T, et al. Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. *Mol Ecol.* 2013;22:5098–111.
- Kruck NC, Innes DI, Ovenden JR. New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). *Mol Ecol Resour.* 2013;13:715–25.



42. Liggins L, Trembl EA, Riginos C. Taking the plunge: an introduction to undertaking seascape genetic studies and using biophysical models. *Geogr Compass*. 2013;7:173–96.
43. Selkoe KA, Henzler CM, Gaines SD. Seascape genetics and the spatial ecology of marine populations. *Fish Fish*. 2008;9:363–77.
44. Albert AY, Sawaya S, Vines TH, Knecht AK, Miller CT, Summers BR, et al. The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution*. 2008;62:76–85.
45. Barrett RD, Paccard A, Healy TM, Bergek S, Schulte PM, Schluter D, et al. Rapid evolution of cold tolerance in stickleback. *P Roy Soc B-Biol Sci*. 2011;278:233–8.
46. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*. 2005;307:1928–33.
47. DeFaveri J, Jonsson PR, Merilä J. Heterogeneous genomic differentiation in marine threespine sticklebacks: adaptation along an environmental gradient. *Evolution*. 2013;67:2530–46.
48. DeFaveri J, Merilä J. Evidence for adaptive phenotypic differentiation in Baltic Sea sticklebacks. *J Evol Biol*. 2013;26:1700–15.
49. DeFaveri J, Shikano T, Shimada Y, Goto A, Merilä J. Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*). *Evolution*. 2011;65:1800–7.
50. Marchinko KB, Schluter D. Parallel evolution by correlated response: lateral plate reduction in threespine stickleback. *Evolution*. 2007;61:1084–90.
51. McCairns RJS, Bernatchez L. Landscape genetic analyses reveal cryptic population structure and putative selection gradients in a large-scale estuarine environment. *Mol Ecol*. 2008;17:3901–16.
52. McCairns RJS, Bernatchez L. Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution*. 2010;64:1029–47.
53. McCairns RJS, Bourget S, Bernatchez L. Putative causes and consequences of MHC variation within and between locally adapted stickleback demes. *Mol Ecol*. 2011;20:486–502.
54. Raeymaekers JA, Van Houdt JK, Larmuseau MH, Geldof S, Volckaert FA. Divergent selection as revealed by  $P_{ST}$  and QTL-based  $F_{ST}$  in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. *Mol Ecol*. 2007;16:891–905.
55. Hendry AP, Peichel CL, Matthews B, Boughman JW, Nosil P. Stickleback research: the now and the next. *Evol Ecol Res*. 2013;15:111–41.
56. Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, et al. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Mol Ecol*. 2013;22:2864–83.
57. Ferchaud AL, Pedersen SH, Bekkevold D, Jian J, Niu Y, Hansen MM, et al. A low-density SNP array for analyzing differential selection in freshwater and marine populations of threespine stickleback (*Gasterosteus aculeatus*). *BMC Genomics*. 2014;15:867.
58. Feulner PGD, Chain FJJ, Panchal M, Eizaguirre C, Kalbe M, Lenz TL, et al. Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Mol Ecol*. 2013;22:635–49.
59. Hohenlohe PA, Bassham S, Currey M, Cresko WA. Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos T R Soc B*. 2012;367:395–408.
60. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA, et al. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 2010;6:e1000862.
61. Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, et al. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr Biol*. 2012;22:83–90.
62. Roesti M, Gavrilets S, Hendry AP, Salzburger W, Berner D. The genomic signature of parallel adaptation from shared genetic variation. *Mol Ecol*. 2014;23:3944–56.
63. Roesti M, Hendry AP, Salzburger W, Berner D. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol Ecol*. 2012;21:2852–62.
64. Roesti M, Moser D, Berner D. Recombination in the threespine stickleback genome – patterns and consequences. *Mol Ecol*. 2013;22:3014–27.
65. Terekhanova NV, Logacheva MD, Penin AA, Neretina TV, Barmintseva AE, Bazhkin GA, et al. Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genet*. 2014;10:e1004696.
66. DeFaveri J, Shikano T, Ab Ghani NI, Merilä J. Contrasting population structures in two sympatric fishes in the Baltic Sea basin. *Mar Biol*. 2012;159:1659–72.
67. Hämmerli A, Reusch TB. Local adaptation and transplant dominance in genets of the marine clonal plant *Zostera marina*. *Mar Ecol Prog Ser*. 2002;242:111–8.
68. Wrangé AL, Andre C, Lundh T, Lind U, Blomberg A, Jonsson PJ, et al. Importance of plasticity and local adaptation for coping with changing salinity in coastal areas: a test case with barnacles in the Baltic Sea. *BMC Evol Biol*. 2014;14:156.
69. Johannesson K, Smolarz K, Grahn M, Andre C. The future of Baltic Sea populations: local extinction or evolutionary rescue? *Ambio*. 2011;40:179–90.
70. Johannesson K, Andre C. Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Mol Ecol*. 2006;15:2013–29.
71. Wennerstrom L, Laikre L, Ryman N, Utter FM, Ab Ghani NI, Andre C, et al. Genetic biodiversity in the Baltic Sea: species-specific patterns challenge management. *Biodivers Conserv*. 2013;22:3045–65.
72. Davey JW, Blaxter ML. RADSeq: next-generation population genetics. *Brief Funct Genomics*. 2010;9:416–23.
73. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. *Genetics*. 2010;185:1411–23.
74. Jeffreys H. *Theory of Probability*. 3rd ed. London: Oxford University Press; 1961.
75. Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D. Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol Ecol*. 2009;18:3128–50.
76. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. *Mol Ecol*. 2009;18:375–402.
77. Yeaman S, Whitlock MC. The genetic architecture of adaptation under migration-selection balance. *Evolution*. 2011;65:1897–911.
78. Flaxman SM, Feder JL, Nosil P. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*. 2013;67:2577–91.
79. Jacobsen MW, Pujolar JM, Bernatchez L, Munch K, Jian J, Niu Y, et al. Genomic footprints of speciation in Atlantic eels (*Anguilla anguilla* and *A. rostrata*). *Mol Ecol*. 2014;23:4785–98.
80. Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL, et al. Widespread genomic divergence during sympatric speciation. *P Natl Acad Sci USA*. 2010;107:9724–9.
81. Via S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos T R Soc B*. 2012;367:451–60.
82. Feder JL, Egan SP, Nosil P. The genomics of speciation-with-gene-flow. *Trends Genet*. 2012;28:342–50.
83. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, et al. Genomics and the origin of species. *Nat Rev Genet*. 2014;15:176–92.
84. Hauser L, Carvalho GR. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish Fish*. 2008;9:333–62.
85. Beaumont MA. Selection and sticklebacks. *Mol Ecol*. 2008;17:3425–7.
86. Mäkinen HS, Cano JM, Merilä J. Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Mol Ecol*. 2008;17:3565–82.
87. Hedrick PW. A standardized genetic differentiation measure. *Evolution*. 2005;59:1633–8.
88. Nichols RA, Freeman KL. Using molecular markers with high mutation rates to obtain estimates of relative population size and to distinguish the effects of gene flow and mutation: a demonstration using data from endemic Mauritian skinks. *Mol Ecol*. 2004;13:775–87.
89. Newman RM, Hall L, Connole M, Chen GL, Sato S, Yuste E, et al. Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5a. *P Natl Acad Sci USA*. 2006;103:19134–9.
90. Kamiya T, O'Dwyer K, Westerdahl H, Senior A, Nakagawa S. A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. *Mol Ecol*. 2014;23:5151–63.
91. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool*. 2005;2:16.
92. Altizer S, Harvell D, Friedle E. Rapid evolutionary dynamics and disease threats to biodiversity. *Trends Ecol Evol*. 2003;18:589–96.
93. Budria A, Candolin U. Human-induced eutrophication maintains high parasite prevalence in breeding threespine stickleback populations. *Parasitology*. 2014. doi:10.1017/S0031182014001814.
94. Zander CD. Parasite diversity of sticklebacks from the Baltic Sea. *Parasitol Res*. 2007;100:287–97.

95. Wang G, Yang E, Smith KJ, Zeng Y, Ji G, Connon R, et al. Gene expression responses of threespine stickleback to salinity: implications for salt-sensitive hypertension. *Front Genet.* 2014;5:312.
96. Deagle BE, Jones FC, Chan YGF, Absher DM, Kingsley DM, Reimchen TE, et al. Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *P Roy Soc B-Biol Sci.* 2012;279:1277–86.
97. Ojaveer H, Jaanus A, Mackenzie BR, Martin G, Olenin S, Radziejewska T, et al. Status of biodiversity in the Baltic Sea. *PLoS One.* 2010;5:e12467.
98. Schluter D, Conte GL. Genetics and ecological speciation. *P Natl Acad Sci USA.* 2009;106:9955–62.
99. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, et al. The genome sequence of Atlantic cod reveals a unique immune system. *Nature.* 2011;477:207–10.
100. Futschik A, Schlötterer C. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics.* 2010;186:207–18.
101. Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol.* 2013;22:3179–90.
102. Lynch M, Bost D, Wilson S, Maruki T, Harrison S. Population-genetic inference from pooled-sequencing data. *Genome Biol Evol.* 2014;6:1210–8.
103. Anderson EC, Skaug HJ, Barshis DJ. Next-generation sequencing for molecular ecology: a caveat regarding pooled samples. *Mol Ecol.* 2014;23:502–12.
104. Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, et al. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One.* 2011;6:e15925.
105. Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, et al. Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Mol Ecol.* 2012;21:4748–69.
106. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 2011;6, e19379.
107. FASTX-Toolkit. [http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html).
108. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
109. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
110. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics.* 2011;27:3435–6.
111. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19:1639–45.
112. Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, et al. Tracking footprints of artificial selection in the dog genome. *P Natl Acad Sci USA.* 2010;107:1160–5.
113. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics.* 2008;180:977–93.
114. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
115. Kasprzyk A. BioMart: driving a paradigm change in biological data management. *Database.* 2011;2011:bar049.
116. Bluthgen N, Brand K, Cajavec B, Swat M, Herzel H, Beule D, et al. Biological profiling of gene groups utilizing Gene Ontology. *Genome Inform.* 2005;16:106–15.
117. LabDSV. <http://ecology.msu.montana.edu/labdsv/R/>.
118. Latter BD. Selection in finite populations with multiple alleles. 3. Genetic divergence with centripetal selection and mutation. *Genetics.* 1972;70:475–90.
119. Populations 1.2.32. <http://bioinformatics.org/~tryphon/populations/>.
120. Ye J, Fang L, Zheng HK, Zhang Y, Chen J, Zhang ZJ, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 2006;34:W293–7.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

