Genome-Wide Linkage Disequilibrium in Nine-Spined Stickleback Populations

Ji Yang,* Takahito Shikano,† Meng-Hua Li,*,1 and Juha Merilä†

*CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences (CAS), Beijing 100101, China, and †Ecological Genetics Research Unit, Department of Biosciences, FIN-00014 University of Helsinki, Finland

ABSTRACT Variation in the extent and magnitude of genome-wide linkage disequilibrium (LD) among populations residing in different habitats has seldom been studied in wild vertebrates. We used a total of 109 microsatellite markers to quantify the level and patterns of genome-wide LD in 13 Fennoscandian nine-spined stickleback (*Pungitius pungitius*) populations from four (*viz.* marine, lake, pond, and river) different habitat types. In general, high magnitude ($D' > 0.5$) of LD was found both in freshwater and marine populations, and the magnitude of LD was significantly greater in inland freshwater than in marine populations. Interestingly, three coastal freshwater populations located in close geographic proximity to the marine populations exhibited similar LD patterns and genetic diversity as their marine neighbors. The greater levels of LD in inland freshwater compared with marine and costal freshwater populations can be explained in terms of their contrasting demographic histories: founder events, long-term isolation, small effective sizes, and population bottlenecks are factors likely to have contributed to the high levels of LD in the inland freshwater populations. In general, these findings shed new light on the patterns and extent of variation in genome-wide LD, as well as the ecological and evolutionary factors driving them.

During the processes of population differentiation and local adaptation, evolutionary forces of selection, drift, gene flow, and mutation jointly influence the structure and patterning of genetic variation in the genome. Ultimately, this influences the extent and strength of associations among different parts of the genome. Such genetic associations are reflected in nonrandom coinheritance of alleles at different loci, a phenomenon known as linkage disequilibrium (LD; Lewontin and Kojima 1960). Interest toward LD recently has been fueled by its fundamental role in determining the required marker density and feasibility of gene mapping approaches (Jorde 2000; Zondervan and Cardon 2004). Knowledge about the extent and magnitude of LD also has the potential to provide valuable insights into an organism's evolutionary past (Nordborg and Tavaré 2002; Slatkin 2008). For

instance, the degree and extent of genome-wide LD can help to identify population substructuring and demographic events such as bottlenecks and admixture (*e.g.*, Nei and Li 1973; Golding and Strobeck 1980). Similarly, patterns of local LD can help to uncover the history of mutation, gene conversion, and selection (*e.g.*, Karlin and Feldman 1970; Frisse *et al.* 2001). In this perspective, studies of LD also can be viewed as bridging evolutionary biology to genomics.

During the past few years, molecular markers across the whole genome have become available in many species, facilitating progress in quantifying the magnitude and patterns of genome-wide LD, for example in human (*e.g.*, Reich *et al.* 2001; Shifman *et al.* 2003), livestock (*e.g.*, Corbin *et al.* 2010; Badke *et al.* 2012; García-Gámez *et al.* 2012; Espigolan *et al.* 2013), crop (*e.g.*, Hao *et al.* 2011; Van Inghelandt *et al.* 2011; Delourme *et al.* 2013; Fang *et al.* 2013), and model species (*e.g.*, Mukai *et al.* 1971; Nordborg *et al.* 2002; Branca *et al.* 2011). However, the information about genome-wide LD in wild vertebrate populations remains limited to a few studies of mammals (*e.g.*, Hernandez *et al.* 2007; Laurie *et al.* 2007), birds (*e.g.*, Backström *et al.* 2006; Li and Merilä 2010; Kawakami *et al.* 2014), and fishes (*e.g.*, Hohenlohe *et al.* 2012). Yet, studies of LD in the wild are important, because they can address biological questions that are not approachable by use of laboratory or domestic populations. These include, for instance, mapping quantitative trait loci (QTL) or candidate genes for ecologically and environmentally important traits in the wild

[1]Corresponding author: CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences (CAS), Beichen West Road No. 1-5, Beijing 100101, China. E-mail: menghua.li@ioz.ac.cn

(e.g., Slate 2005, 2013; Laurie et al. 2007; Ellegren and Sheldon 2008; Gratten et al. 2008; Slate et al. 2009), and disclosing the relative contributions of different factors like natural selection and demography shaping organism's genome (e.g., Cutter 2006). Furthermore, knowledge about interpopulation and interhabitat variation in genomic LD can be helpful in advancing our understanding of evolutionary processes in nature (Gould and Johnston 1972; Roesti et al. 2013). Several earlier studies have described differences in the degree and extent of LD among populations of humans (e.g., Service et al. 2006), domestic animals (e.g., Sutter et al. 2004; Badke et al. 2012), and cultivated plants (Hao et al. 2011; Fang et al. 2013). However, interpopulation comparisons of LD in wild vertebrates are scarce (but see: Li and Merilä 2011; Miller et al. 2011; Hohenlohe et al. 2012). Hence, more empirical studies are needed to advance our understanding of variation in the extent and magnitude of LD in the wild.

The nine-spined stickleback (Pungitius pungitius) is a small cold-water adapted fish with a circumpolar distribution in the northern hemisphere (Wootton 1976). Fennoscandian nine-spined stickleback populations have been derived from a common ancestral population and became established after the last glacial maximum (Shikano et al. 2010a; Teacher et al. 2011). They occur in both freshwater and marine habitats along the coastal areas of the White Sea and the Baltic Sea (Shikano et al. 2010a; Defaveri et al. 2012). Due to differing selection pressures among habitats, the species has undergone marked adaptive differentiation and, thus, shows pronounced morphological, physiological, and behavioral differentiation across habitat types (Merilä 2013). For instance, freshwater populations display reduced body armor (e.g., Herczeg et al. 2010; Shikano et al. 2013), gigantism (e.g., Herczeg et al. 2009), increased aggression (e.g., Herczeg and Välimäki 2011), and divergent brain architecture

(e.g., Gonda et al. 2012) compared with marine populations. Earlier population genetic and phylogeographic studies (Shikano et al. 2010a; Teacher et al. 2011; Bruneaux et al. 2013) also suggest that postglacial recolonization and associated founder events have strongly affected the genetic variability and structure of current populations. Despite this progress in understanding local adaptation and differentiation among nine-spined stickleback populations (see also: Karhunen et al. 2014), possible differences in the extent and levels of genome-wide LD among populations and habitat types remain unknown.

The main aim of this study was to quantify and compare the patterns and extent of genome-wide LD in nine-spined stickleback populations from different habitats (viz. marine, river, lake, and pond). To this end, we used genotypic data on 109 microsatellite loci from 13 different nine-spined stickleback populations. Because isolated freshwater populations have very low levels of genetic variability (Shikano et al. 2010a; Bruneaux et al. 2013) and thus, are likely to have smaller effective population sizes and be more susceptible to stochastic demographic events than open and more genetically variable marine populations, we expected to find greater levels of genomic LD in freshwater compared with marine populations.

MATERIALS AND METHODS

Study populations and samples

A total of 312 nine-spined stickleback individuals (24 per population) from three marine and 10 freshwater populations were included in the analyses. The sampling sites covered a large part of the Fennoscandian area and encompassed a diverse array of habitats (viz. marine, river, lake, and pond populations; Figure 1 and Table 1). Marine fish were collected from the White Sea (Lev) and the Baltic Sea (Sbol and Hel),
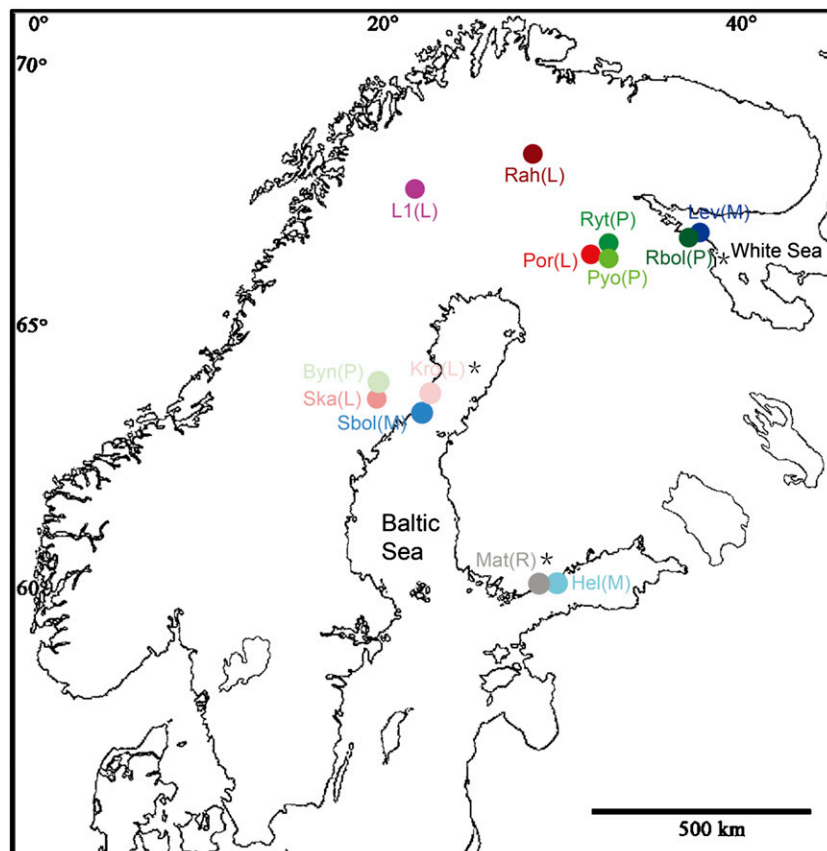

Figure 1 Map showing the locations of 13 nine-spined stickleback populations used in this study. The abbreviations of the populations are defined in Table 1. The letter in brackets stands for habitat type (M = marine; R = river; L = lake; P = pond). Asterisks indicate coastal freshwater populations.

Population	Habitat	n	npl	A	Ar	Pr	H_O	H_E	F_{IS} (95% CI)
Helsinki (Hel)	Marine	24	106	760	6.97	0.72	0.554	0.569	0.027 (−0.027−0.034)
Bölesviken (Sbol)	Marine	24	103	757	6.94	0.69	0.556	0.573	0.030 (−0.025−0.038)
Levin Navolok (Lev)	Marine	24	103	765	7.02	0.63	0.531	0.545	0.026 (−0.025−0.029)
Kroktjärnen (Kro)	Lake	24	103	647	5.94	0.44	0.562	0.574	0.021 (−0.024−0.020)
Västre-Skavträsket (Ska)	Lake	24	52	266	2.44	0.34	0.201	0.199	−0.012 (−0.088−0.019)
Iso-Porontima (Por)	Lake	24	90	397	3.64	0.20	0.300	0.320	0.064 (0.012−0.066)
Lake 1 (L1)	Lake	24	79	266	2.44	0.17	0.309	0.309	−0.002 (−0.064−0.015)
Rahajärvi (Rah)	Lake	24	89	524	4.81	0.37	0.358	0.368	0.030 (−0.041−0.046)
Bynastjärnen (Byn)	Pond	24	67	240	2.20	0.04	0.240	0.239	−0.004 (−0.073−0.019)
Pyöreälampi (Pyo)	Pond	24	33	164	1.50	0.03	0.084	0.085	0.002 (−0.092−0.047)
Bolotnoje (Rbol)	Pond	24	104	656	6.02	0.31	0.523	0.533	0.020 (−0.029−0.020)
Rytilampi (Ryt)	Pond	24	68	253	2.32	0.27	0.232	0.230	−0.009 (−0.081−0.016)
Matinoja (Mat)	River	24	103	507	4.65	0.24	0.522	0.530	0.015 (−0.040−0.023)

n, number of sampled individuals; npl, number of polymorphic loci; A, number of alleles; Ar, allelic richness; Pr, private allelic richness; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, departure from panmixia; CI, confidence interval.

whereas freshwater fish were collected from one river (Mat), five lakes (Rah, L1, Por, Ska, and Kro) and four ponds (Ryt, Rbol, Pyo, and Byn; Figure 1). Three of the freshwater populations (Mat, Kro, and Rbol) located in close proximity to coastlines (Figure 1) were referred to as coastal freshwater populations, while the other seven freshwater populations (Rah, L1, Por, Ska, Ryt, Pyo, and Byn; Figure 1) were considered as inland freshwater populations.

Molecular analyses

Total genomic DNA for the samples was extracted from fin clips using the phenol–chloroform method (Taggart *et al.* 1992) following proteinase K digestion. The same panel of 112 microsatellites as used by Shikano *et al.* (2010b) was used in all analyses. The genotyping data of the microsatellite markers for eight populations (Lev, Sbol, Hel, Mat, L1, Kro, Rbol, and Pyo) were taken from Shikano *et al.* (2010b,c), whereas the data for other five populations (Rah, Por, Ska, Ryt, and Byn) were produced in the present study (Supporting Information, File S1). Polymerase chain reactions (PCRs) were carried out using the QIAGEN multiplex PCR Kit (QIAGEN) in a reaction volume of 10 μL containing 1× QIAGEN multiplex PCR Master Mix, 0.5× Q-Solution, 2 pmol of each primer, and 10–20 ng of genomic DNA. The PCR amplifications were performed using the following cycle: initial activation at 95° for 15 min, followed by 30 s at 94°, 90 s at 53 or 55°, and 60 s at 72° for 30 cycles, ending with a final extension at 60° for 5 min. PCR products were resolved on 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.

Population genetic analyses

Within-population observed heterozygosities (H_O), expected heterozygosities (H_E), inbreeding coefficient (F_{IS}), and allele frequencies were calculated with FSTAT v2.9.3.2 (Goudet 2002). The proportion of rare alleles (allele frequency <5%) in each population was estimated using Microsoft Excel. Measures of allelic richness and private allelic richness for each population were calculated using HP-RARE (Kalinowski 2005), accounting for rarefaction.

Three approaches were used to investigate population genetic structure. First, pairwise F_{ST} among populations was calculated using GENETIX v4.03 (Belkhir *et al.* 2004), and the significance of F_{ST} values was evaluated via 10,000 permutations. Second, principal

component analysis was performed at the individual level using the program GenAlEx 6.501 (Peakall and Smouse 2006, 2012). Third, to assess the relative contributions of potential factors to population differentiation, a hierarchical analysis of molecular variance was performed using the program Arlequin v3.5 (Excoffier and Lischer 2010), based on three different grouping patterns of populations: habitat type I (marine, lake, pond, and river), habitat type II (marine and freshwater) and geographic proximity (Hel and Mat; Sbol and Kro; Ska and Byn; Por, Pyo, and Ryt; Rbol and Lev; L1; and Rah; see Figure 1). Statistical significance was assessed with 10,000 permutations. As population substructure tends to inflate LD (Nei and Li 1973; Pritchard and Przeworski 2001), we performed Bayesian clustering analyses in STRUCTURE v2.3.4 (Pritchard *et al.* 2000) to examine whether the observed high levels of LD (see the section *Results*) were due to within-population substructuring. We conducted three independent runs for each K-value ranging from 1 to 20. The admixture model and correlated allele frequencies model (Falush *et al.* 2003; Excoffier *et al.* 2005) were used, with 500,000 iterations after a 100,000 burn-in for each run. Also hidden family structure could amplify LD, and thus, we used Queller and Goodnight's method (Queller and Goodnight 1989) implemented in program IDENTIX v1.1.5 (Belkhir *et al.* 2002) to estimate pairwise relatedness coefficient between individuals within each population.

Signatures of genetic bottlenecks were tested for each population using two methods. First, we used the heterozygosity excess method (Luikart *et al.* 1998) as implemented in the program Bottleneck v1.2.02 (Piry *et al.* 1999) to test for recent reductions in population size. We ran the program under the two-phased mutation model (TPM) with 90% single-step mutations. Statistical significance of the results was evaluated by 1000 iterations with a one-tailed Wilcoxon signed-rank test. Second, we used the M-ratio method (Garza and Williamson 2001) to detect historical population contractions (Garza and Williamson 2001; Williamson-Natesan 2005). Population-specific values of M (the number of alleles / the allele size range) and M_c (the critical value of M) were estimated using the programs M_P_VAL and CRITICAL_M (Garza and Williamson 2001), respectively. For each run, the simulations consisted of 10,000 iterations with the average mutation rate (μ) of 1.5×10^{-4} per generation (Shimoda *et al.* 1999), a TPM with 10% multistate change and 3.5 base steps for the mean size of multistep mutations (Garza and Williamson 2001). We tested three conservative values of theta ($\theta = 4N_e\mu$) that equate to a prebottleneck effective population size (N_e) of 1000, 5000, and 10,000 for the three marine and three coastal freshwater populations, and a prebottleneck

N_e of 100, 500, and 1000 for the seven inland freshwater populations. The observed value of M was compared with the corresponding M_c, and a lower value of M relative to M_c indicated a historical population bottleneck (Garza and Williamson 2001).

Linkage map and haplotype phasing

Since nine-spined and three-spined sticklebacks (*Gasterosteus aculeatus*) have the same number ($n = 21$) of chromosomes (Chen and Reisman 1970) and syntenic locations of microsatellite loci are conserved between these two closely related species (Shapiro *et al.* 2009; Shikano *et al.* 2010c, 2013), we built the genomic distance-based (Mb) linkage map for the nine-spined stickleback through its homology with the three-spined genome assembly (http://www.ensembl.org/Gasterosteus_aculeatus/index.html). BLAST searches were performed to locate the 112 nine-spined stickleback microsatellite markers in the three-spined stickleback genome using the BLASTN tool in the Ensembl database. Initial searches were performed with the default conditions, and a locus was assigned to a genomic location if it provided a unique hit at $E \leq 1e^{-10}$. When a locus provided multiple matches at $E \leq 1e^{-10}$, it was unassigned unless the best hit had an E value at least 10 decimal places lower than the next best one. For ease of comparison, we numbered linkage groups (LGs) for the nine-spined stickleback linkage map in accordance with the syntenic LGs in the three-spined stickleback (Figure 2).

The gametic phase of haplotypes and missing genotypes were inferred from genotype data for each LG in each population and habitat type using a Bayesian statistical method as implemented in PHASE v2.1 (Stephens *et al.* 2001; Stephens and Scheet 2005). In each run, we chose the original model defined in Stephens *et al.* (2001), and set the number of iterations to 1000, thinning interval to 1 and a burn-in to 100. Ten independent runs were performed with different seeds to check for consistency between the results. We considered the PHASE results to be consistent when no less than eight runs gave the same inferred haplotypes, and in such case the consistent haplotypes were used in the subsequent calculations; otherwise, the haplotypes from the run with the highest average value for the goodness of fit statistics were used for the subsequent analyses (Stephens *et al.* 2001).

LD analyses

Two different gametic LD measures, multiallelic D' and r^2, were used. The two LD estimates were derived from the standard measure of LD between two alleles at two different loci: $D_{ij} = p(A_iB_j) - p(A_i)p(B_j)$, where $p(A_i)$ is the frequency of allele A_i at locus A, $p(B_j)$ is the frequency of allele B_j at locus B, and $p(A_iB_j)$ is the frequency of haplotype A_iB_j in the population.

Multiallelic D' was estimated as (Lewontin 1964; Hedrick 1987):

$$D' = \sum_{i=1}^{k} \sum_{j=1}^{l} p(A_i)p(B_j) \left| \frac{D_{ij}}{D_{ij}^{max}} \right|$$

where k and l were the number of alleles for markers A and B, respectively, and

$$\begin{aligned} D_{ij}^{max} &= \min[p(A_i)p(B_j), (1 - p(A_i))(1 - p(B_j))] \\ &\quad \text{when } D_{ij} < 0 \text{ and} \\ D_{ij}^{max} &= \min[p(A_i)(1 - p(B_j)), p(B_j)(1 - p(A_i))] \\ &\quad \text{when } D_{ij} \geq 0. \end{aligned}$$

Multiallelic r^2 was estimated as (Hill and Robertson 1968):

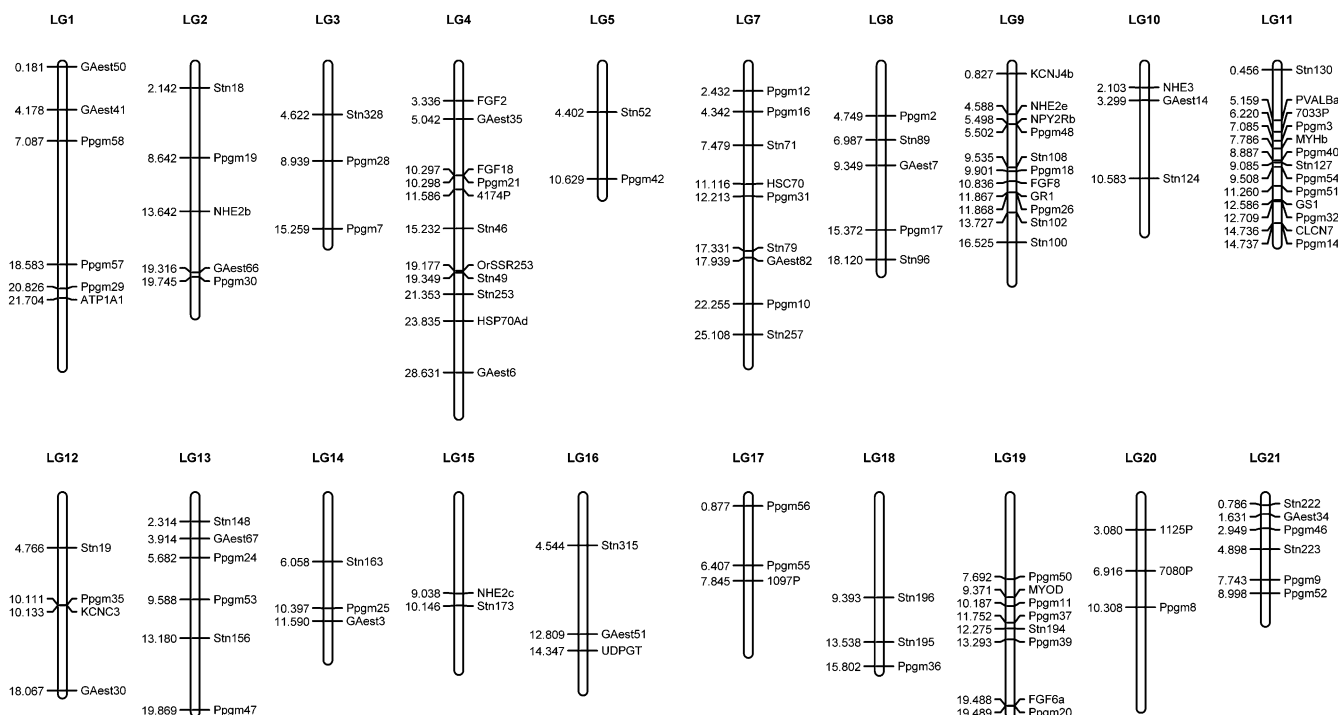$$r^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{D_{ij}^2}{p(A_i)(1 - p(A_i))p(B_j)(1 - p(B_j))}$$

Figure 2 Genome-wide linkage map for nine-spined stickleback based on 109 microsatellite markers. Genomic distances (in Megabases, Mb) are listed on the left side of each linkage group (LG). All 109 loci were involved in linkage disequilibrium (LD) analyses.

We computed D' and r^2 for all pairwise syntenic markers in each population and habitat type using the program PowerMarker v3.25 (Liu and Muse 2005). Pearson's and Kendall's correlation tests were performed to investigate the correlation between D' and r^2 values within population or habitat. Because the measure D' commonly has been used in studies of wild vertebrates (e.g., Backström et al. 2006; Hohenlohe et al. 2012) and has more power to detect LD (Devlin and Risch 1995), it was used in the following analyses to facilitate comparison of our results with those of other studies. Logarithmic regression plots of D' values of all syntenic pairwise markers against genomic distances (Mb) in each population and habitat type were generated in Microsoft Excel. The half-length of LD (Reich et al. 2001), i.e., the distance at which it falls to 0.5, was evaluated.

Mann-Whitney U-tests (Mann and Whitney 1947) were used to assess the statistical significance of differences in D' values between habitat types. Kruskal-Wallis tests (Kruskal and Wallis 1952) were used to assess the significance of differences in D' values across all of the populations or among populations within the same habitat type. Partly different polymorphic markers were involved in different population-specific LD analyses (Table 1), hence the variation in marker distance between populations could potentially influence statistical significance tests of D' values. In order to control for this, we used analysis of covariance (ANCOVA) in which population and habitat were treated as random and fixed factors, respectively, and associated D' values were regarded as dependent variables, with physical distance between markers as a covariate. Furthermore, there were differences in marker density in different LGs. In order to corroborate the LD patterns observed in the genome-wide analyses, we examined LD patterns in four LGs with the greatest marker densities (i.e., LGs 9, 11, 19, and 21; Figure 2) for each population. All statistical tests were conducted in SPSS 16.0 (SPSS Inc, Chicago, IL), and Bonferroni corrections (Rice 1989) were applied to adjust significance levels when multiple testing was involved.

To examine whether observed high levels of LD could be an artifact due to haplotype phasing, we also estimated the composite LD measure (Weir 1996) based on unphased genotypic data using the method described in Zaykin et al. (2008). In addition, to examine the effect of rare alleles (allele frequency <5%) on the levels of LD, we recalculated both haplotypic and composite LD measures with rare alleles excluded.

RESULTS

Population genetic analyses

One-hundred nine microsatellite markers were successfully mapped to the three-spined stickleback genome. The basic indices of within-population genetic variability are given in Table 1. The number of polymorphic loci ranged from 33 (in Pyo) to 106 (in Hel) depending on the population. Allelic richness and expected heterozygosities (H_E) estimated across all loci ranged from 1.50 (in Pyo) to 7.02 (in Lev), and from 0.085 (in Pyo) to 0.574 (in Kro), respectively (Table 1). Private allelic richness for each population ranged from 0.03 (in Pyo) to 0.72 (in Hel; Table 1). The marine (Hel, Sbol, and Lev) and coastal freshwater populations (Mat, Kro, and Rbol) had much greater genetic diversities ($H_E = 0.530–0.574$; Table 1) than the seven inland freshwater populations (Ska, Byn, Por, Pyo, Ryt, L1, and Rah; $H_E = 0.085–0.368$; Table 1). F_{IS} values and their 95% confidence intervals did not deviate significantly from zero in any of the populations (Table 1). A high proportion of rare alleles was observed within populations, ranging from 0.15 in Pyo to 0.53 in Hel (Table S1).

The extent of population differentiation as measured by F_{ST} among population pairs varied greatly ($F_{ST} = 0.003–0.724$), most of which were significant (52/78, $P < 0.05/78 = 0.000641$; Table S2). In general, F_{ST} values between inland freshwater populations were always greater than those between marine or coastal freshwater populations (Table S2). Principal component analysis revealed that the first and second axes accounted for 13.7% and 10.1% of variation in allele frequencies, respectively (Figure S1). The individuals from the inland freshwater populations clustered more tightly than those from the coastal freshwater and marine populations (Figure S1). Analysis of molecular variance analyses suggested that 7.4% of the total genetic variation was explained by geographic proximity ($P < 0.001$), whereas the factors of habitat type (marine vs. lake vs. pond vs. river, −1.9%, $P > 0.05$; marine vs. freshwater, −2.1%, $P > 0.05$; see Table 2) did not contribute to the patterns of genetic differentiation. Based on the value of ΔK (Evanno et al. 2005), STRUCTURE analyses indicated that the most probable K was nine (Figure S2). No substructure was found within any of the populations at both the optimal K value (i.e., 9) and the maximum tested K value (i.e., 20; Figure S2). Thus, population substructuring was unlikely to account for the observed high levels of LD. The estimated pairwise relatedness coefficients were generally small (e.g., < 0.2) for 12 populations except Pyo (File S2), suggesting that most individuals should be unrelated; hence, family structure was not an explanation for the high LD values.

A signal of recent population bottleneck was detected in only one population (L1; $P = 0.03$) under the TPM using the heterozygosity excess method. However, all populations except Pyo showed strong evidence for historical population bottlenecks using the M-ratio method, despite the differences in pre-bottleneck N_e (Table S3). Observed population-specific M-ratio values ranged from 0.670 to 0.898, and most (12/13, except Pyo) were lower than the corresponding M_c values (Table S3). It was unexpected that no bottleneck was detected in Pyo because this population had the lowest genetic diversity of all populations in this study (Table 1). However, this could be due to a small number of polymorphic markers ($n = 33$; Table 1) segregating in the population.

■ **Table 2 Analysis of molecular variance in three different population groupings based on 109 microsatellite markers**

Population Groups Defined	Components	Percentage of Variation
Four groups according to habitat type	Among groups	−1.90
	Among populations within groups	34.14***
	Within populations	67.76***
Seven groups according to geographic proximity	Among groups	**7.35***
	Among populations within groups	25.65***
	Within populations	67.00***
Marine vs. freshwater populations	Among groups	**−2.05**
	Among populations within groups	33.75***
	Within populations	68.30***

***$P < 0.001$. The percentage of genetic variation among groups is indicated by bold type.

Linkage map

Based on homologous positions in the three-spined stickleback genome, the 109 mapped microsatellites defined a total number of 20 LGs of the nine-spined stickleback (Figure 2). Two to 13 markers were mapped to each of the LGs, but none of the markers mapped to LG6 of the three-spined stickleback (Figure 2). Based on the three-spined stickleback genome assembly, the average interval between adjacent markers was 2.738 Mb, with the smallest spacing of 0.001 Mb and the largest of 11.496 Mb. The median distance between adjacent markers was 2.004 Mb. With regard to different LGs, the average inter-marker distance ranged from 1.19 Mb in LG11 to 6.227 Mb in LG5. Inferred haplotypes from the program PHASE were largely consistent across the ten replicate runs, and approximately 90% of the total number of loci had phase probabilities of more than 0.8, indicating that the results were reliable.

Genome-wide LD

Overall, the levels of syntenic LD as measured by D' were relatively high (Table 3), but varied among the 13 populations (Kruskal-Wallis, $\chi^2 = 100.20$, $d.f. = 12$, $P < 0.001$; ANCOVA, $F_{12, 2911} = 10.64$, $P < 0.001$). When different habitat types were considered, lake (Mann-Whitney, $Z = -4.99$, $P < 0.001$; ANCOVA, $F_{1, 650} = 15.37$, $P < 0.001$), pond (Mann-Whitney, $Z = -6.91$, $P < 0.001$; ANCOVA, $F_{1, 646} = 45.75$, $P < 0.001$), and river (Mann-Whitney, $Z = -4.95$, $P < 0.001$; ANCOVA, $F_{1, 646} = 28.13$, $P < 0.001$) habitats, showed significantly greater D' values than the marine habitat. The greatest average D' values were observed in the pond habitat (Table 3). There were no differences in D' values among the different marine populations (viz. Hel, Sbol, Lev; Kruskal-Wallis, $\chi^2 = 2.13$, $d.f. = 2$, $P = 0.34$; ANCOVA, $F_{2, 937} = 0.92$, $P = 0.40$), but significant differences were found among the lake (viz. Kro, Ska, Por, L1, Rah; Kruskal-Wallis, $\chi^2 = 64.94$, $d.f. = 4$, $P < 0.001$; ANCOVA, $F_{4, 1049} = 19.59$, $P < 0.001$) and pond (viz. Byn, Pyo, Rbol, Ryt; Kruskal-Wallis, $\chi^2 = 15.95$, $d.f. = 3$, $P < 0.001$; ANCOVA, $F_{3, 609} = 7.08$, $P < 0.001$) populations. When restricting the comparisons to LGs with high density markers (LG9, LG11, LG19, LG21; 38 markers in total; Figure 2), the D' values were similar to those obtained in the genome-wide analyses (all LGs; 109 markers in total; Figure 2) in 12 populations (Figure S3). This supports the view that the relatively low number of microsatellite markers used in this study can indeed yield information about general patterns of genome-wide LD. When LD was measured with r^2, lower absolute values were observed (Table S4) compared with those of D' (Table 3). However, D' and r^2 values were positively and significantly correlated in most populations and habitat types (Table S5).

Comparison of the patterns of LD decay as a function of genomic distance revealed very weak and statistically nonsignificant ($R^2 < 0.01$, $P > 0.05$; Table S6 and Figure 3) correlations between D' and genomic distance. With regard to LD decay in different habitats, the dataset of all marine populations combined or all freshwater populations combined showed higher correlations and shorter LD half-length compared with the combined lake or pond datasets (Figure 4 and Table S6). Interestingly, we found that the three coastal freshwater populations (Mat, Kro, Rbol; Figure 3B), which were geographically close to the marine populations (Hel, Sbol, Lev; Figure 3A), exhibited similar LD patterns as their marine neighbors, but deviated from the typical LD pattern in the inland freshwater populations (Figure 3C and Table 3). In addition, LD values increased slightly with genomic distance in three inland freshwater populations (Ska, Byn, Pyo; Figure 3D and Table S6), and the level of LD in Por was independent of genomic

■ **Table 3 Linkage disequilibrium estimate (*D′*) and associated estimation error for syntenic markers in 13 nine-spined stickleback populations and five habitat types (marine, lake, pond, river, and coastal freshwater) using 109 microsatellite markers**

Data Set	Physical Distance Interval (Syntenic)					Overall (Syntenic)
	0−5 Mb	5.001−10 Mb	10.001−15 Mb	15.001−20 Mb	>20 Mb	
Hel (M)	0.557 (0.020)	0.549 (0.024)	0.500 (0.039)	0.492 (0.051)	0.649 (0.117)	0.544 (0.014)
Sbol (M)	0.557 (0.020)	0.553 (0.024)	0.534 (0.037)	0.578 (0.048)	0.519 (0.095)	0.553 (0.014)
Lev (M)	0.559 (0.020)	0.590 (0.026)	0.551 (0.037)	0.571 (0.043)	0.579 (0.103)	0.570 (0.014)
Kro (L)	0.509 (0.021)	0.504 (0.021)	0.445 (0.039)	0.442 (0.048)	0.373 (0.101)	0.491 (0.013)
Ska (L)	0.651 (0.050)	0.539 (0.050)	0.596 (0.090)	0.959 (0.042)	0.700 (0.174)	0.631 (0.033)
Por (L)	0.630 (0.031)	0.716 (0.035)	0.564 (0.068)	0.615 (0.088)	0.541 (0.147)	0.648 (0.021)
L1 (L)	0.551 (0.031)	0.428 (0.042)	0.536 (0.074)	0.453 (0.101)	0.471 (0.279)	0.506 (0.023)
Rah (L)	0.646 (0.027)	0.672 (0.029)	0.719 (0.051)	0.616 (0.082)	0.677 (0.053)	0.663 (0.018)
Byn (P)	0.553 (0.046)	0.636 (0.056)	0.628 (0.105)	0.608 (0.102)	0.767 (0.149)	0.605 (0.032)
Pyo (P)	0.804 (0.064)	0.710 (0.101)	0.579 (0.421)	1.000 (0.000)	—	0.781 (0.051)
Rbol (P)	0.537 (0.022)	0.599 (0.023)	0.526 (0.038)	0.455 (0.042)	0.522 (0.085)	0.551 (0.014)
Ryt (P)	0.631 (0.041)	0.604 (0.052)	0.577 (0.092)	0.521 (0.105)	0.768 (0.232)	0.612 (0.029)
Mat (R)	0.509 (0.021)	0.562 (0.027)	0.532 (0.041)	0.426 (0.043)	0.580 (0.122)	0.527 (0.015)
Marine (average[a])	0.558 (0.001)	0.564 (0.013)	0.528 (0.015)	0.547 (0.028)	0.582 (0.038)	0.556 (0.008)
Lake (average[a])	0.597 (0.037)	0.572 (0.069)	0.572 (0.058)	0.617 (0.121)	0.552 (0.080)	0.588 (0.047)
Pond (average[a])	0.631 (0.070)	0.637 (0.029)	0.578 (0.024)	0.646 (0.141)	0.686 (0.082)	0.637 (0.058)
CF (average[a])	0.518 (0.009)	0.555 (0.028)	0.501 (0.028)	0.441 (0.009)	0.492 (0.062)	0.523 (0.017)
Marine (combined[b])	0.433 (0.017)	0.450 (0.021)	0.393 (0.033)	0.357 (0.035)	0.406 (0.082)	0.428 (0.012)
Lake (combined[b])	0.507 (0.017)	0.486 (0.018)	0.469 (0.028)	0.451 (0.042)	0.503 (0.111)	0.491 (0.011)
Pond (combined[b])	0.530 (0.022)	0.578 (0.023)	0.553 (0.037)	0.531 (0.046)	0.547 (0.069)	0.550 (0.014)
CF (combined[b])	0.383 (0.017)	0.397 (0.021)	0.332 (0.020)	0.357 (0.039)	0.413 (0.094)	0.380 (0.011)
River	0.509 (0.021)	0.562 (0.027)	0.532 (0.041)	0.426 (0.043)	0.580 (0.122)	0.527 (0.015)

M, marine; L, lake; P, pond; R, river; CF, Coastal freshwater, including Kro, Rbol, and Mat. The population abbreviations are defined in Table 1. The value in the brackets is the estimation error associated to the mean D' value, obtained by dividing the SD of D' value by the square root of the number of marker pairs used to measure LD in each distance bin (Table S7).

[a] D' value is directly obtained from the averaged D' value of relevant populations.

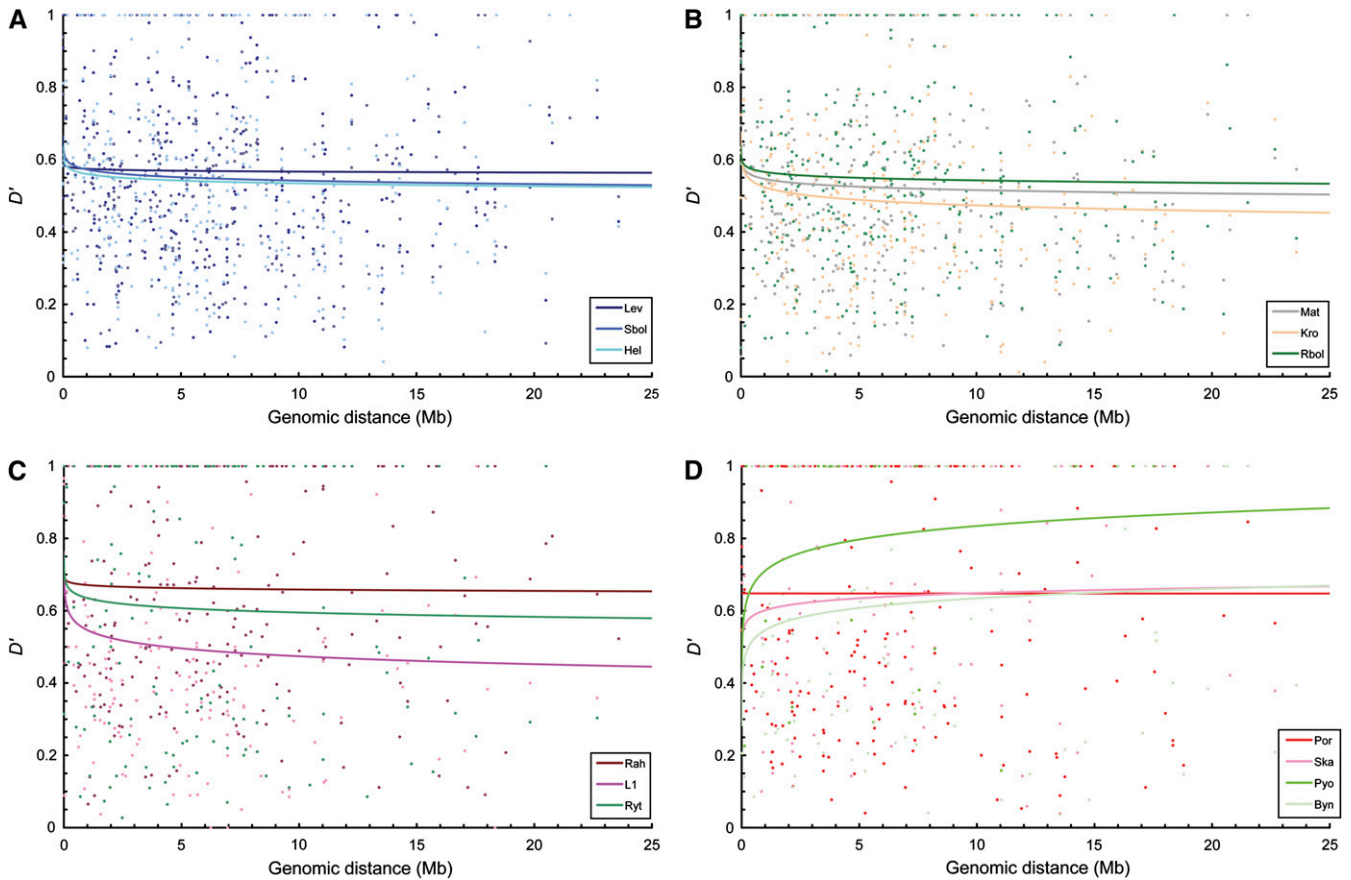[b] D' value is calculated from the combined original haplotype data of relevant populations.

Figure 3 Observed linkage disequilibrium (LD, measured by *D'*) as a function of genomic distance (Megabases, Mb) between all syntenic markers in nine-spined stickleback populations using 109 microsatellite loci. (A) LD decay in three marine populations. (B) LD decay in three coastal freshwater populations. (C) LD decay in three inland freshwater populations with common decay pattern. (D) LD decay in four inland freshwater populations with unusual decay patterns. For population abbreviations, see Table 1.

distance (Figure 3D and Table S6). This finding could be ascribable to stochasticity caused by the small number of marker pairs used to measure LD in each distance bin in these highly homozygous populations (Table S7).

The composite *D'* and *r²* values were relatively high (Table S8) and comparable with the levels of haplotypic LD values (Table 3 and Table S4), indicating that observed high levels of LD were unlikely to be explainable as an effect of haplotype phasing. When the rare alleles
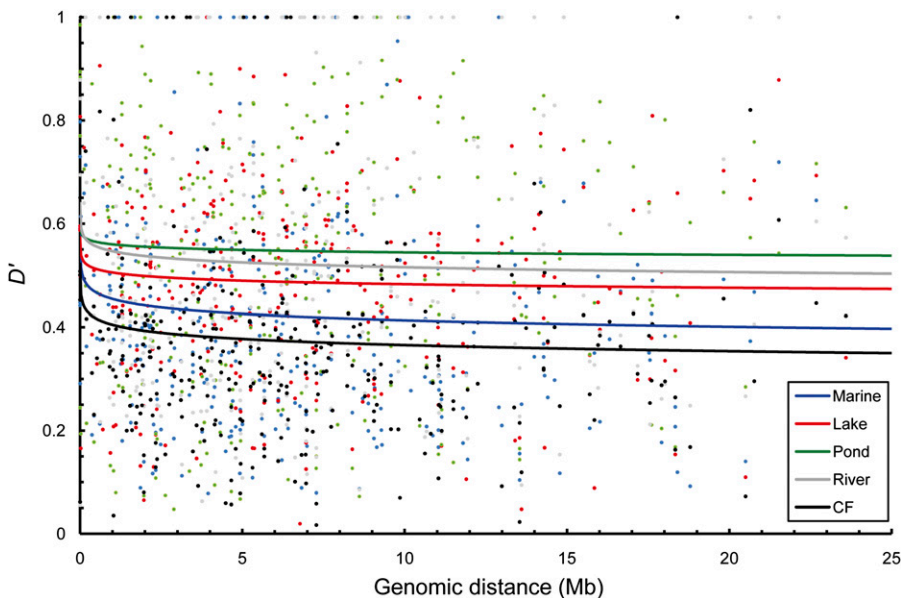

Figure 4 Linkage disequilibrium (LD, measured by *D'*) decay between all syntenic markers in five different habitat types (blue = marine populations, red = lake populations, green = pond populations, gray = river population, black = coastal freshwater [CF] populations). Combined population data of 109 microsatellite loci within the same habitat type were employed to estimate habitat-specific *D'* values.

were excluded, both haplotypic and composite D' values were smaller, but the overall syntenic D' value was still above 0.4 in almost all the populations (Table S8). On the contrary, both haplotypic and composite r^2 values became larger without the rare alleles (Table S8). Notably, irrespective of whether inferred haplotypic data or unphased genotypic data were used and whether the rare alleles were involved in the analyses or not, the findings about the LD patterns among habitat types (*i.e.*, Pond > Lake > Marine; Coastal freshwater is similar to Marine) based on combined data remained largely unchanged (Table 3, Table S4, and Table S8).

DISCUSSION

In general, low-to-moderate genetic diversity, strong genetic differentiation, and high levels of genome-wide LD were observed in Fennoscandian nine-spined stickleback populations. The extent and patterns of LD varied among populations and habitat types. Isolated and small freshwater populations tended to have greater LD compared with open marine populations. In the following, we will discuss these findings and their implications to our understanding of the factors influencing levels and extent of genomic LD in the wild.

Several recent studies have focused on fine-scale LD in commercially important fishes (*e.g.*, Hayes *et al.* 2006), whereas genome-wide levels of LD in wild fish populations remain largely unexplored, with few exceptions (*e.g.*, Hohenlohe *et al.* 2012; Roesti *et al.* 2013). We found high levels of LD in the studied nine-spined stickleback populations, and in this respect the results are comparable with those from the closely related three-spined stickleback (Mattern 2004; Mattern and Mclennan 2004), in which high magnitudes of LD were observed in both freshwater and marine populations (Hohenlohe *et al.* 2012). The high degree of LD in nine-spine sticklebacks did not come as a surprise in the view that earlier population genetic studies of this species (Shikano *et al.* 2010a; Teacher *et al.* 2011; Bruneaux *et al.* 2013) have suggested limited gene flow and low effective population sizes, both of which are factors expected to amplify genetic drift and thus the accumulation of LD (Service *et al.* 2006; Slatkin 2008; Charlesworth 2009). Likewise, demographic events such as founder effects and population bottlenecks can create high LD (*e.g.*, Nei and Li 1973; Zhang *et al.* 2004). In our case, the evidence for genetic bottlenecks in 12 of the 13 populations using *M*-ratio tests indicated that historical bottlenecks most probably have contributed to the high magnitude of genome-wide LD. Given that the stickleback populations studied here have been colonized after the last glacial maximum (<10,000 years ago), founder effects associated with postglacial recolonization also may account for the high LD. It should be noted that we have not taken recombination into account in our LD estimation due to its heterogeneity across the genome. Nevertheless, this should not affect the observed habitat or population differences in LD if the recombination hotspots are congruent in different populations, as has been reported for human populations (Conrad *et al.* 2006). One should also note that marker type can influence observed levels and extent of LD. For instance, microsatellite markers have more alleles per locus than SNP markers, and hence, they generally show higher levels of LD than SNPs (Chapman and Wijsman 1998). Consequently, the strong LD found here could partly be attributed to the high information content of microsatellites (Pritchard and Przeworski 2001). However, it is unlikely that this would be the sole explanation for the high levels of LD in nine-spined sticklebacks, especially in the view that this explanation cannot account for observed habitat or population differences in levels of LD. Other factors such as gene conversion, inversions and chromosome rearrangement could also have influenced

the levels of LD in nine-spined sticklebacks, but the role of these factors remains to be investigated in future studies.

Despite the generally high magnitude of LD within populations, we also found significant differences in the levels and extent of LD between habitat types. The greatest levels of LD were observed in the seven inland freshwater populations, which was not unexpected as these are all population isolates that have been subject to substantial genetic drift due to initial founder effects, subsequent isolation and small effective population sizes. This drift has also led to reduced allelic diversity as reflected by low heterozygosities, low allelic richness, and overrepresentation of monomorphic microsatellite loci and rare alleles in these populations. This finding aligns well with those of earlier studies, which have shown that population isolates typically are characterized by low levels of genetic variation and high levels of LD (*e.g.*, Arcos-Burgos and Muenke 2002; Li and Merilä 2010). Interestingly, the patterns of LD and genetic variation in the three coastal freshwater populations were similar to those in the adjacent marine populations. Similar observations also were reported in an earlier study of Swedish nine-spined sticklebacks, which showed little genetic and morphological differentiation between marine and coastal lake populations in the Baltic Sea region (Herczeg *et al.* 2009; Mobley *et al.* 2011). One plausible explanation for these observations is that the coastal freshwater populations are influenced by admixture/gene flow from adjacent marine populations, or that they have only recently become isolated from the marine populations (Herczeg *et al.* 2009; Mobley *et al.* 2011).

Different metrics have been developed to measure the degree of LD, and we employed both D' and r^2 estimators in this study. We found that the former yielded consistently higher values than the latter; such differences have also been reported in previous LD studies (*e.g.*, Shifman *et al.* 2003; García-Gámez *et al.* 2012; Espigolan *et al.* 2013). Several possible underlying factors could account for such differences, including large allele frequency differences between markers (*e.g.*, Ardlie *et al.* 2002; Wray *et al.* 2011) as was observed in this study (File S1). Likewise, the high proportion of rare alleles (allele frequency <5%; Table S1) and consequent loss of haplotypes in the populations may also yield high D' values yet low r^2 values (Slatkin 2008; Purcell *et al.* 2009). Despite this discrepancy in absolute values of D' and r^2, the two estimators were positively correlated in our data (Table S5), and gave consistent LD patterns in inter-habitat comparisons (Table 3 and Table S4). Thus, conclusions drawn from D' values are qualitatively similar to those obtained using r^2 values in respect to patterns of LD across habitat types.

Rare alleles (allele frequency <5%) tend to elevate D' values (Teare *et al.* 2002); hence, they have often been eliminated from LD analyses. In our study, rare alleles were frequent in many populations, and this partly explains the high D' values in this study. We believe that the inclusion of rare alleles in our LD analyses was reasonable on the following grounds: First, the overall syntenic D' values remained relatively high (>0.4) in all of the 13 populations when the rare alleles were excluded. The differences in LD among habitat types (*i.e.*, Pond > Lake > Marine) remained unchanged even if the rare alleles were excluded. Second, rare variants can convey important information in genome-wide genetic studies (Dickson *et al.* 2010). Thus, given that the high proportion of rare alleles is an inherent characteristic of the nine-spined stickleback populations investigated here, ignoring them might bias the results. Third, given the demographic history of these populations, a high frequency of rare alleles is to be expected. Population genetics theory suggests that rare variants are likely to be recently derived alleles (Watterson and Guess 1977), and a large number of rare variants could derive from recent population expansions

(Pritchard 2001; Gorlov *et al.* 2008). As for Fennoscandian nine-spined sticklebacks, earlier studies (Shikano *et al.* 2010a; Teacher *et al.* 2011; Bruneaux *et al.* 2013) indicated that populations inhabiting this region derived from ancestors in refugia from which the recolonization occurred approximately 10,000 years ago. Population expansions are very likely to have been involved in this re-establishment process, and thus, result in the large number of rare alleles in marine and coastal freshwater populations observed here. Previous studies have also indicated that inland freshwater populations have been established from marine populations recurrently (Teacher *et al.* 2011; Bruneaux *et al.* 2013). This finding, coupled with the fact that much genetic variation including rare alleles has been lost due to drift in inland isolates may explain why fewer rare alleles were observed in inland as compared to marine populations. In fact, within the same geographic region, an excess of rare alleles have also been observed in human (Reich *et al.* 2001) and Norway spruce (*Picea abies*) populations (Larsson *et al.* 2013).

Our findings of genomic LD and genetic variability have several important implications for gene mapping studies in nine-spined sticklebacks. First, given the high level of LD, a relatively small number of markers are required to cover a relatively large genomic region in QTL-mapping studies. Second, given the previous consideration, the mapping resolution will be relatively low because large genomic regions are likely to be inherited as linked clusters. Third, given the high frequency of rare alleles, nine-spined stickleback populations might prove to be suitable for rare variant mapping of complex traits. Nevertheless, although this study provides some preliminary insight on variation in LD across the nine-spined stickleback genome, one should bear in mind that the relatively low number of markers and their non-uniform distribution over the LGs and populations limit the inferences. Further exploration based on a larger number of markers, together with a high-density linkage map would pave the road for more refined inferences.

To sum up, the results provide the first investigation of genome-wide LD patterns in the nine-spined stickleback, and also one of the most extensive studies exploring patterns of habitat related variation in LD in wild vertebrates. In general, high levels of LD were observed in most of the analyzed populations, and more interestingly, higher levels of LD were detected in inland freshwater than in costal populations. This habitat patterning in the levels of LD matches what we discovered—and what has been known from earlier studies—about habitat-specific differences in demographic history and effective population size in these populations. The levels of LD uncovered in present study also suggest that studies seeking to disclose the genetic basis of phenotypic traits using QTL-mapping approaches may face challenges, especially in inland freshwater populations which are low in genetic variability and exhibit high levels of LD: the few polymorphic markers segregating in those populations are likely to be associated for long stretches of linked genes.

LITERATURE CITED

Arcos-Burgos, M., and M. Muenke, 2002   Genetics of population isolates. Clin. Genet. 61: 233–247.

Ardlie, K. G., L. Kruglyak, and M. Seielstad, 2002   Patterns of linkage disequilibrium in the human genome. Nat. Rev. Genet. 3: 299–309.

Backström, N., A. Qvarnström, L. Gustafsson, and H. Ellegren, 2006   Levels of linkage disequilibrium in a wild bird population. Biol. Lett. 2: 435–438.

Badke, Y. M., R. O. Bates, C. W. Ernst, C. Schwab, and J. P. Steibel, 2012   Estimation of linkage disequilibrium in four US pig breeds. BMC Genomics 13: 24.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme, 2004   *GENETIX 4.03, logiciel sous WindowsTM pour la génétique des populations. Laboratoire génome et populations*. Université de Montpellier II, Montpellier, France.

Belkhir, K., V. Castric, and F. Bonhomme, 2002   IDENTIX, a software to test for relatedness in a population using permutation methods. Mol. Ecol. Notes 2: 611–614.

Branca, A., T. D. Paape, P. Zhou, R. Briskine, A. D. Farmer *et al.*, 2011   Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. Proc. Natl. Acad. Sci. USA 108: E864–E870.

Bruneaux, M., S. E. Johnston, G. Herczeg, J. Merilä, C. R. Primmer *et al.*, 2013   Molecular evolutionary and population genomic analysis of the nine-spined stickleback using a modified restriction-site-associated DNA tag approach. Mol. Ecol. 22: 565–582.

Chapman, N. H., and E. M. Wijsman, 1998   Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. Am. J. Hum. Genet. 63: 1872–1885.

Charlesworth, B., 2009   Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10: 195–205.

Chen, T.-R., and H. M. Reisman, 1970   A comparative chromosome study of the North American species of sticklebacks (Teleostei: Gasterosteidae). Cytogenet. Genome Res. 9: 321–332.

Conrad, D. F., M. Jakobsson, G. Coop, X. Wen, J. D. Wall *et al.*, 2006   A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat. Genet. 38: 1251–1260.

Corbin, L. J., S. C. Blott, J. E. Swinburne, M. Vaudin, S. C. Bishop *et al.*, 2010   Linkage disequilibrium and historical effective population size in the Thoroughbred horse. Anim. Genet. 41: 8–15.

Cutter, A. D., 2006   Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. Genetics 172: 171–184.

DeFaveri, J., T. Shikano, N. I. A. Ghani, and J. Merilä, 2012   Contrasting population structures in two sympatric fishes in the Baltic Sea basin. Mar. Biol. 159: 1659–1672.

Delourme, R., C. Falentin, B. F. Fomeju, M. Boillot, G. Lassalle *et al.*, 2013   High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. BMC Genomics 14: 120.

Devlin, B., and N. Risch, 1995   A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29: 311–322.

Dickson, S. P., K. Wang, L. Krantz, H. Hakonarson, and D. B. Goldstein, 2010   Rare variants create synthetic genome-wide associations. PLoS Biol. 8: e1000294.

Ellegren, H., and B. C. Sheldon, 2008   Genetic basis of fitness differences in natural populations. Nature 452: 169–175.

Espigolan, R., F. Baldi, A. A. Boligon, F. R. P. Souza, D. G. M. Gordo *et al.*, 2013   Study of whole genome linkage disequilibrium in Nellore cattle. BMC Genomics 14: 305.

Evanno, G., S. Regnaut, and J. Goudet, 2005   Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611–2620.

Excoffier, L., A. Estoup, and J. M. Cornuet, 2005   Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genetics 169: 1727–1738.

Excoffier, L., and H. E. L. Lischer, 2010   Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10: 564–567.

Falush, D., M. Stephens, and J. K. Pritchard, 2003   Inference of population structure: extensions to linked loci and correlated allele frequencies. Genetics 164: 1567–1587.

Fang, D. D., L. L. Hinze, R. G. Percy, P. Li, D. Deng et al., 2013   A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (Gossypium hirsutum L.) cultivars from major cotton-growing countries. Euphytica 191: 391–401.

Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack et al., 2001   Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium. Am. J. Hum. Genet. 69: 831–843.

García-Gámez, E., G. Sahana, B. Gutiérrez-Gil, and J. J. Arranz, 2012   Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. BMC Genet. 13: 43.

García-Ramos, G., and M. Kirkpatrick, 1997   Genetic models of adaptation and gene flow in peripheral populations. Evolution 51: 21–28.

Garza, J. C., and E. G. Williamson, 2001   Detection of reduction in population size using data from microsatellite loci. Mol. Ecol. 10: 305–318.

Golding, G. B., and C. Strobeck, 1980   Linkage disequilibrium in a finite population that is partially selfing. Genetics 94: 777–789.

Gonda, A., K. Välimäki, G. Herczeg, and J. Merilä, 2012   Brain development and predation: plastic responses depend on evolutionary history. Biol. Lett. 26: 249–252.

Gorlov, I. P., O. Y. Gorlova, S. R. Sunyaev, M. R. Spitz, and C. I. Amos, 2008   Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. Am. J. Hum. Genet. 82: 100–112.

Goudet, J., 2002   FSTAT, a program to estimate and test gene diversities and fixation indices. Version 2.9.3.2. Available at: http://www2.unil.ch/popgen/softwares/fstat.htm. Accessed: August 19, 2014.

Gould, S. J., and R. F. Johnston, 1972   Geographic variation. Annu. Rev. Ecol. Syst. 3: 457–498.

Gratten, J., A. J. Wilson, A. F. McRae, D. Beraldi, P. M. Visscher et al., 2008   A localized negative genetic correlation constrains microevolution of coat color in wild sheep. Science 319: 318–320.

Hao, C., L. Wang, H. Ge, Y. Dong, and X. Zhang, 2011   Genetic diversity and linkage disequilibrium in Chinese bread wheat (Triticum aestivum L.) revealed by SSR markers. PLoS One 6: e17279.

Hayes, B. J., A. Gjuvsland, and S. Omholt, 2006   Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males. Heredity 97: 19–26.

Hedrick, P. W., 1987   Gametic disequilibrium measures: proceed with caution. Genetics 117: 331–341.

Herczeg, G., and K. Välimäki, 2011   Intraspecific variation in behaviour: effects of evolutionary history, ontogenetic experience and sex. J. Evol. Biol. 24: 2434–2444.

Herczeg, G., A. Gonda, and J. Merilä, 2009   Evolution of gigantism in nine-spined stickleback. Evolution 63: 3190–3200.

Herczeg, G., M. Turtiainen, and J. Merilä, 2010   Morphological divergence of North-European nine-spined sticklebacks (Pungitius pungitius): signatures of parallel evolution. Biol. J. Linn. Soc. Lond. 101: 403–416.

Hernandez, R. D., M. J. Hubisz, D. A. Wheeler, D. G. Smith, B. Ferguson et al., 2007   Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. Science 316: 240–243.

Hill, W. G., and A. Robertson, 1968   The effects of inbreeding at loci with heterozygote advantage. Genetics 60: 615–628.

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

Jorde, L. B., 2000   Linkage disequilibrium and the search for complex disease genes. Genome Res. 10: 1435–1444.

Kalinowski, S. T., 2005   HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol. Ecol. Notes 5: 187–189.

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

Karlin, S., and M. W. Feldman, 1970   Linkage and selection: two locus symmetric viability model. Theor. Popul. Biol. 1: 39–71.

Kawakami, T., N. Backström, R. Burri, A. Husby, P. Olason et al., 2014   Estimation of linkage disequilibrium and interspecific gene flow in Ficedula flycatchers by a newly developed 50k single-nucleotide polymorphism array. Mol. Ecol. Resour. Doi: 10.1111/1755-0998.12270.

Kruskal, W. H., and W. A. Wallis, 1952   Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47: 583–621.

Larsson, H., T. Källman, N. Gyllenstrand, and M. Lascoux, 2013   Distribution of long-range linkage disequilibrium and Tajima's D values in Scandinavian populations of Norway spruce (Picea abies). G3 (Bethesda) 3: 795–806.

Laurie, C. C., D. A. Nickerson, A. D. Anderson, B. S. Weir, R. J. Livingston et al., 2007   Linkage disequilibrium in wild mice. PLoS Genet. 3: e144.

Lewontin, R., 1964   The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49: 49–67.

Lewontin, R. C., and K. Kojima, 1960   The evolutionary dynamics of complex polymorphisms. Evolution 14: 458–472.

Li, M. H., and J. Merilä, 2010   Extensive linkage disequilibrium in a wild bird population. Heredity 104: 600–610.

Li, M.-H., and J. Merilä, 2011   Population differences in levels of linkage disequilibrium in the wild. Mol. Ecol. 20: 2916–2928.

Liu, K., and S. V. Muse, 2005   PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128–2129.

Luikart, G., F. W. Allendorf, J. M. Cornuet, and W. B. Sherwin, 1998   Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. 89: 238–247.

Mann, H. B., and D. R. Whitney, 1947   On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat. 18: 50–60.

Mattern, M. Y., 2004   Molecular phylogeny of the Gasterosteidae: the importance of using multiple genes. Mol. Phylogenet. Evol. 30: 366–377.

Mattern, M. Y., and D. A. McLennan, 2004   Total evidence phylogeny of Gasterosteidae: combining molecular, morphological and behavioural data. Cladistics 20: 14–22.

Merilä, J., 2013   Nine-spined stickleback (Pungitius pungitius): an emerging model for evolutionary biology research. Ann. N. Y. Acad. Sci. 1289: 18–35.

Miller, J. M., J. Poissant, J. W. Kijas, and D. W. Coltman The International Sheep Genomics Consortium, 2011   A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Mol. Ecol. Resour. 11: 314–322.

Mobley, K. B., D. Lussetti, F. Johansson, G. Englund, and F. Bokma, 2011   Morphological and genetic divergence in Swedish postglacial stickleback (Pungitius pungitius) populations. BMC Evol. Biol. 11: 287.

Mukai, T., L. E. Mettler, and S. I. Chigusa, 1971   Linkage disequilibrium in a local population of Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 68: 1065–1069.

Nei, M., and W. H. Li, 1973   Linkage disequilibrium in subdivided populations. Genetics 75: 213–219.

Nordborg, M., and S. Tavaré, 2002   Linkage disequilibrium: what history has to tell us. Trends Genet. 18: 83–90.

Nordborg, M., J. O. Borevitz, J. Bergelson, C. C. Berry, J. Chory et al., 2002   The extent of linkage disequilibrium in Arabidopsis thaliana. Nat. Genet. 30: 190–193.

Peakall, R., and P. E. Smouse, 2006   GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6: 288–295.

Peakall, R., and P. E. Smouse, 2012   GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28: 2537–2539.

Piry, S., G. Luikart, and J. M. Cornuet, 1999   Bottleneck: a computer program for detecting recent reductions in the effective size using allele frequency data. J. Hered. 90: 502–503.

Pritchard, J. K., 2001   Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. 69: 124–137.

Pritchard, J. K., and M. Przeworski, 2001   Linkage disequilibrium in humans: Models and data. Am. J. Hum. Genet. 69: 1–14.

Pritchard, J. K., M. Stephens, and P. Donnelly, 2000   Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Purcell, S. M., N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan *et al.*, 2009   Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460: 748–752.

Queller, D. C., and K. F. Goodnight, 1989   Estimating relatedness using genetic markers. Evolution 43: 258–275.

Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti *et al.*, 2001   Linkage disequilibrium in the human genome. Nature 411: 199–204.

Rice, W. R., 1989   Analyzing tables of statistical tests. Evolution 43: 223–225.

Roesti, M., D. Moser, and D. Berner, 2013   Recombination in the threespine stickleback genome—patterns and consequences. Mol. Ecol. 22: 3014–3027.

Service, S., J. DeYoung, M. Karayiorgou, J. L. Roos, H. Pretorious *et al.*, 2006   Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat. Genet. 38: 556–560.

Shapiro, M. D., B. R. Summers, S. Balabhadra, J. T. Aldenhoven, A. L. Miller *et al.*, 2009   The genetic architecture of skeletal convergence and sex determination in ninespine sticklebacks. Curr. Biol. 19: 1140–1145.

Shifman, S., J. Kuypers, M. Kokoris, B. Yakir, and A. Darvasi, 2003   Linkage disequilibrium patterns of the human genome across populations. Hum. Mol. Genet. 12: 771–776.

Shikano, T., Y. Shimada, G. Herczeg, and J. Merilä, 2010a   History *vs.* habitat type: explaining the genetic structure of European nine-spined stickleback (*Pungitius pungitius*) populations. Mol. Ecol. 19: 1147–1161.

Shikano, T., J. Ramadevi, and J. Merilä, 2010b   Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). Mol. Biol. Evol. 27: 2775–2789.

Shikano, T., J. Ramadevi, Y. Shimada, and J. Merilä, 2010c   Utility of sequenced genomes for microsatellite marker development in non-model organisms: a case study of functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). BMC Genomics 11: 334.

Shikano, T., V. N. Laine, G. Herczeg, J. Vilkki, and J. Merilä, 2013   Genetic architecture of parallel pelvic reduction in ninespine sticklebacks. G3 (Bethesda) 3: 1833–1842.

Shimoda, N., E. W. Knapik, J. Ziniti, C. Sim, E. Yamada *et al.*, 1999   Zebrafish genetic map with 2000 microsatellite markers. Genomics 58: 219–232.

Slate, J., 2005   Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Mol. Ecol. 14: 363–379.

Slate, J., 2013   From beavis to beak colour: a simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. Evolution 67: 1251–1262.

Slate, J., J. Gratten, D. Beraldi, J. Stapley, M. Hale *et al.*, 2009   Gene mapping in the wild with SNPs: guidelines and future directions. Genetica 136: 97–107.

Slatkin, M., 2008   Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9: 477–485.

Stephens, M., and P. Scheet, 2005   Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. Am. J. Hum. Genet. 76: 449–462.

Stephens, M., N. J. Smith, and P. Donnelly, 2001   A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68: 978–989.

Sutter, N. B., M. A. Eberle, H. G. Parker, B. J. Pullar, E. F. Kirkness *et al.*, 2004   Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. Genome Res. 14: 2388–2396.

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

Teacher, A. G. F., T. Shikano, M. E. Karjalainen, and J. Merilä, 2011   Phylogeography and genetic structuring of European nine-spined sticklebacks (*Pungitius pungitius*)—mitochondrial DNA evidence. PLoS ONE 6: e19476.

Teare, M. D., A. M. Dunning, F. Durocher, G. Rennart, and D. F. Easton, 2002   Sampling distribution of summary linkage disequilibrium measures. Ann. Hum. Genet. 66: 223–233.

Van Inghelandt, D., J. C. Reif, B. S. Dhillon, P. Flament, and A. E. Melchinger, 2011   Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor. Appl. Genet. 123: 11–20.

Waples, R. S., 2006   A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conserv. Genet. 7: 167–184.

Watterson, G. A., and H. A. Guess, 1977   Is the most frequent allele the oldest? Theor. Popul. Biol. 11: 141–160.

Weir, S. B., 1996   Disequilibrium, pp. 91–138 in *Genetic Data Analysis II*. Sinauer Associates, Inc., Sunderland, MA.

Williamson-Natesan, E. G., 2005   Comparison of methods for detecting bottlenecks from microsatellite loci. Conserv. Genet. 6: 551–562.

Wootton, R. J., 1976   *The Biology of the Sticklebacks*. Academic Press, New York.

Wray, N. R., S. M. Purcell, and P. M. Visscher, 2011   Synthetic associations created by rare variants do not explain most GWAS results. PLoS Biol. 9: e1000579.

Zaykin, D. V., A. Pudovkin, and B. S. Weir, 2008   Correlation-based inference for linkage disequilibrium with multiple alleles. Genetics 180: 533–545.

Zhang, W., A. Collins, J. Gibson, W. J. Tapper, S. Hunt *et al.*, 2004   Impact of population structure, effective bottleneck time, and allele frequency on linkage disequilibrium maps. Proc. Natl. Acad. Sci. USA 101: 18075–18080.

Zondervan, K. T., and L. R. Cardon, 2004   The complex interplay among factors that influence allelic association. Nat. Rev. Genet. 5: 89–100.

Communicating editor: Y. Kim