Does eutrophication cause directional genetic selection in three-spined stickleback (Gasterosteus aculeatus)?

- A study of multiple Baltic Sea populations.


Abstract

Human-induced eutrophication is indirectly affecting aquatic organisms by altering their environment. This brings on altered selective pressures and could thereby cause changes in the genetic composition of exposed populations. Since anthropogenic environmental changes are usually occurring at a much higher rate than naturally occurring changes, they force populations to adapt to the new conditions faster than normal. Here, I have studied populations of three-spined sticklebacks (Gasterosteus aculeatus) from four eutrophicated and four adjacent reference sites, along the coast of Finland, to investigate if this species has responded genetically to the human-induced eutrophication of the Baltic Sea. For this purpose I used amplified fragment length polymorphism (AFLP) and found distinctions in genetic composition between the two habitats, as well as similarities between populations from eutrophicated sites. This suggests a similar genetic response to eutrophicated conditions by stickleback populations from different geographical areas. Moreover I found a distinct geographic structure among three-spined sticklebacks in the Baltic Sea.

Introduction

In the last centuries, human activities have been altering natural habitats at an accelerating rate, and presently the changes are faster than ever (Tuomainen & Candolin, 2011). This could affect the evolutionary processes by altering and increasing the selective pressures on populations (Heuschele et al., 2009). The ongoing human-induced eutrophication is an example of an environmental change that is altering the living conditions for aquatic organisms worldwide. Consequences that have been linked to the over-enrichment of nutrients are increased fluctuations in dissolved oxygen levels and pH due to increased growth and decomposition of primary producers (Pollock et al., 2007; Reddy, 1981; Larsson, 1985). Other consequences are increased water turbidity, changed habitat structure and decreased visibility for the associated fauna (Nielsen et al., 2002; Sandén & Håkansson, 1996; Bonsdorff et al., 2002).

The genetic structure of populations is created by the evolutionary processes of gene flow, mutation, random genetic drift and natural selection. While mutations, genetic drift and divergent natural selection drive populations genetically apart, gene flow retains similarities between them (Freeman & Herron, 2007). According to the model of isolation by distance, populations living close to each other should, because of a higher rate of gene flow, have a more similar genetic composition than geographically distant populations of the same species (Wright, 1943). However, if geographically distant populations are living in areas with similar environmental conditions they should experience similar selective pressures. This could lead to the same adaptations among these populations and therefore result in a more similar genetic composition than expected by the model of isolation by distance. Natural selection can therefore act the opposite way, by reducing genetic distance between populations (Lind & Grahn, 2011). In contrast to a previously well established view, that evolutionary
and ecological times differ in orders of magnitude, studies have shown that the evolutionary response of populations can be fast (Thompson, 1998). A definition of rapid evolution was formulated by Hairston et al. (2005) as “a genetic change occurring rapidly enough to have a measurable impact on simultaneous ecological change”. One example is Daphnia galeata in the Lake Constance, which in less than ten years evolved a greater ability to grow on the low quality food cyanobacteria, which became more abundant due to eutrophication (Hairston et al., 1999, 2001). More, guppies that were introduced to a stream with a smaller predator in 11 years evolved similar life-history traits as guppies that had coexisted with the smaller predator for a long time (Reznick et al., 1990). And Pereyra et al. (2009) estimated that the brown macroalga Fucus radicans, endemic to the Baltic Sea, diverged from Fucus vesiculosus about 400 years ago and has since developed into a new species. The most rapid evolution occurs in large populations with ample genetic variation, under strong selective pressure (Stearns & Hoekstra, 2005). For a species to be able to survive and reproduce in a new environment, it is assumed that a high degree of phenotypic plasticity is needed. This will give room for subsequent genetic adaptations to the novel conditions (Candolin, 2009).

The Baltic Sea is a small, semi-enclosed sea with a large catchment area and low annual mean temperature. These characteristics result in a long residence time of water and slow degradation processes, which makes it susceptible to eutrophication. The low salinity (brackish water) also leads to a low biodiversity (http://www.helcom.fi/environment2/nature/en_GB/nature/). The three-spined stickleback (Gasterosteus aculeatus) is an example of an ecologically flexible species that lives in the Baltic Sea. It shows a high degree of phenotypic plasticity in a range of traits, such as morphology (Day et al., 1994; Sharpe et al., 2008), physiology (Schaarschmidt et al., 1999), life-history traits (Baker & Foster, 2002) and resource use (Svanbäck & Bolnick, 2007), and has a distribution across the entire northern hemisphere. Increased growth of primary producers due to eutrophication has been shown to have complex effects on reproductive behaviors and sexual selection in this species (Candolin et al., 2007). One study found that increased water turbidity - caused by an increased phytoplankton growth - induce selection for intense courtship activity in stickleback males (Engström-Öst & Candolin, 2007). On the other hand, another study showed that in dense vegetation (another effect of eutrophication), selection for male courtship behavior and red nuptial color is relaxed, despite an increased effort in male courtship activity and increased female mate choice, since these characters do not enhance the attractiveness of males in this habitat (Candolin et al., 2007). Moreover, increased vegetation, water turbidity and macroalgae cover have a positive impact on the sticklebacks as it reduce predation risk (Candolin & Voigt, 1998) as well as aggressiveness between nesting males, improve their parenting ability and enhance the reproductive output. In turbid water, there is also a reduced selection for large male size, since more males are successful in reproducing (Candolin et al., 2008). During daytime the increased photosynthesis, which is a result of increased growth of primary producers, leads to an augmented pH value, making male olfactory cues more important in female mate choice (Heuschele & Candolin, 2007).

Determination of the evolutionary responses of populations to human-induced environmental change, and their effects, is of crucial importance, to make prediction of the future of both populations and ecosystems (Candolin & Heuschele, 2008). This study aims to investigate if the large scale human-induced eutrophication of the Baltic Sea is causing directional selection in the three-spined stickleback. For this purpose I will use amplified fragment length polymorphism (AFLP) to compare allele frequencies and genetic variation within and between populations of three-spined stickleback from four eutrophicated and four adjacent reference sites. My hypothesis is that populations living close to each other will have more similar genetic composition due to a more recent common
evolutionary history and a higher rate of gene flow, and that eutrophication cause similar selective pressures in populations at different geographical sites leading to a more similar genetic composition. Further, I will screen for $F_{ST}$-outlier loci that are assumed to be associated to genes that are coding for traits under directional selection. Since eutrophication seems to reduce predation risk and relax sexual selection in the three-spined stickleback, I want to investigate if the genetic diversity within populations from eutrophicated sites is maintained, despite the assumed increased selection on some alleles or haplotypes in eutrophicated habitat.

**Methods**

**Sampling**

Tail fin clips from a total of 233 three-spined sticklebacks were sampled by Ulla Tuomainen (University of Helsinki, Finland), from four eutrophicated and four reference sites in the Baltic Sea, along the coast of Finland (see Fig. 1 & Table 1). The sites were chosen based on long-term data provided by the Finnish Environment Institute and on the ecological classification of surface waters (pers. comm. Ulla Tuomainen).

![Fig. 1 Sampling sites along the coast of Finland in the Baltic Sea. From each of the four sites - Raahe, Turku, Tvärmne and Kotka – samples were taken from three-spined sticklebacks from both eutrophicated (Raa-E, Tur-E, Tva-E and Kot-E) and reference (Raa-C, Tur-C, Tva-C and Kot-C) conditions. All sites were chosen based on long-term environmental data and on ecological classification of the surface water.](image-url)
Table 1 Sampling site, name of population (Pop), habitat (Reference/Eutrophicated), number of sampled individuals (N), Coordinates for sampling site, gene diversity (Hj) also known as Nei’s gene diversity, derived from AFLP-Surv, with standard error (S.E).

<table>
<thead>
<tr>
<th>Sampling site</th>
<th>Pop</th>
<th>Habitat</th>
<th>N</th>
<th>Coordinates</th>
<th>Hj ± S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raah</td>
<td>Raah-C</td>
<td>Reference</td>
<td>36</td>
<td>64°36’42.36”N 24°22’44.63”E</td>
<td>0.14426 ± 0.01410</td>
</tr>
<tr>
<td>Raah</td>
<td>Raah-E</td>
<td>Eutrophicated</td>
<td>34</td>
<td>64°40’44.33”N 24°27’5.98”E</td>
<td>0.15867 ± 0.01504</td>
</tr>
<tr>
<td>Turku</td>
<td>Turku-C</td>
<td>Reference</td>
<td>16</td>
<td>60°10’46.23”N 21°38’19.63”E</td>
<td>0.17982 ± 0.01554</td>
</tr>
<tr>
<td>Turku</td>
<td>Turku-E</td>
<td>Eutrophicated</td>
<td>34</td>
<td>60°9’53.95”N 21°42’45.83”E</td>
<td>0.16450 ± 0.01465</td>
</tr>
<tr>
<td>Tvärminne</td>
<td>Tvärminne-C</td>
<td>Reference</td>
<td>33</td>
<td>59°49’41.46”N 23°12’33.21”E</td>
<td>0.18194 ± 0.01405</td>
</tr>
<tr>
<td>Kotka</td>
<td>Kotka-C</td>
<td>Reference</td>
<td>34</td>
<td>60°26’23.68”N 26°50’56.45”E</td>
<td>0.16981 ± 0.01490</td>
</tr>
<tr>
<td>Kotka</td>
<td>Kotka-E</td>
<td>Eutrophicated</td>
<td>17</td>
<td>60°27’1.70”N 26°56’57.11”E</td>
<td>0.18171 ± 0.01554</td>
</tr>
</tbody>
</table>

**DNA-extraction**

Total genomic DNA was extracted from a 2 x 2 mm tail fin clip according to Laird et al. (1991) with some modifications. The fin was incubated for four hours in 56°C, together with 100µl lysis buffer (10µl 1M TRIS-HCL pH 8.5; 1µl 0.5M EDTA; 2µl 10% SDS; 4µl 5M NaCl and ddH2O) and 2µl Proteinase K 18.9 mg/ml (Fermentas AB), after which the sample was centrifuged. DNA from the resulting supernatant was precipitated with 200µl freezing EtOH 99.7% and 20µl NaAc 3M and the aggregated precipitation was washed in 100µl EtOH 70%. The sample was air-dried before resolved in 50µl 1xTE-buffer. NanoDrop spectrophotometer (ND-1000 V3.2.1) was used to measure the DNA concentration and thereafter the samples were diluted to 50ng/µl with ddH2O.

**Generation of AFLP markers**

Amplified fragment length polymorphism (AFLP) is a fingerprinting technique which is based on the selective amplification of DNA fragments. It is a nondestructive method which is also relatively cheap and time-saving and it generates a large number of dominant markers from multiple regions in the genome. The high resolution makes it possible to detect fine scale genetic differences between groups (Bensch & Åkesson, 2005). Genomic DNA is first cut with specific restriction enzymes, followed by ligation of double stranded oligonucleotide adapters to the ends of the resulting fragments. The selective PCR amplification is achieved by using single stranded primers that are not only corresponding to the nucleotide sequences of the adapters, but also extending into the actual restriction fragments. Depending on the sequences and number of the selective nucleotides of the primers, different numbers of AFLP markers will be generated. AFLP can be used without any prior sequence knowledge of the organism being analyzed (Vos et al. 1995).

The generation of AFLP markers was achieved by following the protocol by Vos et al. (1995), with minor modifications, as described by Bensch et al. (2002). 10µl of DNA sample was incubated for 1 hour at 37°C together with 10µl digestion cocktail consisting of 2.5U EcoR1 (5’-G1JAATTC-3’), 2.5U TruI (5’-T↓TAA-3’) (Fermentas AB), 1µg BSA, 2µl 10xTA-buffer and ddH2O. For the ligation of adapters to the digested genomic DNA, 5µl cocktail, consisting of 0.5µl 10x-ligation buffer, 2.5pmol E-adapter (5’-CTCGTAGACTCGTACC-3’ and 3’-CATCTGACGCATGGTAA-5’), 0.025mmol M-adapter (5’-GACGATGATCGTCGAGT-3’ and 3’-TACTCAGGACTCAT-5’) (eurofins mwg/operon), 0.5U T4-ligase (Fermentas AB) and ddH2O, was added. Incubation then continued for another 3 hours.
before the sample was diluted 10 times with ddH2O and stored at -20°C. A pre-amplification step was performed to increase the accuracy of the amplification of AFLP markers. 10 μl of DNA template was added to a cocktail consisting of 6pmol E-primer (5'-GACTGCGTACCAATTCTC-3') with an additional T at the 3’ end, 6pmol M-primer (5'-GATGAGTCCTGAGTAAN-3') with an additional C at the 3’ end (eurofins mwg/operon), 0.8μg BSA, 0.05μmol MgCl2, 2μl 10x PCR-buffer, 4nmol dNTPs, 0.4U AmpliTaq® DNA Polymerase (Roche) and ddH2O to a total sample volume of 20μl. The temperature profile for the PCR was set on [94°C 2 min] + [94°C 30s, 56°C 30s, 72°C 60s] * 20 cycles + [72°C 10 min] and the PCR products were diluted 10 times with ddH2O before storing them at -20°C. In the selective amplification step, two different primer combinations were used. 2.5μl of the DNA template was mixed with 6pmol E-primer (5'-GACTGCGTACCAATTCTC-3') with the three additional bases TCT or TAG at the 3’ end, 6pmol M-primer (5'-GATGAGTCCTGAGTAAN-3') with the three additional bases CAC at the 3’ end (eurofins mwg/operon), 0.4μg BSA, 0.025μmol MgCl2, 1μl 10x PCR-buffer, 2nmol dNTPs, 0.4U AmpliTaq® DNA Polymerase (Roche) and ddH2O to a total volume of 7.5μl. The temperature profile was set on [94°C 2 min] + [94°C 30s, 65-0.7°C/cycle 30s, 72°C 60s] * 12 cycles + [94°C 30s, 56°C 30s, 72°C 60s] * 23 cycles + [72°C 10 min]. Samples were loaded on 96-well PCR plates, sealed with parafilm and sent to Uppsala Genome Center for separation of the AFLP markers through capillary electrophoresis with ABI3730XL DNA Analyzer.

**Statistical analyses**

Gene marker® v. 2.2.0 Demo (Softgenetics LLC®) was employed to visually evaluate the quality of the individual samples, whereupon defective samples were discarded. The final genotype data matrix was generated using Genemapper® software version 4.0 (Applied Biosystems), with analysis range settings at 150-500 bp and bin width 1 bp. The peak amplitude threshold was set to 200 RFU. To estimate the error rates within the data matrix a duplicate analysis was performed with 67 individuals from primer combination Ectc-Mcac and 6 individuals from primer combination Etag-Mcac. To estimate the gene diversity within populations as well as between habitats and to calculate the pairwise FST-values between populations, AFLP-Surv 1.0 (Vekemans, 2002) was used, using the approach of Lynch and Milligan (1994), assuming Hardy-Weinberg equilibrium and using the Bayesian method with a non-uniform prior distribution of allele frequencies and 1000 permutations to test the significance of the FST. Genetic distances (FST) and geographical distances (waterway) between sites (see Table. 2) were log-transformed and combined to test for isolation by distance using a mantel test with Isolation by Distance Web Service version 3.16 (Jensen et al. 2005). An ANOVA was performed in R commander (Rcmdr) 1.7-3 (R Developmental Core Team, 2007) to test for differences in gene diversity between sites as well as between habitats.

STRUCTURE 2.3.2.1 (Earl, 2009) was used to group the individual samples into clusters, based on their genotypes, and to find the most probable number of clusters when under Hardy-Weinberg equilibrium, burnin was set to 50000 with 70000 additional cycles. Each run was iterated 3 times, and number of clusters (K) set from 1-10, assuming admixture and uncorrelated allele frequencies. The number of clusters was calculated and output files visualized by STRUCTURE HARVESTER (Earl & von Holdt, 2011). Further testing of differences in population genetic structure was done with a constrained principal coordinate analysis (CoPCoA) performed in the Vegan package (Oksanen et al., 2011) in R 2.14.0 (R Developmental Core Team, 2007), as well as a constrained principal coordinate analysis, where population was conditioned to remove the raw effect of geography (Conditioned CoPCoA). To test for a genetic divergence between the two habitat groups a locus-by-locus Analysis of Molecular Variance (AMOVA) was performed in Arlequin Ver. 3.5 (Excoffier, 2005), which
assumes that the loci are unlinked and adjust sample sizes for each locus. An $F_{ST}$-outlier loci analysis was also performed in Arlequin (Excoffier, 2009), in which $F_{ST}$ values for each locus were tested to find $F_{ST}$-outlier loci with higher $F_{ST}$ than expected under Hardy-Weinberg equilibrium. The eight populations were grouped into four groups with eutrophicated-reference population pairs (Raa-C/Raa-E, Tur-C/Tur-E, Tva-C/Tva-E, Kot-C/Kot-E) and 50 groups with 100 demes each for 50000 runs was simulated. To test which loci contributed most to the genotypic differences between habitats, individuals were grouped according to habitat (i.e. eutrophicated or reference), and a Random Forest classification (Breiman, 2001) was performed in R 2.14.0 (R Developmental Core Team, 2007). After removal of identified $F_{ST}$-outlier loci the Conditioned CoPCoA and the locus-by-locus AMOVA were repeated to test if the distinction between habitats would disappear.

**Results**

The final data matrix consisted of 219 individual samples and 140 loci - 61 loci derived from Etet-Mcac and 79 loci from Etag-Mcac. The number of individuals from each of the eight sites ranged from 15 to 37 (see Table 1). 14% of Etet-Mcac loci and 33% of Etag-Mcac loci showed more than 10% mismatches between duplicates.

Results from AFLP-Surv analyses showed a distinct genetic structure where 15% of the total genetic variation was appointed among populations ($F_{ST} = 0.15, P < 0.01$). The genetic variation between populations in eutrophicated sites was 16% ($F_{ST} = 0.16, P < 0.01$) and 17% between populations in reference sites ($F_{ST} = 0.17, P < 0.01$). No significant differences in mean within-population gene diversity between geographical sites were found with the ANOVA (DF = 3, F-value = 4.6341, P = 0.1200). Neither was there any significant differences in mean within-population gene diversity when comparing habitats (DF = 1, F-value = 0.6797, P = 0.4701; gene diversity $Hw$ ± S.E.: eutrophicated 0.1751 ± 0.00841; reference 0.1690 ± 0.00865), but all pairwise $F_{ST}$-values were significant except for those between Kot-E/Tur-C and Tur-C/Tur-E (see Table 2).

<table>
<thead>
<tr>
<th>Raa-C</th>
<th>Raa-E</th>
<th>Tur-C</th>
<th>Tur-E</th>
<th>Tva-C</th>
<th>Tva-E</th>
<th>Kot-C</th>
<th>Kot-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raa-C</td>
<td></td>
<td>0.009*</td>
<td>0.0358**</td>
<td>0.0158**</td>
<td>0.2971**</td>
<td>0.2853**</td>
<td>0.2562**</td>
</tr>
<tr>
<td>Raa-E</td>
<td>9.2</td>
<td></td>
<td>0.0319**</td>
<td>0.0144**</td>
<td>0.2998**</td>
<td>0.2798**</td>
<td>0.2508**</td>
</tr>
<tr>
<td>Tur-C</td>
<td>598.1</td>
<td>607.3</td>
<td></td>
<td>0.0051**</td>
<td>0.2161**</td>
<td>0.2147**</td>
<td>0.1631**</td>
</tr>
<tr>
<td>Tur-E</td>
<td>593</td>
<td>602.2</td>
<td>5.1</td>
<td></td>
<td>0.2544**</td>
<td>0.2445**</td>
<td>0.2031**</td>
</tr>
<tr>
<td>Tva-C</td>
<td>704.8</td>
<td>714</td>
<td>106.7</td>
<td>111.8</td>
<td></td>
<td>0.0409**</td>
<td>0.0259**</td>
</tr>
<tr>
<td>Tva-E</td>
<td>701.5</td>
<td>710.7</td>
<td>103.4</td>
<td>108.5</td>
<td>3.31</td>
<td></td>
<td>0.0768**</td>
</tr>
<tr>
<td>Kot-C</td>
<td>935.6</td>
<td>944.8</td>
<td>317.5</td>
<td>322.6</td>
<td>326.2</td>
<td>216.6</td>
<td></td>
</tr>
<tr>
<td>Kot-E</td>
<td>925.6</td>
<td>934.8</td>
<td>327.5</td>
<td>327.5</td>
<td>326.2</td>
<td>226.6</td>
<td>224</td>
</tr>
</tbody>
</table>

Table 2 Under the diagonal are the shortest waterway distances between all sites in kilometers and over the diagonal are the Pairwise $F_{ST}$-values between populations (Phylip format) obtained from AFLP-Surv. $P < 0.05^*$, $P < 0.01^{**}$. The two $F_{ST}$-values marked with red are nonsignificant.

The Isolation by Distance test showed that the $F_{ST}$-values increased with geographical distance ($r^2 = 0.204; P = 0.0154$) and the CoPCoA showed that most of the genetic variation within the data matrix was due to geography (DF = 7, F-value = 9.9558, $P < 0.005$) (see Fig. 2). The proportion explained by the first axis was 20.4%, and the second axis 1.69%. When population was conditioned, the genotypes separated according to habitat (DF = 1, F-value = 8.3955, $P < 0.005$) (Fig. 5a & b). Proportion explained by the first axis was 3.7%, and the second axis 21.1%.
Constrained principal coordinate analysis (CoPCoA), based on 140 amplified fragment length polymorphism (AFLP) loci from 219 individuals and eight populations – four from eutrophicated and four from reference sites, showing the distribution of both individuals and population centroids (DF = 7, F-value = 9.9558, P < 0.005). The first axis (CAP1), explains 20.4% of the variation and the second axis (CAP2) explains 1.69%.

STRUCTURE results showed that the most probable number of genetic clusters were two (K = 2; Mean LnP = -12150; SD = 4.05) (Fig. 3). All individuals from the Raahé-populations and 94% and 97% of the individuals from Tur-C and Tur-E respectively, were assigned to Cluster1. All individuals from Tva-C and -E and Kot-C were assigned to Cluster2, as were 69% of Kot-E individuals.

The locus-by-locus AMOVA showed that the two habitats, when comparing population pairs from the four geographical sites (Kotka, Raahé, Turku & Tvärminne), were significantly separated by 3.7% of the total genetic variation. 20.6% of the variation was between the four geographical sites (F<sub>ST</sub> = 0.24325, P < 0.0001, 1023 permutations). Eight F<sub>ST</sub>-outlier loci were detected with the F<sub>ST</sub>-outlier loci analysis. Seven of these and two additional loci were detected with the Random Forest classification (Fig. 4a & b). After removal of these loci, the amount of variation between habitats had decreased to 2.6% and to 12.3% among sites according to the locus-by-locus AMOVA. The Conditioned CoPCoA, after removal of F<sub>ST</sub>-outlier loci, still showed some distinction between habitats, even though somewhat lessened (DF = 1, F-value = 8.3955, P < 0.005) (Fig. 5c).
Fig. 4a) $F_{ST}$-outlier loci analysis in Arlequin between eutrophicated and reference population pairs from four geographical sites. 140 AFLP markers from two primer combinations – Ectc-Mcac and Etag-Mcac – were used. Blue ($P < 0.05$) and red ($P < 0.01$) dots indicate loci with significantly higher or lower $F_{ST}$-values than expected under Hardy-Weinberg equilibrium and are assumed to be under directional (higher $F_{ST}$-values) and stabilizing (lower $F_{ST}$-values) selection. Loci with attached names were singled out and removed in subsequent analyses.

4b) Result from Random Forest classification showing the 20 loci of highest importance for separating individuals from eutrophicated and reference sites when they were ordered in the two habitat groups. All $F_{ST}$-outlier loci from the analysis in Arlequin are present here except for Etag-Mcac97. Two red stars indicate loci at or over the 99% confidence interval level ($P \leq 0.01$) and one blue star indicate loci at or over the 95% confidence interval level ($P \leq 0.05$) in the $F_{ST}$-outlier loci analysis in Arlequin. These were singled out and removed in subsequent analyses, as were the two loci with surrounding black squares.
Fig. 5a) A constrained principal coordinate analysis, where the genetic variation caused by geography was removed by conditioning population (Conditioned CoPCoA), shows that the samples can be separated due to habitat. The first axis explains 3.74% and the second axis explains 21.1% of the variation. b) When adding population identity to the samples, it becomes visible that the populations mainly divide into two clusters, with individuals from Kotka and Tvärminne in one and individuals from Raathe and Turku in the other. c) After removal of 10 $F_{ST}$-outlier loci some of the distinction between habitats disappeared.

**Discussion**

The results from this study suggest that the three-spined sticklebacks in the Baltic Sea have responded genetically to human-induced eutrophication. This was expected and corresponds to a similar study of three-spined sticklebacks in pulp mill polluted areas of the Baltic Sea (Lind & Grahn, 2011). Despite the fact that sticklebacks live in an open environment (a sea) and that the eutrophicated-reference population pairs were located very close to each other, approximately 3-10 kilometers, three out of four population pairs had significant $F_{ST}$-values (Table 2). The locus-by-locus AMOVA further showed that some of the total genetic variation was between habitats when analyzing population pairs. This suggests different selective pressures between them, which is in concordance with multiple ethological studies of Baltic Sea populations of three-spined sticklebacks in eutrophicated conditions (Candolin et al., 2007; Engström-Öst & Candolin, 2007; Candolin & Voigt, 1998; Candolin et al., 2008; Heuschele...
& Candolin, 2007). In addition to these results, the Conditioned CoPCoA (Fig. 5a) showed that there were enough differences in genetic variation between eutrophicated and reference sites to separate the two habitat groups. This point toward a diverging natural selection between habitats and a similar genetic response in populations from eutrophicated sites to the new environmental conditions. Moreover, the \( F_{ST} \)-outlier loci analysis identified eight out of 140 loci, assumed to be under directional selection, i.e. loci with significantly higher \( F_{ST} \)-values than expected between habitats when under Hardy-Weinberg equilibrium (Fig. 4a). The incidence of loci indicated to be under directional selection in this study - 5.7% in the \( F_{ST} \)-outlier loci analysis - is similar to the proportion reported in earlier studies of a range of species where random molecular markers, including AFLP, were used (Shimada et al., 2011 and references therein). For three-spined sticklebacks in pulp mill polluted sites in the Baltic Sea, Lind & Grahn (2011) found a proportion of 8.4% of loci indicated to be under directional selection. The loci identified in the \( F_{ST} \)-outlier loci analysis were largely the same as the ones found in the Random Forest classification (Fig. 4b), which increases the likelihood of correct identification of loci under divergent selection. Occurrence of \( F_{ST} \) outlier loci is a straightforward indication of different selection between groups, since the specific differences in genetic composition are pinpointed. Some of the distinction between habitat groups remained however, after 10 \( F_{ST} \)-outlier loci were removed from the analysis (Fig. 5c), suggesting a more complex effect of eutrophication on genetic composition.

The gene diversity did not differ between the two habitat groups. This result was not surprising since Candolin et al. (2008) found that the consequences of eutrophication relaxes sexual selection and increase the reproductive output of three-spined sticklebacks. A higher proportion of successfully reproducing individuals increase the possibility of maintaining genetic diversity within populations in eutrophicated conditions, despite altered and/or increased selective pressures. Also, Candolin & Voigt (1998) found that predation risk was reduced due to eutrophication, which could have further positive effects on the genetic diversity within stickleback populations.

In accordance with Lind & Grahn, (2011), I found a distinct geographic structure in the three-spined sticklebacks of the Baltic Sea using AFLP. The result is however in contrast with earlier studies, performed with microsatellite data. Cano et al., (2008) found, for instance, that coastal populations of the Baltic Sea were genetically relatively uniform (\( F_{ST} = 0.003 \)), and no isolation by distance pattern was found among Baltic Sea populations by Mäkinen et al. (2006). When testing all eight populations together, I found a significant isolation by distance pattern and most of the genetic variation in the data matrix was explained by geography (Fig. 2). In addition to the general isolation by distance pattern, the CoPCoA (Fig. 2), the Conditioned CoPCoA (Fig. 4b), as well as the cluster analysis in STRUCTURE (Fig. 3), showed a further geographical division into two genetic clusters with individuals from Raahe and Turku mainly in one and individuals from Tvärminne and Kotka mainly in the other (see Fig. 1). This indicates that a factor other than distance (e.g. biogeographical) is causing a distinction between the two cluster-groups. At the same time there was only a weak or no genetic structure between Tur-C, Tur-E and Kot-E suggesting a sufficient gene flow between these populations to erase genetic differentiation.

That eutrophication is causing changes in the genetic composition of three-spined stickleback populations has been shown in this study. Since the human-induced eutrophication of the Baltic Sea started about 200 years ago (http://www.helcom.fi/environment2/eutrophication/en_GB/front/) the genetic response has been relatively fast in evolutionary terms. More, it is possible that the populations in eutrophicated conditions are still in the process of adaptation and will continue to diverge from
populations in non-eutrophicated conditions. If this is the case, continued genetic changes are likely to further influence interactions with other species, such as competitors and predators. For now, it seems that the three-spined stickleback is handling the eutrophication of the Baltic Sea well, due to a high degree of plasticity (Day et al., 1994; Sharpe et al., 2008; Schaarschmidt et al., 1999; Baker & Foster, 2002; Svanbäck & Bolnick, 2007) and evolutionary potential (Taylor & McPhail, 1986; Law & Blake, 1996; Bergström, 2002; Schaarschmidt & Jurss, 2003; Blake et al., 2005; McPhail, 1977; Snyder, 1991; Nagel & Schluter, 1998; Write et al., 2004; Walker & Bell, 2000; Leinonen et al., 2006; Sharpe et al., 2008; Aguirre, 2009, Lind & Grahn, 2011; Mäkinen et al., 2006). If eutrophication of aquatic environments will continue, this species could therefore become a strong competitor to other species with a poorer ability to handle environmental changes.

More research is needed for a better understanding of the consequences of human-induced eutrophication on populations. Additional sites and primer combinations, as well as temporal studies of multiple populations of three-spined stickleback from eutrophicated conditions, are needed. This would increase the resolution of differentiation between habitats, and clarify the evolutionary patterns within and between populations. It would also be valuable to investigate the seasonal movements and patterns of gene flow in sticklebacks in the Baltic Sea. Since the three-spined stickleback's genome has been sequenced, isolation and sequencing of $F_{ST}$-outlier loci would help us understand which traits are under genetic selection in eutrophicated conditions. Further, comparing genes under selection in eutrophicated and other polluted environments could be another approach to achieve a better knowledge of the mechanisms of adaptation in three-spined sticklebacks in the Baltic Sea.

Acknowledgements

I would like to thank my supervisors, Emma Lind and Josefine Larsson, for invaluable help, guidance and support through this whole process – all the way from lab-work and subsequent statistics to the writing of this article. I would not have made it without you! I would also like to thank PhD Ulla Tuomainen at the University of Helsinki for the stickleback samples, as well as Emma Lind and Josefine Larsson, for the permission to use them. I take a big bow to Mats Grahn and Mikael Lönn for precious comments and inspiration. Thanks also to Patrik Dinnétz and Kari Lehtilä for statistical advise. Last but not least I thank Andrej Pierow for being there for me at all times!

References


Vekemans, X. (2002). AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.


