

Neutral and adaptive differentiation in pike (*Esox lucius*) populations from contrasting environments

Johanna Sunde¹, Yeşerin Yıldırım¹, Petter Tibblin¹, Dorte Bekkevold², Christian Skov², Oscar Nordahl¹, Per Larsson¹, and Anders Forsman¹

¹Linnaeus University

²Technical University of Denmark National Institute of Aquatic Resources Silkeborg

October 1, 2020

Abstract

Understanding how eco-evolutionary processes and environmental factors drive population differentiation and adaptation are key challenges in evolutionary biology and of relevance for biodiversity protection. Differentiation requires at least partial reproductive separation, which may result from geographic isolation (allopatry), isolation by distance (IBD), environment (IBE), adaptation (IBA), and time (IBT). We investigate how ecological and evolutionary processes influence genetic diversity and structure in 11 populations of pike (*Esox lucius*) using Restriction-site Associated DNA sequencing (RADseq). Study populations represented three ecotypes (freshwater, anadromous, and brackish water resident) along a latitudinal gradient (54.9 - 63.6 °N). Genetic diversity and structure were investigated both for the full RADseq dataset (5993 loci) and for an adaptive subset consisting outlier loci. Both neutral and adaptive processes influenced genetic structure, and their contributions differed between allopatric and sympatric populations, and also within and among ecotypes. Signatures of neutral processes were pronounced among geographically isolated freshwater populations, likely reflecting long time since divergence combined with low gene flow. For sympatric populations, ecotype (anadromous versus resident) and geography influenced both neutral and adaptive genetic structure, consistent with IBE. Outlier analyses pointed to a role of selection associated with salinity and temperature, consistent with IBA. Results provide rare evidence that separate analyses of neutral and adaptive loci can help illuminate how different, potentially interacting, processes jointly contribute to shaping spatiotemporal patterns of biodiversity. It is argued that data on adaptive rather than neutral genetic variation should inform management and policy development.

Running title: Neutral and adaptive differentiation in pike

Johanna Sunde^{1*}, Yeşerin Yıldırım¹, Petter Tibblin¹, Dorte Bekkevold², Christian Skov², Oscar Nordahl¹, Per Larsson¹ & Anders Forsman¹

Affiliations:

¹ Ecology and Evolution in Microbial Model Systems, EEMiS, Department of Biology and Environmental Science, Linnaeus University, 392 31 Kalmar, Sweden

² National Institute of Aquatic Resources, Technical University of Denmark, 8600 Silkeborg, Denmark

ORCID

Johanna Sunde: <https://orcid.org/0000-0002-3145-1475>

Yeşerin Yıldırım: <https://orcid.org/0000-0002-0431-8417>

Petter Tibblin: <https://orcid.org/0000-0001-6804-5342>

Dorte Bekkevold: <https://orcid.org/0000-0002-5297-032X>

Christian Skov: <https://orcid.org/0000-0002-8547-6520>

Oscar Nordahl: <https://orcid.org/0000-0001-9667-1228>

Per Larsson: <https://orcid.org/0000-0003-0344-1939>

Anders Forsman: <https://orcid.org/0000-0001-9598-7618>

* Corresponding author. E-mail: johanna.sunde@lnu.se, Phone: +46-(0)480-446743

Number of text pages (including references, tables and figures): 47

Number of figures: 5

Number of tables: 3

Supplementary material: 10 pages (including 6 figures)

Manuscript for Molecular Ecology

Version 2020-09-28

Abstract

Understanding how eco-evolutionary processes and environmental factors drive population differentiation and adaptation are key challenges in evolutionary biology and of relevance for biodiversity protection. Differentiation requires at least partial reproductive separation, which may result from geographic isolation (allopatry), isolation by distance (IBD), environment (IBE), adaptation (IBA), and time (IBT). We investigate how ecological and evolutionary processes influence genetic diversity and structure in 11 populations of pike (*Esox lucius*) using Restriction-site Associated DNA sequencing (RADseq). Study populations represented three ecotypes (freshwater, anadromous, and brackish water resident) along a latitudinal gradient (54.9 - 63.6 °N). Genetic diversity and structure were investigated both for the full RADseq dataset (5993 loci) and for an adaptive subset consisting outlier loci. Both neutral and adaptive processes influenced genetic structure, and their contributions differed between allopatric and sympatric populations, and also within and among ecotypes. Signatures of neutral processes were pronounced among geographically isolated freshwater populations, likely reflecting long time since divergence combined with low gene flow. For sympatric populations, ecotype (anadromous versus resident) and geography influenced both neutral and adaptive genetic structure, consistent with IBE. Outlier analyses pointed to a role of selection associated with salinity and temperature, consistent with IBA. Results provide rare evidence that separate analyses of neutral and adaptive loci can help illuminate how different, potentially interacting, processes jointly contribute to shaping spatiotemporal patterns of biodiversity. It is argued that data on adaptive rather than neutral genetic variation should inform management and policy development.

Keywords: adaptation, *Esox lucius*, genetic differentiation, RADseq, pike, population structure, selection, outlier loci

Introduction

Understanding how eco-evolutionary processes and environmental factors drive population differentiation and adaptation remain key challenges in evolutionary biology. Besides informing about the origin and dynamics of biodiversity, knowledge about drivers can promote successful management and protection of populations and species. For differentiation to occur populations must be at least partially reproductively separated. That geographically isolated (allopatric) populations experience such reproductive separation is self-evident. However, reproductive isolation may also arise through other mechanisms that restrict gene flow. For example, the Isolation By Distance model (IBD; Wright, 1943) proposes that there is a negative association between geographic distance and gene flow, which should translate into clinal population structures with higher degrees of differentiation on longer distances. Apart from geographical dispersal boundaries, ecological factors

can influence the degree of gene flow (Rundle & Nosil, 2005). Differences in ecology and environmental preferences (*e.g.* different ecotypes) can instigate population differentiation via Isolation By Environment (IBE; Sexton, Hangartner, & Hoffmann, 2014; Wang & Bradburd, 2014). There is also potential for population genetic structure to evolve as a result of Isolation By Time (IBT), whereby reproductive separation occurs between groups of individuals due to differences in reproductive timing (Hendry & Day, 2005).

Differences in ecology and/or environmental preferences are generally accompanied by divergent selection. The evolution of different local adaptations can cause a feed-back loop that reinforces the reproductive separation and speeds up the differentiation process (a case of IBE often referred to as Isolation by Adaptation (IBA); Nosil, Egan, & Funk, 2008). The reinforcing effect is likely of particular importance when such divergent selection acts upon traits directly connected to reproductive isolation and reproductive success, *e.g.* habitat preferences (Hendry, Nosil, & Rieseberg, 2007), spawning segregation (Nosil, 2012), or early life history traits (Momigliano et al., 2017), and it might ultimately lead to ‘ecological speciation’ (Schluter & Rambaut, 1996).

Northern temperate freshwater fish species present good opportunities to study differentiation and evolutionary dynamics (Hume, Recknagel, Bean, Adams, & Mable, 2018). It was not until after the retreat of the glaciers (15,000 – 10,000 years ago) that many areas were colonized via range expansions (Petit et al., 2003; Rowe, Heske, Brown, & Paige, 2004). The geographical isolation in glacial refugia, in combination with the subsequent range expansions, means that there has been opportunities for founder events, bottlenecks, divergent selection, and adaptations to local conditions to influence genetic structure and diversity, and the different contemporary ecotypes that can be observed today have likely evolved during this time. In aquatic systems, gene flow is generally expected to be high due to the lack of apparent dispersal boundaries (Gagnaire et al., 2015; Puebla, Bermingham, & McMillan, 2012). Despite this, genetic structuring has been reported within open and connected waterbodies (Bergek & Björklund, 2007; Momigliano et al., 2017; Nordahl et al., 2019; Rosenbaum et al., 2009; Willing et al., 2010;), even on small spatial scales, which is likely due to ecological and environmental heterogeneity imposing limits to gene flow. One example of such a system is the Baltic Sea, one of the largest brackish-water ecosystems in the world, that exhibits steep environmental north-south gradients in salinity and temperature (Bendtsen, Söderkvist, Dahl, Hansen, & Reker, 2007). This makes the Baltic Sea an excellent system for studies of differentiation and local adaptation (*e.g.* in Guo, DeFaveri, Sotelo, Nair, & Merilä, 2015; Guo, Li, & Merilä, 2016).

One of the most common, coastal, predatory fish species in the Baltic Sea is pike (*Esox lucius*). Pike plays an important role in many aquatic ecosystems by regulating abundances of species in lower trophic levels (Donadi et al., 2017). It is an economically important species for both recreational and commercial fishing (Lehtonen, Leskinen, Selen, & Reinikainen, 2009; Pierce, Tomcko, & Schupp, 1995), and an established model species in studies of ecology and evolution (Forsman et al., 2015). It is originally a freshwater species, but it has managed to colonise brackish waters with salinities up to approximately 15 ppt (Craig 2008). This range expansion has been accompanied by the evolution of three different ecotypes that differ with regard to migratory behaviour. The original ecotype (freshwater) spend the entire life in freshwater, whilst the other ecotypes (anadromous and resident) spend either part of, or their entire life, in saline (brackish) waters. The two latter ecotypes co-exist in the coastal waters of the Baltic Sea during most of the year (Westin & Limburg, 2002), and separate only for a short period during spawning when the anadromous individuals migrate to freshwater localities whilst the resident ecotype stays in the brackish coastal waters (Engstedt, Stenroth, Larsson, Ljunggren, & Elfman, 2010; Larsson et al., 2015; Muller, 1986). An important feature of pike in the Baltic Sea, which affects the population structure of the species (Nordahl et al., 2019), is their homing behaviour (Engstedt, Engkvist, & Larsson, 2014; Larsson et al., 2015; Tibblin, Forsman, Borger, & Larsson, 2016), *i.e.* the adults return to their natal spawning ground for reproduction (Engstedt et al., 2014; Engstedt et al., 2010; Jacobsen et al., 2017; Muller, 1986).

While a previous study of pike in the Baltic Sea has indicated genetic differentiation between the resident and anadromous ecotypes (Nordahl et al., 2019), studies of the genetic structuring within the ecotypes show conflicting results. Some studies suggest weak structuring (Laikre et al., 2005; Wennerström, Olsson, Ryman,

& Laikre, 2016), whilst others report fine-scaled genetic structuring among anadromous populations, and point to a role of isolation by distance (Möller, Winkler, Richter, & Bastrop, 2020; Nordahl et al., 2019; Sunde, Yildirim, Tibblin, & Forsman, 2020a). There is also evidence to suggest that environmental differences among spawning locations have resulted in adaptive phenotypic differentiation in various traits, including salinity tolerance (Sunde, Tamario, Tibblin, Larsson, & Forsman, 2018), temperature tolerance (Sunde, Larsson, & Forsman, 2019), vertebral number (Tibblin, Berggren, Nordahl, Larsson, & Forsman, 2016), body size and growth rate (Tibblin et al., 2015), early life history traits, and reproductive investment (Berggren, Nordahl, Tibblin, Larsson, & Forsman, 2016). Pike in the Baltic Sea therefore offers possibilities to study differentiation at different levels (between allopatric and sympatric populations, and within and among ecotypes), to evaluate the contributions to genetic structure of different types of reproductive isolation (such as IBD, IBE, IBA, and IBT), and to investigate the potential genetic underpinnings of local adaptations.

Increased knowledge into these issues is also important for management. Declines in pike populations in the Baltic Sea have been observed during the last decades (Lehtonen et al., 2009; Ljunggren et al., 2010; Nilsson, Flink, & Tibblin, 2019; Olsson, 2019), and extensive management actions, such as restoration of spawning habitats and fisheries regulations, have been implemented to counteract this negative trend (Larsson et al., 2015; Nilsson et al., 2019). Despite this, abundances remain low, and there is a need to increase the understanding to aid successful management. Previous population assignments have mainly been based on studies utilizing neutral microsatellite markers (Bekkevold, Jacobsen, Hemmer-Hansen, Berg, & Skov, 2015; Eschbach et al., 2019; Moller et al., 2020; Nordahl et al., 2019; Wennerstrom et al., 2016; but see Sunde et al., 2020b). Informed decisions regarding management and conservation efforts require knowledge about how the combination of stochastic and deterministic processes shape genetic structure and diversity. This in turn requires that population genetic studies use molecular markers that capture variation also in coding regions (Sunde et al., 2020a). To investigate genetic structure and decipher the roles of different ecological and evolutionary processes for differentiation and adaptation in pike, we therefore performed a population genetic study utilizing Restriction Site Associated DNA sequencing (RADseq). For this we used 11 populations spanning 8.7 degrees in latitude, from Denmark (Askeby) in the south to Umea in northern Sweden (**Figure 1**, **Table 1**), that experience different environmental conditions (*e.g.* ice cover, light, salinity, temperature, prey species, competition, and predation). The populations represented all three ecotypes (freshwater, anadromous and resident) and, to our knowledge, this is the first population genetic study of pike that includes populations of all three ecotypes.

The RADseq method yields thousands of Single Nucleotide Polymorphisms (SNPs) residing in both coding (functional) and non-coding (mainly selectively neutral) genomic regions (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). Thus, comparisons of genetic structure and diversity obtained using separate analyses of neutral and functional SNPs, can inform about the roles of both stochastic and deterministic processes (Andrews et al., 2016; de Villemereuil & Gaggiotti, 2015;). In the present study, we therefore investigated genetic structure using both neutral and adaptive SNP datasets. In addition, we performed outlier analyses to identify loci putatively under selection and to pinpoint environmental factors contributing to evolutionary divergence among ecotypes and populations.

Material and methods

Study populations and sampling procedure

A total of 234 pike from 11 populations ($N = 8 - 27$) were sampled for this study. The locations of the populations ranged from 63.6 – 54.9 degN, and represented all three spawning ecotypes (freshwater, anadromous, and resident; for details see **Figure 1** and **Table 1**). Our initial plan was to include both anadromous and resident populations throughout the latitudinal range. However, due to the aforementioned decreases of pike in the Baltic Sea, many of the resident populations are no longer found in their former spawning locations, and we were therefore not able to include more resident populations. Individuals were

captured using fyke nets, rod-and-reel fishing, or electrofishing, and a non-lethal DNA sample (fin clip) was collected for each individual before they were released back into the water. The fin clips were placed in separate 1.5 mL Eppendorf tubes with 70 % ethanol, which were stored in a freezer (-20 degC) until the molecular work was conducted.

Molecular workflow

DNA was extracted from fin tissue with DNeasy blood and tissue kit (Qiagen, USA), and digested with HF EcoRI (New England Biolabs, USA). Size-selection, library preparation, sequencing (using either Illumina NovaSeq 6000 2 x 150 bp or Illumina HiSeq 2500 2 x 125 bp, **Table 1**), demultiplexing and quality control with MultiQC (Ewels, Magnusson, Lundin, & Kaller, 2016) were performed by Science for Life Laboratory (Stockholm, Sweden). Sequence data for Harfjarden, Okne and Lervik was previously published (Sunde et al., 2020a) and retrieved from the NCBI Sequence Read Archive (Sunde, Yildirim, Tibblin, & Forsman, 2020b). In total this yielded a dataset of ~ 3,200 M raw reads (mean 13.6 M per sample). Quality filtering of the raw reads was conducted using Trimmomatic (Bolger, Lohse, & Usadel, 2014) and process_radtags in the Stacks pipeline (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). The ~ 3,000 M reads (95 %) that passed the quality filtering, were further processed with the integrated approach (Paris, Stevens, & Catchen, 2017) of the Stacks *de novo* pipeline, using parameter settings determined by an initial parameter optimization (**Figure S1-S2**). This yielded a final dataset of 5993 biallelic SNPs, for which missing data was imputed using Beagle (Browning, Zhou, & Browning, 2018) before it was used in the downstream analyses. For details on the molecular workflow see section ‘Supplementary methods’ in the Supplementary materials.

Investigations of neutral genetic variation

Analysis of neutral genetic diversity and population structure

The Populations software in the Stacks pipeline was used to obtain estimates of number of private alleles, observed heterozygosity (H_O), expected heterozygosity (H_E), and fixation index (F_{is}) for each population. The same software was used to obtain estimates of the fixation index (F_{ST}) by Weir and Cockerham (1984), that was used to assess pairwise population differentiation.

To assess genetic structuring, the full dataset was analysed using multiple approaches. To determine the most likely number of genetic clusters (K), fastSTRUCTURE (Raj, Stephens, & Pritchard, 2014) was used. Because fine-scaled differentiation might be concealed by differentiation on higher levels (Evanno, Regnaut, & Goudet, 2005), we also ran the analysis with a subset of only the Baltic Sea populations (anadromous and resident) to test if further differentiation became evident. A Principal Component Analysis (PCA) based on pairwise genetic similarity between individuals (proportion of alleles shared) was visualized with the prcomp function in RStudio 2 (RStudio Team, 2015) with R (R Core Team, 2012). For details about settings see ‘Supplementary methods’ in the supplementary materials.

To determine whether the full dataset was representative of neutrally evolving diversity, the results were compared to those generated by a subset ‘neutral’ dataset. This ‘neutral’ dataset was created by excluding the 5 % tails of the F_{ST} distribution from the full dataset, which should exclude loci under strong selection (both divergent and balancing). Patterns of genetic structure indicated by such a neutral dataset should therefore be reflective of only stochastic processes. The comparison revealed that the results obtained for the full and the ‘neutral’ datasets were qualitatively similar (not shown), indicating that the full dataset was representative of neutral diversity. We therefore chose to proceed with the full dataset for the comparisons with adaptive genetic variation and structure (see below).

Isolation by distance

The full dataset for only the Baltic Sea samples (anadromous and resident populations) was used to test for the presence of a correlation between geographical and genetic distances (F_{ST}) between the populations. The freshwater populations were excluded because we did not expect them to exhibit an IBD pattern because of geographical isolation among the lakes. We performed a Mantel test (Mantel, 1967) in Arlequin (Fdist; Excoffier & Lischer, 2010), the significance was tested with 1,000 randomization, and the correlation was visualized in ggplot2 (Wickham, 2016) in R. An additional Mantel test was performed for only the anadromous populations to test for potential correlation between geographical and genetic distances within this ecotype.

Testing for effects of environmental variables

To investigate the roles of salinity (ecotype) and/or latitude for neutral genetic structure, we used the similarity matrix (for the full dataset) created for the PCA in a distance-based redundancy analysis (db-RDA), which is a constrained version of a PCA (db-RDA; Legendre & Anderson, 1999). Because potential adaptations to salinity either could have resulted from differences in the spawning habitats, or reflect differences in salinity in the foraging areas, we wanted to account also for the salinity gradient in the Baltic Sea in the analysis. To this end, we chose to use midrange salinity (mean value of minimum and maximum salinity experienced during the life-cycle) as a proxy for ecotype (salinity for each population was obtained from Bendtsen et al. (2007); see **Table 1**). This resulted in that all freshwater populations were assigned a value of 0, the resident population the highest value of 12, and the anadromous values in the intermediate range of 1.5 to 6.0.

To make the variables comparable for the db-RDA, both environmental variables (latitude and midrange salinity) were normalized (to a standard deviation of 1; by first subtracting the normal value from the observed value, and then dividing by the standard deviation of all observed values) before running the db-RDA. For the normalization we used 0 (psu) as the normal value for salinity (because pike is of freshwater origin; Craig, 1996; Raat, 1988), and the mean latitude as normal value for the latitude. The db-RDA was run in the vegan package (Oksanen et al., 2019) in R, and the statistical significance were assessed with 999 permutations.

Phylogenetic analysis

The evolutionary relationship among the samples was investigated with a maximum likelihood (ML) based phylogenetic inference in RAxML v8.2.10 (Stamatakis, 2014). This was done by using the full dataset following the steps described in 'The Hard & Slow Way' in the RAxML manual (Stamatakis, 2016), which includes optimizing the parameter settings for number of rate categories and initial rearrangement. After the optimization, the final analysis with 200 inferences and 1,000 bootstraps was run using 10 as initial rearrangement, 25 rate categories, the GTRCAT model of nucleotide substitution, and using the *E. lucius* genome published by Rondeau et al. (2014) as outgroup. The phylogenetic tree was then visualized using Interactive Tree Of Life (iTOL)(Letunic & Bork, 2019).

Investigations of adaptive genetic variation

Identification of loci putatively under selection

The full dataset was used in the outlier analyses to search for loci putatively under selection. To test for locus-specific effects, populations were introduced as separate groups, and the data was analysed using multiple approaches, in three different software (BayeScan, Fdist, and LOSITAN; for details about settings see Supplementary methods in the supplementary materials). Only loci identified as outliers by all three software were retained. This was done because differences in the algorithms and assumptions of the approaches might lead to different SNPs being identified as outliers; and by conservative selection of only overlapping ones,

it is more likely that the identified outliers are true positives (de Villemereuil, Frichot, Bazin, Francois, & Gaggiotti, 2014).

To test for correlations between selection and ecotype, we utilized BayeScEnv (de Villemereuil & Gaggiotti, 2015), which aims at differentiating signals of selection from those of demographic processes by searching for associations between co-variables (environmental variables) and allele frequencies. A benefit with BayeScEnv when searching for outlier loci associated with co-variables is that the software can simultaneously evaluate the roles of two environmental variables in a single analysis. Thus, using this software enabled us to search for outlier loci associated with midrange salinity after statistically accounting for variation due to latitudinal differences. For this, we used the same normalized values for the two environmental variables (midrange salinity and latitude) as in the db-RDA (for details see subsection 'Testing for effects of environmental variables' above).

Analysis of adaptive genetic diversity and structure, and testing for effects of environmental variables

To compare the effects of neutral and adaptive processes on patterns of genetic diversity, a subset dataset that should represent adaptive variation (henceforth referred to as the 'adaptive dataset') was also analyzed. This adaptive dataset was created by selecting the 17 loci that were identified as outliers by all three software used for the outlier analyses with population grouping, and it should therefore be reflective of deterministic processes such as selection. To obtain estimates of H_O , and H_E , the populations software in Stacks was used. The adaptive dataset was also analyzed using the same fastSTRUCTURE and db-RDA procedures as used for the full dataset. For details on procedures see 'Supplementary methods' in the supplementary materials.

Results

Patterns of neutral genetic variation

Neutral genetic diversity and population structure

The analyses of genetic diversity with the Populations software yielded H_O estimates between 0.17 - 0.28, expected heterozygosity H_E estimates between 0.19 - 0.28, and F_{IS} estimates between 0.002 - 0.079 (**Table 1**). The populations had 2 - 137 private alleles, and the number of private alleles were higher in the three allopatric lake populations (63 - 137 per population) compared to all other populations (2 - 42 per population) (**Table 1**). Only a small proportion of the loci were not in HWE (100 - 283 loci for each population, $P < 0.05$), and the F_{IS} distributions were unimodal and peaked at zero for all populations (**Figure S3**), suggesting that the majority of the loci were not affected by null alleles and that the populations were in HWE (Ravinet et al., 2016).

Analysis of pairwise population differentiation revealed that all populations were significantly differentiated from each other ($P < 0.0001$; F_{ST} -value range: 0.06 to 0.25; **Figure 2**). The highest F_{ST} -values were found in the pairwise comparisons including the anadromous population Harfjarden, or at least one of the freshwater populations.

The fastSTRUCTURE analysis of all populations suggested that the most likely number of genetic clusters (K) was either 6 (for model complexity that maximizes marginal likelihood) or 7 (for model used to explain structure in the data) (**Figure 1**). The results obtained using $K = 6$ and $K = 7$ were generally similar, and only differed with regard to $K = 7$ assigning Snackstavig to a population-specific cluster, whilst it was assigned to a shared cluster for $K = 6$. All three freshwater lake populations were mainly assigned to population-specific clusters, indicating strong differentiation from all other populations. Within the Baltic Sea samples (anadromous and resident populations), both ecotype and region seemed to influence the patterns of genetic structuring. The majority of the anadromous populations (5 and 4 of 7 for $K = 6$ and $K = 7$, respectively) were assigned to a shared 'anadromous genetic cluster'. All the populations assigned to this

anadromous cluster were located on the Swedish mainland, whilst the two anadromous populations that were not included in the main anadromous cluster (for neither $K = 6$ nor $K = 7$), were Askeby from Denmark and Harfjorden from the island of Oland. Notably, the two populations from Denmark (Askeby and Stege Nor) were assigned to a shared genetic cluster despite belonging to different ecotypes.

The fastSTRUCTURE analysis of the subset including only the Baltic Sea populations (anadromous and resident) suggested that the most likely number of K was either 3 (for model complexity that maximizes marginal likelihood) or 4 (for model used to explain structure in the data), and revealed the same patterns as were found when analysing all populations (**Figure S4**). The results from the PCA were in agreement with the findings from fastSTRUCTURE, showing that the freshwater lake populations formed separate distinct clusters (suggestive of differentiation). They further indicated that the anadromous mainland populations formed an overlapping cluster, and that the two Danish populations were genetically close to each other (**Figure S5**).

Isolation by distance

Based on the Mantel tests, there was no association between geographic and genetic distance for either all Baltic Sea populations (anadromous and resident; $R = -0.161$, $P = 0.678$) or within the anadromous ecotype ($R = -0.204$, $P = 0.640$; **Figure S6**).

Associations with environmental variables

The db-RDA based on the full dataset revealed significant effects on genetic distance of both midrange salinity ($F_{1,232} = 9.46$, $P < 0.001$) and latitude ($F_{1,232} = 6.02$, $P < 0.001$). In addition, a significant interaction effect was found between latitude and midrange salinity ($F_{1,231} = 8.60$, $P < 0.001$). The biplot revealed that the individuals clustered within populations, and that most of the populations were separated (with the exception of some of the more closely located anadromous populations overlapping). The direction of separation associated with midrange salinity was almost parallel with the CAP1 axis, and that the direction of separation associated with latitude corresponded fairly well with the CAP2 axis. In addition, the biplot revealed that the interaction effect seems to reflect that the direction of separation associated with midrange salinity explains the differentiation between the ecotypes, whilst the direction of separation associated with latitude explains the differentiation among the anadromous populations (the arrow aligns along the separation among the anadromous populations; **Figure 3a**).

Phylogenetic analysis

The phylogenetic analysis revealed overall low levels of variation and shallow branching, which suggests that the populations are closely related (**Figure 4**). The samples were, however, clearly grouped into separate populations, and recent nodes within populations were supported by relatively high bootstrap values. The outgroup (reference genome published by Rondeau et al. (2014)) was most closely related to one of the anadromous populations (Angeran), and the anadromous populations constituted a paraphyletic group. However, bootstrap values associated with the deeper nodes were low, and the relationships between the populations could therefore not be reliably resolved.

Patterns of adaptive genetic variation

Loci putatively under selection

All approaches utilized for the outlier analyses identified loci putatively under selection. Locus-specific effects were found for 28 loci with BayeScan (q -values < 0.05), for 231 loci with Fdist ($P < 0.01$), and for 635 loci with LOSITAN ($P < 0.01$). Of all these loci, 17 were identified by all three software (**Figure 5, Table 2**). These 17 loci were used to search for candidate genes putatively under selection, which showed that 10 of the loci were located in previously annotated genes, and because two of the loci resided in the same gene, nine candidate genes were identified (**Table 2**).

When utilizing BayeScEnv to search for associations between allele frequencies and the environmental variables (midrange salinity and latitude), 13 loci were identified as outliers (q -value < 0.05 ; **Figure 5, Table 3**), of which eight were located in previously annotated genes. Because two of the loci resided in the same gene, seven candidate genes were identified (**Table 3**), of which three were also identified in the test of locus-specific effects mentioned in the previous paragraph. The annotation further revealed that some of the identified candidate genes have previously been found to be associated with salinity. One of these genes was potassium voltage-gated channel subfamily A member 10 (KCNA10), which encodes a voltage-dependent potassium-selective channel that has been found to be associated with salinity stress in blue mussels (*Mytilus* spp.) (Lockwood & Somero, 2011). Another gene was vesicle-associated membrane protein 7 (VAMP7), which is crucial for calcium regulated lysosomal exocytosis, and has been found to be involved in salt-tolerance in *Arabidopsis thaliana* (Leshem et al., 2006). These genes therefore seem particularly interesting as candidate genes involved in adaptation to salinity. The results also revealed some genes that might be associated with temperature. One gene that seems extra interesting is zinc-finger protein 436-like (ZNF436-like), which was also identified as putatively under selection in the previous study of three of the populations in the present study (Sunde et al., 2020a). The exact function of ZNF436 in fish is not known, but it is a transcription factor that has been found to have a critical role in regulating early cardiac development in humans (Fu et al., 2018). Other transcription factors have also previously been found to be important in responses to heat-stress in sea urchin (*Strongylocentrotus intermedius*) (Zhan et al., 2019), and some ZNF genes associated with acclimation to low salinity in the euryhaline fish half-smooth tongue sole (*Cynoglossus semilaevis*) (Si et al., 2018) and migratory behaviour in brown trout (*Salmo trutta*) (Lemopoulos, Uusi-Heikkilä, Huusko, Vasemagi, & Vainikka, 2018).

Adaptive genetic diversity and structure, and effects of environmental variables

The Populations software yielded H_O estimates in the range between 0.08 - 0.26 and H_E estimates in the range between 0.09 - 0.27 (**Table 1**). The fastSTRUCTURE analysis revealed that the pattern of adaptive genetic structure differed from that of neutral genetic structure (**Figure 1**). The grouping of the mainland anadromous populations in a shared genetic cluster observed for the full (neutral) dataset was not detected in the adaptive dataset. The adaptive data set instead revealed a pattern of adaptive structuring associated with latitude, and further suggested that two of the freshwater populations (that were strongly neutrally differentiated) shared a genetic cluster (**Figure 1**).

As for the full dataset, the db-RDA based on the adaptive dataset revealed significant effects on genetic distance of both midrange salinity ($F_{1,232} = 11.90$, $P < 0.001$) and latitude ($F_{1,232} = 14.48$, $P < 0.001$), and a significant interaction effect ($F_{1,231} = 17.29$, $P < 0.001$). The biplot revealed that there was considerably more overlap between populations and ecotypes compared to the full dataset. The direction of separation associated with latitude corresponded fairly well with the CAP1 axis, and the direction of separation associated with midrange salinity did not correspond with either of the first two CAP axes. As for the full dataset, the direction of separation associated with latitude aligned with the separation of the anadromous populations (**Figure 3b**). However, the separation among ecotypes was not as clear for the adaptive as for the full dataset, and one of the freshwater populations overlapped with some of the anadromous populations.

Discussion

We report on the first investigation of population genetic diversity and structure among all three spawning ecotypes of pike (freshwater, anadromous and brackish water resident) using information based on both neutral and adaptive molecular markers (RADseq SNPs). We document patterns of genetic structure at different spatial scales and identify the eco-evolutionary drivers of the genetic differentiation among and within ecotypes. Besides providing rare evidence of contrasting patterns of neutral and adaptive genetic structure, results exemplify how separate analyses of coding and non-coding variation can help disentangle the complex interplay of different stochastic and deterministic contributing processes. Specifically, we found

that for neutral differentiation allopatry seems to play an important role. For the sympatric Baltic Sea populations, IBD did not explain the genetic structure. Results instead pointed to effects of ecotype (IBE) and spatial sorting (Berggren, Tinnert, & Forsman, 2012; Shine, Brown, & Phillips, 2011), and indicated that IBT might also influence the neutral genetic structure in pike. For adaptive differentiation, temperature and salinity appear to be two key environmental factors driving local adaptations (IBA). Below, we discuss our findings in relation to previous studies and predictions from theory, and their implications for management.

Within population diversity

Observed within population heterozygosity (H_O) for the full dataset (0.17 - 0.28, **Table 1**) were considerably lower than reported in previous microsatellite studies of Baltic Sea pike (0.22 - 0.66; Bekkevold et al., 2015; Nordahl et al., 2019; Sunde et al., 2020a; Wennerstrom et al., 2016). This may in part reflect the use of different markers. This interpretation is supported by the results from the study by Sunde et al. (2020a), in which both microsatellites and RADseq were utilized to study genetic diversity and structure of three pike populations (that are also included in the present study). The results revealed similar estimates of SNP heterozygosity as in the present study, and microsatellite heterozygosity estimates that are comparable to those obtained in the other studies utilizing microsatellites (0.40 - 0.57). Regardless of marker type, both the present (**Table 1**) and previous studies show that within-study estimates of H_O and H_E are similar, which indicates that the populations do not interbreed to a large extent (indicated by $H_O > H_E$), and do not show any clear signs of inbreeding (indicated by $H_O < H_E$).

Diversity is imperative for eco-evolutionary success. Theory and empirical evidence concur that larger genetic and phenotypic variation among individuals within populations may promote establishment success, stabilize population dynamics, allow for faster range expansions, and reduce extinction risk (Des Roches et al., 2018; Forsman, 2014; Forsman & Wennersten, 2016; Hughes, Inouye, Johnson, Underwood, & Vellend, 2008; Waldman, Wilson, Mather, & Snyder, 2016). Our present estimates do not indicate any clear differences in intrapopulation diversity among the ecotypes, neither for estimates based on the full nor on the adaptive dataset (**Table 1**). However, within ecotypes, some populations seem to harbor larger genetic diversity than others, which might reflect more heterogeneous environments or result from a higher degree of gene flow. Yet, these conclusions must be tentative because of the relatively low number of adaptive loci and the unbalanced sampling design, with varying numbers of populations representing the three ecotypes.

Allopatric freshwater populations were strongly neutrally differentiated

We found that all populations were significantly differentiated (based on F_{ST} ; **Figure 2**), which is consistent with previous studies (Bekkevold et al., 2015; Larsson et al., 2015; Moller et al., 2020; Nordahl et al., 2019; Sunde et al., 2020a; Wennerstrom et al., 2016). Levels of pairwise population differentiation ranged from low to high (F_{ST} 0.06 - 0.25), and the highest levels of differentiation were found in the comparisons including either the anadromous population Harfjorden, or at least one freshwater lake population (**Figure 2**). This, together with the finding that all three freshwater populations were strongly differentiated from each other, and distinct from the other two ecotypes point to an important effect of allopatry.

It has been suggested that pike experienced drastic population declines and/or bottlenecks following post-glacial recolonization across the Northern Hemisphere (Jacobsen, Hansen, & Loeschcke, 2004), and that the succeeding isolation of local river and lake systems might have resulted in genetic drift and differentiation among freshwater populations (Bekkevold et al., 2015). Along this line of argument, the substantial differentiation among the freshwater populations likely reflects the combined effects of founder events, genetic drift, low gene flow, more distinct reproductive isolation, divergent selection, and longer time since divergence (because the freshwater ecotype is the original ecotype).

Sympatric Baltic Sea populations were also differentiated

Estimates of population differentiation in Baltic Sea pike differ markedly among studies. Notably, studies on smaller spatial scales report higher levels of differentiation (0.026 - 0.394 in Bekkevold et al., 2015; 0.013 - 0.396 in Nordahl et al., 2019) than those conducted on large spatial scales (0.005 - 0.135 in Laikre et al., 2005; and -0.003 - 0.14 in Wennerstrom et al., 2016). The values in the present study are relatively high compared to the other large-scale studies (0.06 - 0.22). While it is possible that the use of different markers influence the estimates of F_{ST} , Sunde et al. (2020a) arrived at similar F_{ST} -values for microsatellites and RADseq SNPs, indicating consistency across the markers. So far, few attempts have been made to compare the resolution yielded by various genetic markers, and little is therefore known about whether the use of different marker types contributes to heterogeneity of results among studies of other organisms (Sunde et al., 2020a).

Accurate estimation of genetic differentiation requires correct population assignments of individuals. In our study system, where individuals only separate during spawning, sampling during foraging season entail the risk of erroneously grouping individuals from several populations, which results in underestimations of differentiation. In the present study, the anadromous individuals were therefore sampled in their freshwater spawning habitats to assure accurate population assignment. Previous studies differ in their sampling regimes/designs, which might have affected differentiation estimates, and complicates comparisons among studies. That we found stronger population differentiation than reported before for large-scale studies of Baltic Sea pike nevertheless indicates that the populations are more isolated than previously believed. This can have implications for management, as it indicates that each population should be considered as a separate unit. The importance of sampling scheme and associated implications for interpretation of result pertaining to population structure potentially extend also to other migrating species

(such as whitefish *Coregonus maraena*, Olsson, Florin, Mo, Aho, & Ryman, 2012; and perch *Perca fluviatilis*, Olsson, Mo, Florin, Aho, & Ryman, 2011), but to our knowledge this issue has not been systematically evaluated.

The analyses of genetic structure based on the full dataset showed signs of genetic clustering for the anadromous ecotype (**Figure 1 and S4-S5**), which is indicative of gene flow and/or recent divergence. All populations that were assigned to the shared 'anadromous genetic cluster' spawn in localities on the Swedish mainland, and even the population from Angeran (which resides in the north of the Baltic Sea) was assigned to this cluster. The two anadromous populations that were not assigned to the shared cluster (Askeby from Denmark and Harfjorden from the island of Oland) were the only two anadromous populations not spawning in localities on the Swedish mainland. Genetic clustering of populations along the coast have also been reported in previous large-scale studies of pike (Laikre et al., 2005; Wennerstrom et al., 2016). It has also been shown for pikeperch (*Sander lucioperca*) in the northern part of the Baltic Sea in a study by Saisa, Salminen, Koljonen, and Ruuhijarvi (2010), who showed that coastal populations formed one genetic cluster, whilst freshwater lake populations showed strong genetic differentiation and formed distinct clusters.

All the anadromous populations were differentiated from each other (**Figure 2**), and previous studies report low levels of gene flow among anadromous pike populations (Nordahl et al., 2019; Sunde et al., 2020a; Tibblin et al., 2015). It is therefore unlikely that gene flow is sufficient to explain the clustering of anadromous populations. Coherent with previous studies based on mitochondrial DNA from pike across Northern Europe (Maes, Van Houdt, De Charleroy, & Volckaert, 2003; Skog, Vollestad, Stenseth, Kasumyan, & Jakobsen, 2014), the results from the phylogenetic analysis revealed low levels of genetic variation and shallow branching among the populations (**Figure. 5**), which is indicative of recent divergence. However, the results did not provide any firm evidence for more recent divergence among the mainland populations. The use of more dense SNP data or longer reads might allow for the detection of clearer phylogenetic signals and higher resolution (Cariou, Duret, & Charlat, 2013).

That the anadromous populations from Oland and Denmark were distinct from the other anadromous populations may reflect different evolutionary histories, and a combination of founder events followed by

divergent selection and stochastic processes. That the populations inhabiting the East Coast of Oland show strong differentiation from the mainland populations in the Kalmar sound region has been reported previously (Nordahl et al., 2019), and it has been suggested to result from the open water between the Sweish mainland and the island acting as a reproductive barrier. Similarly, the open water between Denmark and Sweden may constitute a reproductive barrier that has facilitated differentiation. In addition, previous work has shown that the population from Oland and one of the mainland anadromous populations experience different environmental conditions during spawning, and that this has resulted in the evolution of local adaptations during early fry development (temperature, Sunde et al., 2019; salinity, Sunde et al., 2018). It is therefore possible that the high level of differentiation partly reflects IBE/IBA, in addition to geographic separation. The population on Oland generally spawns earlier than the mainland populations (Sunde et al., 2019), and it is possible that IBT also has contributed to the genetic differentiation among anadromous pike populations reproducing in localities on the mainland and the island of Oland. To the extent that the timing of spawning migration is heritable (Tibblin et al. 2016), differences in timing of reproduction among populations may impair the success of individuals that attempt to spawn in a population different from where they were born, and thereby reduce gene flow. An alternative explanation is that this pattern reflects multiple evolutionary origins of anadromy. Skog et al. (2014) suggest that there are two clades of pike in the Baltic Sea, and it is possible that anadromy has evolved independently within the clades, but our present results do not provide conclusive evidence.

Taken together, more recent divergence of the anadromous mainland populations, different evolutionary histories, or multiple evolutionary origins all remain plausible explanations for the observed structure and differentiation, but we are unable to discriminate among them based on existing data. Future studies that include data for additional resident, anadromous and freshwater populations from other regions around the Baltic Sea are required to formally evaluate the competing hypotheses.

That the two Denmark populations (Stegø and Askeby) clustered together, despite belonging to different ecotypes, suggests that, in addition to ecotype, geography influence the genetic structure. Unlike previous large-scale studies of pike (Laikre et al., 2005; Wennerstrom et al., 2016), we found no evidence of IBD. This lack of IBD is perhaps partly explained by the long geographic distances between many of the populations. Patterns of IBD are more pronounced for more closely located populations, and gradually disappear with increasing geographic separation (Hutchison & Templeton, 1999; Meirmans, 2012; Tinnert, Hellgren, Lindberg, Koch-Schmidt, & Forsman, 2016; van Strien, Holderegger, & Van Heck, 2015). Consistent with this notion, the results from fastSTRUCTURE showed some signs of gene flow between the more closely located populations (**Figure 1**). It is therefore likely that IBD is of importance for local genetic structuring, but that other processes such as selection and drift have stronger effects on large scales.

The db-RDA for the full dataset indicated that both ecotype and latitude influence neutral genetic structuring. Results further indicate that whereas ecotype might be one of the main factors influencing neutral genetic structure, latitude appear to explain variation among populations within the anadromous ecotype.

Taken together, the findings reported in the present and previous studies suggest that the patterns of genetic structure observed in Baltic Sea pike have been shaped by an interplay between geography and divergent selection associated with the environments occupied by the different ecotypes (i.e., combined contributions of IBD, IBE, IBA, and IBT), as discussed below.

Adaptive genetic variation and structure

When the adaptive dataset (comprising outlier loci) was analysed, some contrasting patterns of structuring emerged. The clustering of the anadromous mainland populations that was evident for the neutral dataset was not present in the adaptive dataset (**Figure 1**). Instead, a main pattern of structuring associated with latitude appeared. The importance of latitude on adaptive differentiation was also evidenced by the db-RDA, which revealed a direction of separation associated with latitude that corresponded relatively well with the CAP1 axis (**Figure 3b**). The effect of latitude appeared to be especially important within the

anadromous ecotype, which was indicated by the direction of separation associated with latitude aligning with the separation among the anadromous individuals. This likely reflects that the anadromous populations included in this study covered a latitudinal range along the environmental clines in the Baltic Sea.

The significant interaction effect between midrange salinity and latitude indicates that the importance of these two factors differs according to the level of structuring, and further shows that salinity alone does not explain the patterns. Instead, selection associated with multiple environmental factors that co-vary with latitude (*e.g.* temperature and salinity) probably contributes to adaptive genetic structure. This is in agreement with previous findings that salinity and temperature regimes have resulted in locally adapted populations (Sunde et al., 2019; Sunde et al., 2018). Previous studies have also reported on evolution associated with salinity tolerance in other fish species in the Baltic Sea, including three-spined stickleback (Guo et al., 2015; Hasan et al., 2017), European flounder (Momigliano et al., 2017), and Atlantic herring (Berg, Slotte, Andersson, & Folkvord, 2019; Lamichhaney et al., 2012), emphasizing the general importance of salinity.

The role of environmental conditions in shaping adaptive structure was further supported by the finding that two of the freshwater populations (Kosta and Hamnaryd) that showed strong neutral differentiation shared an adaptive genetic cluster (**Figure 1**). This suggests that, despite being geographically and reproductively separated, similarities in environmental conditions between these two freshwater lakes have resulted in adaptive similarity via convergent evolution. The contribution of convergent evolution was further supported by the finding that the db-RDA based on the adaptive dataset showed considerably more overlap between populations (including the two freshwater populations sharing a genetic cluster in fastSTRUCTURE) and ecotypes compared to analysis of the full dataset (**Figure 3**).

The conclusion that environmental conditions, and in particular salinity, influence genetic structure was corroborated by the outlier analyses (**Table 2 and 3**). The results specifically revealed that outliers residing in genes that have been suggested to be associated with salinity tolerance were identified in both analyses. The remainder of the SNPs might reside in genes that have not yet been identified in the pike genome, but it is also possible that these are non-coding loci linked to regions under selection.

Implications for management

Biodiversity is under threat worldwide by natural and anthropogenic environmental makeovers, climate change, and overexploitation. The level of genetic diversity within and among populations can influence the eco-evolutionary success of species, as well as the functioning of ecosystems, and this must inform management and protection of fish and the ecosystem services they provide. Genetic and phenotypic variation is required for populations to respond to selection and adapt to changing and novel environmental conditions (Charlesworth & Charlesworth, 2017; Roff, 1997; Wennersten & Forsman, 2012). There is also potential for the consequences of genetic variation to go beyond the level of the species, as it can influence community structure and ecosystem functioning (Des Roches et al., 2018; Hughes et al., 2008). Being important predators, competitors, and prey to other species, there are many ways by which pike and other species of fish can affect the functioning of lakes, rivers, coastal ecosystems and open oceans (Brodersen, Howeth, & Post, 2015; Donadi et al., 2017; Nilsson et al., 2019; Post, Palkovacs, Schielke, & Dodson, 2008; Tamario, Sunde, Petersson, Tibblin, & Forsman, 2019). There are thus several reasons as to why the genetic diversity among and within ecotypes of pike reported in this and previous studies must not be depleted.

A key challenge for conservation is to design management actions that maintain functional genetic and phenotypic diversity both within and among populations (Hutchinson, 2008; Larsson et al., 2015; Nordahl et al., 2019; Stephenson, 1999; Tamario et al., 2019; Wright, Bishop, Matthee, & von der Heyden, 2015). The rates and directions of genetic exchange between populations may be a natural outcome of dispersal or result from management actions, such as removal of migration barriers, compensatory breeding, supplementary stocking, (re-)introductions, and translocations (Gjedrem, Gjoen, & Gjerde, 1991; McClelland & Naish, 2007; McGinnity et al., 2009; Seddon, Armstrong, & Maloney, 2007). While genetic diversity is beneficial,

restoration efforts may not always generate the desirable outcome (McClelland & Naish, 2007; Verhoeven, Macel, Wolfe, & Biere, 2011; Whitlock et al., 2013). Our finding in the present study that patterns of neutral and adaptive genetic diversity differed, which has also been reported in previous studies (Leinonen, O'Hara, Cano, & Merila, 2008; Reed & Frankham, 2001), indicates that neutral variation is not necessarily reflective of adaptive variation. Given that it is adaptive, not neutral, variation that determines the evolvability of populations and influences their capability of coping with changed environmental conditions, this emphasizes the importance for management to base decisions on analyses of adaptive genetic diversity.

Conclusions

We utilized RADseq to study neutral and adaptive genetic variation and structuring in allopatric and sympatric populations of pike belonging to freshwater, anadromous and brackish water resident spawning ecotypes. To our knowledge, this is the first population genetic study of pike that has included populations of all three ecotypes. We utilized a genetic marker that can inform on both neutral and functional diversity and structure, and provide insights about which processes and environmental factors that influence neutral and adaptive diversity and differentiation.

The populations in lakes formed separate distinct genetic clusters, indicating an important effect of allopatry. For the sympatric Baltic Sea populations, we found that most anadromous populations were assigned to a shared genetic cluster, and that the two populations from Denmark, despite belonging to different ecotypes, were assigned to a separate cluster; pointing to important roles of geographical separation, IBE, IBA and spatial sorting (Berggren et al., 2012; Shine et al., 2011). The differentiation among anadromous populations may also in part reflect an effect of IBT (Sunde et al., 2019). Unlike previous studies (Laikre et al., 2005; Wennerstrom et al., 2016), the population genetic structure detected in the present study could not be adequately explained by IBD. We also found markedly higher levels of differentiation than previous large-scaled studies in the Baltic Sea, which indicates that the populations are more isolated than previously believed, and suggests that each population should be considered as a separate unit.

Besides the importance of within population diversity for various aspects of ecological success (Des Roches et al., 2018; Forsman, 2014; Forsman & Wennersten, 2016; Hughes et al., 2008; Waldman et al., 2016), there is evidence to suggest that portfolio effects associated with variation among populations may increase the stability, productivity, and resilience of species in variable environments (Hui, Fox, & Gurevitch, 2017; Schindler et al., 2010; Waldman et al., 2016). Given ongoing climate change and other anthropogenic environmental makeovers, it is therefore imperative that the genetic structure of sympatric anadromous populations of pike is maintained and that management actions are not implemented that contribute to increased homogenization of populations.

The analysis of adaptive variation revealed different patterns of structuring than did the neutral variation, and indicated adaptive similarity between some reproductively isolated populations - a sign of convergent evolution. Adaptive structuring was associated with latitude, indicating that salinity and temperature drive adaptive differentiation. The importance of salinity and temperature was also evidenced by the outlier analyses, which identified candidate genes that have previously been shown to be associated with these two environmental factors.

Utilization of neutral and functional markers together has potential to provide a comprehensive understanding of the eco-evolutionary processes that jointly influence genetic diversity and shape genetic structure of natural populations. Advances in next-generation sequencing has enabled the use of molecular methods that yield thousands of loci including functional regions. Adaptive variation has a pivotal role for the evolvability of populations, their ability to cope with environmental changes, and how they respond to admixture, management and conservation efforts. The finding that neutral genetic variation was not reflective of adaptive variation thus points to the need for management and conservation efforts to be based not only on neutral, but also functional, genetic diversity, to achieve desirable outcomes and ensure successful protection of

biodiversity.

Ethics note

The study was carried out in accordance with all relevant applicable national guidelines for the care and use of animals. The study was granted ethical approval (approval Dnr ID 83, 39-10, and 33-15), by the Ethical Committee on Animal Experiments in Linköping, Swedish Board of Agriculture, Sweden. For Danish samples, collections were carried out in accordance with ethical guidelines and animal welfare legislation described in permission (2005/561-987) from the Danish Ministry of Justice.

Author contributions

JS, PT, PL, and AF conceived the study. JS, AF, PT, DB, CS and PL participated in the sampling of fishes and collection of DNA samples. JS, YY and DB performed the laboratory work. JS and YY performed the bioinformatical analyses. All authors contributed to interpretation of results. JS, YY and AF drafted the first manuscript. All authors commented on and agreed to the final version of the manuscript.

Acknowledgements

We would like to thank Anssi Laurila, Dennis Hasselquist, Ane Timenes Laugen, Bjorn Rogell and Markus Zottl for discussion and comments on an earlier version of the manuscript, and everyone who helped us by collecting and proving samples. Peter Soderling, Carl Tamario, Kristofer Bergstrom and Per Koch-Schmidt for Harfjarden, Lervik, Okne and Kosta, Par Bystrom and Rickard Linderot for Angeran, Staff at DTU Aqua, and in particular Henrik Dalby Ravn for the Danish samples, Henrik C Andersson for Yngern and Snackstavik, Marcus Hall for Hamnaryd, and Olof Engstedt and Tobias Berger for Vambasa. We would also like to thank Emelie Nilsson and Diego Brambilla for help with the bioinformatical analyses. The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. Funding was provided by the Linnaeus University (to AF), The Swedish Research Council FORMAS (to AF and PT grant Dnr. 2017-00346), and Nilsson-Ehle-donationerna Kungliga Fysiografiska Sallskapet i Lund (to JS and YY grant Dnr 39409).

Data accessibility

Raw sequence reads and metadata will be available in the Sequence Read Archive (SRA) in NCBI (BioProject: PRJNA639676 and BioID: SAMN 15246073 - 15246242; Sunde et al.). Temporal link for review: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA639676?reviewer=dqh8vj30ucio6tm5hqrcf4r1kf>

References

- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature reviews. Genetics*, 17 (2), 81-92. doi:10.1038/nrg.2015.28
- Bekkevold, D., Jacobsen, L., Hemmer-Hansen, J., Berg, S., & Skov, C. (2015). From regionally predictable to locally complex population structure in a freshwater top predator: river systems are not al-

ways the unit of connectivity in Northern Pike *Esox lucius* . *Ecology of Freshwater Fish*, 24 (2), 305-316. doi:10.1111/eff.12149

Bendtsen, J., Soderkvist, J., Dahl, K., Hansen, J. L., & Reker, J. (2007). Model simulations of blue corridors in the Baltic Sea. *BALANCE interim Report*, 9:26 .

Berg, F., Slotte, A., Andersson, L., & Folkvord, A. (2019). Genetic origin and salinity history influence the reproductive success of Atlantic herring. *Marine Ecology Progress Series*, 617 , 81-94. doi:10.3354/meps12680

Bergek, S., & Bjorklund, M. (2007). Cryptic barriers to dispersal within a lake allow genetic differentiation of Eurasian perch. *Evolution*, 61 (8), 2035-2041. doi:10.1111/j.1558-5646.2007.00163.x

Berggren, H., Nordahl, O., Tibblin, P., Larsson, P., & Forsman, A. (2016). Testing for local adaptation to spawning habitat in sympatric subpopulations of pike by reciprocal translocation of embryos. *Plos One*, 11 (5), e0154488. doi:10.1371/journal.pone.0154488

Berggren, H., Tinnert, J., & Forsman, A. (2012). Spatial sorting may explain evolutionary dynamics of wing polymorphism in pygmy grasshoppers. *Journal of Evolutionary Biology*, 25 (10), 2126-2138. doi:10.1111/j.1420-9101.2012.02592.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30 (15), 2114-2120. doi:10.1093/bioinformatics/btu170

Brodersen, J., Howeth, J. G., & Post, D. M. (2015). Emergence of a novel prey life history promotes contemporary sympatric diversification in a top predator. *Nature Communications*, 6 , 8115. doi:10.1038/ncomms9115

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, 103 (3), 338-348. doi:10.1016/j.ajhg.2018.07.015

Cariou, M., Duret, L., & Charlat, S. (2013). Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution*, 3 (4), 846-852. doi:10.1002/ece3.512

Catchen, J., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes/Genomes/Genetics*, 1 (3), 171-182. doi:10.1534/g3.111.000240

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22 (11), 3124-3140. doi:10.1111/mec.12354

Charlesworth, B., & Charlesworth, D. (2017). Population genetics from 1966 to 2016. *Heredity*, 118 , 2-9. doi:10.1038/hdy.2016.55

Craig, J. F. (1996). *Pike – Biology and exploitation* . London: Chapman & Hall.

de Villemereuil, P., Frichot, E., Bazin, E., Francois, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology*, 23 (8), 2006-2019. doi:10.1111/mec.12705

de Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, 6 (11), 1248-1258. doi:10.1111/2041-210x.12418

Des Roches, S., Post, D. M., Turley, N. E., Bailey, J. K., Hendry, A. P., Kinnison, M. T., . . . Palkovacs, E. P. (2018). The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, 2 (1), 57-64. doi:10.1038/s41559-017-0402-5

Donadi, S., Austin, A. N., Bergstrom, U., Eriksson, B. K., Hansen, J. P., Jacobson, P., . . . Eklof, J. S. (2017). A cross-scale trophic cascade from large predatory fish to algae in coastal ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, 284 (1859), 20170045. doi:10.1098/rspb.2017.0045

- Engstedt, O., Engkvist, R., & Larsson, P. (2014). Elemental fingerprinting in otoliths reveals natal homing of anadromous Baltic Sea pike (*Esox lucius* L.). *Ecology of Freshwater Fish*, 23 (3), 313-321. doi:10.1111/eff.12082
- Engstedt, O., Stenroth, P., Larsson, P., Ljunggren, L., & Elfman, M. (2010). Assessment of natal origin of pike (*Esox lucius*) in the Baltic Sea using Sr:Ca in otoliths. *Environmental Biology of Fishes*, 89 (3), 547-555. doi:10.1007/s10641-010-9686-x
- Eschbach, E., Nolte, A. W., Kohlmann, K., Alos, J., Schoning, S., & Arlinghaus, R. (2019). Intraspecific population admixture of a top piscivore correlates with anthropogenic alteration of freshwater ecosystems. *bioRxiv*, 677856. doi:10.1101/677856
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14 (8), 2611-2620. doi:10.1111/j.1365-294X.2005.02553.x
- Ewels, P., Magnusson, M., Lundin, S., & Kaller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32 (19), 3047-3048. doi:10.1093/bioinformatics/btw354
- Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10 (3), 564-567. doi:10.1111/j.1755-0998.2010.02847.x
- Forsman, A. (2014). Effects of genotypic and phenotypic variation on establishment are important for conservation, invasion and infection biology. *Proceedings of the National Academy of Sciences USA*, 111 (1), 302-307. doi:10.1073/pnas.1317745111
- Forsman, A., Tibblin, P., Berggren, H., Nordahl, O., Koch-Schmidt, P., & Larson, P. (2015). Pike *Esox lucius* as an emerging model organism for studies in ecology and evolutionary biology: A review. *Journal of Fish Biology*, 87 (2), 472-479. doi:10.1111/jfb.12712
- Forsman, A., & Wennersten, L. (2016). Inter-individual variation promotes ecological success of populations and species: evidence from experimental and comparative studies. *Ecography*, 39, 630-648. doi:10.1111/ecog.01357
- Fu, K., Nakano, H., Morselli, M., Chen, T., Pappoe, H., Nakano, A., & Pellegrini, M. (2018). A temporal transcriptome and methylome in human embryonic stem cell-derived cardiomyocytes identifies novel regulators of early cardiac development. *Epigenetics*, 13 (10-11), 1013-1026. doi:10.1080/15592294.2018.1526029
- Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., . . . Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, 8 (8), 769-786. doi:10.1111/eva.12288
- Gjedrem, T., Gjoen, H. M., & Gjerde, B. (1991). Genetic origin of Norwegian farmed Atlantic salmon. *Aquaculture*, 98 (1), 41-50. doi:10.1016/0044-8486(91)90369-I
- Guo, B., DeFaveri, J., Sotelo, G., Nair, A., & Merila, J. (2015). Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biology*, 13 (1), 19. doi:10.1186/s12915-015-0130-8
- Guo, B., Li, Z., & Merila, J. (2016). Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Molecular Ecology*, 25 (12), 2833-2852. doi:10.1111/mec.13657
- Hasan, M. M., DeFaveri, J., Kuure, S., Dash, S. N., Lehtonen, S., Merila, J., & McCairns, R. J. S. (2017). Sticklebacks adapted to divergent osmotic environments show differences in plasticity for kidney morphology and candidate gene expression. *The Journal of Experimental Biology*, 220 (12), 2175-2186. doi:10.1242/jeb.146027

- Hendry, A. P., & Day, T. (2005). Population structure attributable to reproductive time: isolation by time and adaptation by time. *Molecular Ecology*, *14* (4), 901-916. doi:10.1111/j.1365-294X.2005.02480.x
- Hendry, A. P., Nosil, P., & Rieseberg, L. H. (2007). The speed of ecological speciation. *Functional Ecology*, *21* (3), 455-464. doi:10.1111/j.1365-2435.2007.01240.x
- Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, *11* , 609-623. doi:10.1111/j.1461-0248.2008.01179.x
- Hui, C., Fox, G. A., & Gurevitch, J. (2017). Scale-dependent portfolio effects explain growth inflation and volatility reduction in landscape demography. *Proceedings of the National Academy of Sciences*, *114* (47), 12507-12511. doi:10.1073/pnas.1704213114
- Hume, J. B., Recknagel, H., Bean, C. W., Adams, C. E., & Mable, B. K. (2018). RADseq and mate choice assays reveal unidirectional gene flow among three lamprey ecotypes despite weak assortative mating: Insights into the formation and stability of multiple ecotypes in sympatry. *Molecular Ecology*, *27* (22), 4572-4590. doi:10.1111/mec.14881
- Hutchinson, W. F. (2008). The dangers of ignoring stock complexity in fishery management: the case of the North Sea cod. *Biology Letters*, *4* (6), 693-695. doi:10.1098/rsbl.2008.0443
- Hutchison, D. W., & Templeton, A. R. (1999). Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, *53* (6), 1898-1914. doi:10.1111/j.1558-5646.1999.tb04571.x
- Jacobsen, Bekkevold, D., Berg, S., Jepsen, N., Koed, A., Aarestrup, K., . . . Skov, C. (2017). Pike (*Esox lucius* L.) on the edge: consistent individual movement patterns in transitional waters of the western Baltic. *Hydrobiologia*, *784* (1), 143-154. doi:10.1007/s10750-016-2863-y
- Jacobsen, Hansen, M. M., & Loeschcke, V. (2004). Microsatellite DNA analysis of northern pike (*Esox lucius* L.) populations: insights into the genetic structure and demographic history of a genetically depauperate species. *Biological Journal of the Linnean Society*, *84* (1), 91-101. doi:10.1111/j.1095-8312.2005.00416.x
- Laikre, L., Miller, L. M., Palme, A., Palm, S., Kapuscinski, A. R., Thoresson, G., & Ryman, N. (2005). Spatial genetic structure of northern pike (*Esox lucius*) in the Baltic Sea. *Molecular Ecology*, *14* (7), 1955-1964. doi:10.1111/j.1365-294X.2005.02570.x
- Lamichhaney, S., Martinez Barrio, A., Rafati, N., Sundstrom, G., Rubin, C.-J., Gilbert, E. R., . . . Andersson, L. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences of the United States of America*, *109* (47), 19345-19350. doi:10.1073/pnas.1216128109
- Larsson, P., Tibblin, P., Koch-Schmidt, P., Engstedt, O., Nilsson, J., Nordahl, O., & Forsman, A. (2015). Ecology, evolution, and management strategies of northern pike populations in the Baltic Sea. *Ambio*, *44* , 451-461. doi:10.1007/s13280-015-0664-6
- Legendre, P., & Anderson, M. J. (1999). Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, *69* (1), 1-24. doi:10.1890/0012-9615(1999)069[0001:Dbratm]2.0.Co;2
- Lehtonen, H., Leskinen, E., Selen, R., & Reinikainen, M. (2009). Potential reasons for the changes in the abundance of pike, *Esox lucius* , in the western Gulf of Finland, 1939–2007. *Fisheries Management and Ecology*, *16* (6), 484-491. doi:10.1111/j.1365-2400.2009.00701.x
- Leinonen, T., O'Hara, R. B., Cano, J. M., & Merila, J. (2008). Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, *21* (1), 1-17. doi:10.1111/j.1420-9101.2007.01445.x

- Lemopoulos, A., Uusi-Heikkilä, S., Huusko, A., Vasemagi, A., & Vainikka, A. (2018). Comparison of migratory and resident populations of brown trout reveals candidate genes for migration tendency. *Genome Biology and Evolution*, 10 (6), 1493-1503. doi:10.1093/gbe/evy102
- Leshem, Y., Melamed-Book, N., Cagnac, O., Ronen, G., Nishri, Y., Solomon, M., . . . Levine, A. (2006). Suppression of *Arabidopsis* vesicle-SNARE expression inhibited fusion of H₂O₂-containing vesicles with tonoplast and increased salt tolerance. *Proceedings of the National Academy of Sciences*, 103 (47), 18008-18013. doi:10.1073/pnas.0604421103
- Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, 47 (W1), W256-W259. doi:10.1093/nar/gkz239
- Ljunggren, L., Sandström, A., Bergström, U., Mattila, J., Lappalainen, A., Johansson, G., . . . Eriksson, B. K. (2010). Recruitment failure of coastal predatory fish in the Baltic Sea coincident with an offshore ecosystem regime shift. *Ices Journal of Marine Science*, 67 (8), 1587-1595. doi:10.1093/icesjms/fsq109
- Lockwood, B. L., & Somero, G. N. (2011). Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Molecular Ecology*, 20 (3), 517-529. doi:10.1111/j.1365-294X.2010.04973.x
- Maes, G. E., Van Houdt, J. K. J., De Charleroy, D., & Volckaert, F. A. M. (2003). Indications for a recent Holarctic expansion of pike based on a preliminary study of mtDNA variation. *Journal of Fish Biology*, 63 (1), 254-259. doi:10.1046/j.1095-8649.2003.00140.x
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27 , 209-220.
- McClelland, E., & Naish, K. (2007). What is the fitness outcome of crossing unrelated fish populations? A meta-analysis and an evaluation of future research directions. *Conservation Genetics*, 8 (2), 397-416. doi:10.1007/s10592-006-9178-x
- McGinnity, P., Jennings, E., deEyto, E., Allott, N., Samuelsson, P., Rogan, G., . . . Cross, T. (2009). Impact of naturally spawning captive-bred Atlantic salmon on wild populations: depressed recruitment and increased risk of climate-mediated extinction. *Proceedings of the Royal Society B: Biological Sciences*, 276 (1673), 3601-3610. doi:10.1098/rspb.2009.0799
- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21 (12), 2839-2846. doi:10.1111/j.1365-294X.2012.05578.x
- Møller, S., Winkler, H. M., Richter, S., & Bastrop, R. (2020). Genetic population structure of pike (*Esox lucius* Linnaeus, 1758) in the brackish lagoons of the southern Baltic Sea. *Ecology of Freshwater Fish*, 00 , 1-10. doi:10.1111/eff.12571
- Momigliano, P., Jokinen, H., Fraimout, A., Florin, A.-B., Norkko, A., & Merilä, J. (2017). Extraordinarily rapid speciation in a marine fish. *Proceedings of the National Academy of Sciences*, 114 (23), 6074-6079. doi:10.1073/pnas.1615109114
- Müller, K. (1986). Seasonal anadromous migration of the pike (*Esox lucius* L) in coastal areas of the northern Bothnian Sea. *Archiv Für Hydrobiologie*, 107 (3), 315-330.
- Nilsson, J., Flink, H., & Tibblin, P. (2019). Predator-prey role reversal may impair the recovery of declining pike populations. *Journal of Animal Ecology*, 88 (6), 927-939. doi:10.1111/1365-2656.12981
- Nordahl, O., Koch-Schmidt, P., Sundé, J., Yildirim, Y., Tibblin, P., Forsman, A., & Larsson, P. (2019). Genetic differentiation between and within ecotypes of pike (*Esox lucius*) in the Baltic Sea. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29 , 1923-1935. doi:10.1002/aqc.3196
- Nosil, P. (2012). *Ecological speciation*. Oxford, UK: Oxford University Press.

- Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, *62* (2), 316-336. doi:10.1111/j.1558-5646.2007.00299.x
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2019). *vegan: Community Ecology Package*. R package version 2.5-5. Retrieved from: <https://CRAN.R-project.org/package=vegan>.
- Olsson, J. (2019). Past and current trends of coastal predatory fish in the Baltic Sea with a focus on perch, pike, and pikeperch. *Fishes*, *4* (1), 7. doi:10.3390/fishes4010007
- Olsson, J., Florin, A.-B., Mo, K., Aho, T., & Ryman, N. (2012). Genetic structure of whitefish (*Coregonus maraena*) in the Baltic Sea. *Estuarine, Coastal and Shelf Science*, *97* , 104-113. doi:<https://doi.org/10.1016/j.ecss.2011.11.032>
- Olsson, J., Mo, K., Florin, A. B., Aho, T., & Ryman, N. (2011). Genetic population structure of perch *Perca fluviatilis* along the Swedish coast of the Baltic Sea. *Journal of Fish Biology*, *79* (1), 122-137. doi:10.1111/j.1095-8649.2011.02998.x
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, *8* (10), 1360-1373. doi:10.1111/2041-210X.12775
- Petit, R. J., Aguinalalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., . . . Vendramin, G. G. (2003). Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, *300* (5625), 1563-1565. doi:10.1126/science.1083264
- Pierce, R. B., Tomcko, C. M., & Schupp, D. H. (1995). Exploitation of northern pike in seven small north-central Minnesota lakes. *North American Journal of Fisheries Management*, *15* (3), 601-609. doi:10.1577/1548-8675(1995)015<0601:eonpis>2.3.co;2
- Post, D. M., Palkovacs, E. P., Schielke, E. G., & Dodson, S. I. (2008). Intraspecific variation in a predator affects community structure and cascading trophic interactions. *Ecology*, *89* (7), 2019-2032. doi:10.1890/07-1216.1
- Puebla, O., Bermingham, E., & McMillan, W. O. (2012). On the spatial scale of dispersal in coral reef fishes. *Molecular Ecology*, *21* (23), 5675-5688. doi:10.1111/j.1365-294X.2012.05734.x
- R. (2012). *R: A language and environment for statistical computing*. In. Vienna, Austria: R Foundation for Statistical Computing.
- Raat, A. J. P. (1988). Synopsis of biological data on the northern pike *Esox lucius* . In: *FAO Fisheries Synopsis No. 30 Rev. 2* .
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197* (2), 573-589. doi:10.1534/genetics.114.164350
- Ravinet, M., Westram, A., Johannesson, K., Butlin, R., Andre, C., & Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, *25* (1), 287-305. doi:10.1111/mec.13332
- Reed, D. H., & Frankham, R. (2001). How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution*, *55* (6), 1095-1103. doi:10.1111/j.0014-3820.2001.tb00629.x
- Roff, D. A. (1997). *Evolutionary quantitative genetics* . New York: Chapman & Hall.
- Rondeau, E. B., Minkley, D. R., Leong, J. S., Messmer, A. M., Jantzen, J. R., von Schalburg, K. R., . . . Koop, B. F. (2014). The genome and linkage map of the Northern Pike (*Esox lucius*): Conserved syntenic revealed between the salmonid sister group and the Neoteleostei. *Plos One*, *9* (7). doi:10.1371/journal.pone.0102089

- Rosenbaum, H. C., Pomilla, C., Mendez, M., Leslie, M. S., Best, P. B., Findlay, K. P., . . . Kiszka, J. (2009). Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. *Plos One*, 4 (10), e7318. doi:10.1371/journal.pone.0007318
- Rowe, K. C., Heske, E. J., Brown, P. W., & Paige, K. N. (2004). Surviving the ice: Northern refugia and postglacial colonization. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (28), 10355-10359. doi:10.1073/pnas.0401338101
- RStudio. (2015). RStudio: Integrated development for R. In R. S. Team (Ed.), (1.1.838 ed.). Boston, MA: RStudio, Inc.
- Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8 (3), 336-352. doi:10.1111/j.1461-0248.2004.00715.x
- Saisa, M., Salminen, M., Koljonen, M.-L., & Ruuhijarvi, J. (2010). Coastal and freshwater pikeperch (*Sander lucioperca*) populations differ genetically in the Baltic Sea basin. *Hereditas*, 147 (5), 205-214. doi:10.1111/j.1601-5223.2010.02184.x
- Schindler, D. E., Hilborn, R., Chasco, B., Boatright, C. P., Quinn, T. P., Rogers, L. A., & Webster, M. S. (2010). Population diversity and the portfolio effect in an exploited species. *Nature*, 465 (7298), 609-612. doi:10.1038/nature09060
- Schluter, D., & Rambaut, A. (1996). Ecological speciation in postglacial fishes [and Discussion]. *Philosophical Transactions: Biological Sciences*, 351 (1341), 807-814.
- Seddon, P. J., Armstrong, D. P., & Maloney, R. F. (2007). Developing the science of reintroduction biology. *Conservation Biology*, 21 (2), 303-312. doi:10.1111/j.1523-1739.2006.00627.x
- Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, 68 (1), 1-15. doi:10.1111/evo.12258
- Shine, R., Brown, G. P., & Phillips, B. L. (2011). An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences*, 108 (14), 5708-5711. doi:10.1073/pnas.1018989108
- Si, Y., Wen, H., Li, Y., He, F., Li, J., Li, S., & He, H. (2018). Liver transcriptome analysis reveals extensive transcriptional plasticity during acclimation to low salinity in *Cynoglossus semilaevis* . *BMC genomics*, 19 (1), 464. doi:10.1186/s12864-018-4825-4
- Skog, A., Vollestad, L. A., Stenseth, N. C., Kasumyan, A., & Jakobsen, K. S. (2014). Circumpolar phylogeography of the northern pike (*Esox lucius*) and its relationship to the Amur pike (*E. reichertii*). *Frontiers in Zoology*, 11 (1), 67. doi:10.1186/s12983-014-0067-8
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, 30 (9), 1312-1313. doi:10.1093/bioinformatics/btu033
- Stamatakis, A. (2016). The RAxML v8.2.X Manual. Heidelberg Institute for Theoretical Studies. Retrieved from: <https://cme.h-its.org/exelixis/resource/download/NewManual.pdf>.
- Stephenson, R. L. (1999). Stock complexity in fisheries management: a perspective of emerging issues related to population sub-units. *Fisheries Research*, 43 (1), 247-249. doi:10.1016/S0165-7836(99)00076-4
- Sunde, J., Larsson, P., & Forsman, A. (2019). Adaptations of early development to local spawning temperature in anadromous populations of pike (*Esox lucius*). *Bmc Evolutionary Biology*, 19 (1), 148. doi:10.1186/s12862-019-1475-3
- Sunde, J., Tamario, C., Tibblin, P., Larsson, P., & Forsman, A. (2018). Variation in salinity tolerance between and within anadromous subpopulations of pike (*Esox lucius*). *Scientific Reports*, 8 (22). doi:10.1038/s41598-017-18413-8

- Sunde, J., Yildirim, Y., Tibblin, P., Bekkevold, D., Skov, C., Nordahl, O., . . . Forsman, A. *Data from: Neutral and adaptive differentiation in pike (Esox lucius) populations from contrasting environments* .
- Sunde, J., Yildirim, Y., Tibblin, P., & Forsman, A. (2020a). Comparing the performance of microsatellites and RADseq in population genetic studies: analysis of data for pike (*Esox lucius*) and a synthesis of previous studies. *Frontiers in Genetics*, *11* (218). doi:10.3389/fgene.2020.00218
- Sunde, J., Yildirim, Y., Tibblin, P., & Forsman, A. (2020b). *Data from: Comparing the performance of microsatellites and RADseq in population genetic studies: analysis of data for pike (Esox lucius) and a synthesis of previous studies* .
- Tamario, C., Sunde, J., Petersson, E., Tibblin, P., & Forsman, A. (2019). Ecological and evolutionary consequences of environmental change and management actions for migrating fish: a review. *Frontiers in Ecology and Evolution*, *7* (271). doi:10.3389/fevo.2019.00271
- Tibblin, P., Berggren, H., Nordahl, O., Larsson, P., & Forsman, A. (2016). Causes and consequences of intra-specific variation in vertebral number. *Scientific Reports*, *6* , 26372 doi:10.1038/srep26372
- Tibblin, P., Forsman, A., Borger, T., & Larsson, P. (2016). Causes and consequences of repeatability, flexibility and individual fine-tuning of migratory timing in pike. *The Journal of animal ecology*, *85* (1), 136-145. doi:10.1111/1365-2656.12439
- Tibblin, P., Forsman, A., Koch-Schmidt, P., Nordahl, O., Johannessen, P., Nilsson, J., & Larsson, P. (2015). Evolutionary divergence of adult body size and juvenile growth in sympatric subpopulations of a top predator in aquatic ecosystems. *The American Naturalist*, *186* (1), 98-110.
- Tinnert, J., Hellgren, O., Lindberg, J., Koch-Schmidt, P., & Forsman, A. (2016). Population genetic structure, differentiation, and diversity in *Tetrix subulatapygmy* grasshoppers: roles of population size and immigration. *Ecology and Evolution*, *6* (21), 7831-7846. doi:10.1002/ece3.2520
- van Strien, M. J., Holderegger, R., & Van Heck, H. J. (2015). Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity*, *114* (1), 27-37. doi:10.1038/hdy.2014.62
- Verhoeven, K. J. F., Macel, M., Wolfe, L. M., & Biere, A. (2011). Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proceedings of the Royal Society B*, *278* (1702), 2-8. doi:10.1098/rspb.2010.1272
- Waldman, J., Wilson, K. A., Mather, M., & Snyder, N. P. (2016). A resilience approach can improve anadromous fish restoration. *Fisheries*, *41* (3), 116-126. doi:10.1080/03632415.2015.1134501
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23* (23), 5649-5662. doi:10.1111/mec.12938
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38* (6), 1358-1370. doi:10.2307/2408641
- Wennersten, L., & Forsman, A. (2012). Population-level consequences of polymorphism, plasticity and randomized phenotype switching: a review of predictions. *Biological Reviews*, *87* , 756-767. doi:10.1111/j.1469-185X.2012.00231.x
- Wennerstrom, L., Olsson, J., Ryman, N., & Laikre, L. (2016). Temporally stable, weak genetic structuring in brackish water northern pike (*Esox lucius*) in the Baltic Sea indicates a contrasting divergence pattern relative to freshwater populations. *Canadian Journal of Fisheries and Aquatic Sciences*, *74* (4), 562-571. doi:10.1139/cjfas-2016-0039
- Westin, L., & Limburg, K. E. (2002). Newly discovered reproductive isolation reveals sympatric populations of *Esox lucius* in the Baltic. *Journal of Fish Biology*, *61* (6), 1647-1652. doi:10.1111/j.1095-8649.2002.tb02505.x

- Whitlock, R., Stewart, G. B., Goodman, S. J., Pierny, S. B., Butlin, R. K., Pullin, A. S., & Burke, T. (2013). A systematic review of phenotypic responses to between-population outbreeding. *Environmental Evidence*, 2 (1), 13. doi:10.1186/2047-2382-2-13
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis* : Springer.
- Willing, E.-M., Bentzen, P., Van Oosterhout, C., Hoffmann, M., Cable, J., Breden, F., . . . Dreyer, C. (2010). Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Molecular Ecology*, 19 (5), 968-984. doi:10.1111/j.1365-294X.2010.04528.x
- Wright. (1943). Isolation by distance. *Genetics*, 28 (2), 114-138.
- Wright, Bishop, J. M., Matthee, C. A., & von der Heyden, S. (2015). Genetic isolation by distance reveals restricted dispersal across a range of life histories: implications for biodiversity conservation planning across highly variable marine environments. *Diversity and Distributions*, 21 (6), 698-710. doi:10.1111/ddi.12302
- Yıldırım, Y., Anderson, M. J., Hansson, B., Patel, S., Millar, C. D., & Rainey, P. B. (2018a). Genetic structure of the grey side-gilled sea slug (*Pleurobranchaea maculata*) in coastal waters of New Zealand. *Plos One*, 13 (8), e0202197. doi:10.1371/journal.pone.0202197
- Yıldırım, Y., Tinnert, J., & Forsman, A. (2018b). Contrasting patterns of neutral and functional genetic diversity in stable and disturbed environments. *Ecology and Evolution*, 8 (23), 12073-12089. doi:10.1002/ece3.4667
- Zhan, Y., Li, J., Sun, J., Zhang, W., Li, Y., Cui, D., . . . Chang, Y. (2019). The impact of chronic heat stress on the growth, survival, feeding, and differential gene expression in the sea urchin *Strongylocentrotus intermedius*. *Frontiers in Genetics*, 10 (301). doi:10.3389/fgene.2019.00301

Table 1. Genetic diversity indices based on the full and adaptive subset datasets for 234 individuals from 11 pike, *Esox lucius*, populations (**Figure 2**).

						Full dataset	Full dataset
Population	N	Coordinates	Ecotype	Salinity (psu)	Seq. run	PA	H_O (SE)
Ängerån	24	63.63132, 19.74329	Anadromous	3	2018	9	0.25899 (0.00470)
Yngern	23	59.135795, 17.419853	Freshwater	0	2019	63	0.25367 (0.00415)
Snäckstavig	26	59.121394, 17.774334	Anadromous	6	2018	7	0.22297 (0.00457)
Hamnaryd	27	57.686717, 14.845233	Freshwater	0	2018	137	0.21858 (0.00358)
Lervik	21	57.072796, 16.522322	Anadromous	7	2018	4	0.22658 (0.00460)
Okne	22	57.01936, 16.449451	Anadromous	7	2018	2	0.23646 (0.00452)
Harfjärden	22	56.816595, 16.812537	Anadromous	7	2018	8	0.17143 (0.00425)
Kosta	14	56.841368, 15.406008	Freshwater	0	2018	131	0.26200 (0.00348)
Vambåsa	8	56.174753, 15.44991	Anadromous	8	2018	42	0.24444 (0.00429)
Steg Nor	23	54.978127, 12.294718	Resident	12	2019	40	0.27652 (0.00503)
Askeby	25	54.942353, 12.163566	Anadromous	12	2019	36	0.25061 (0.00490)

N : number of individuals sampled. Seq. run: samples sequenced in first (2018) or second (2019) run. PA : private alleles. H_O : observed heterozygosity. H_E : expected heterozygosity. F_{is} : fixation index. SE: standard error.

Table 2. Annotated overlapping outlier SNPs (BayeScan $q < 0.05$; Fdist and LOSITAN $P < 0.01$) for locus-specific effects (populations introduced as separate groups). '-' indicates that the annotation did not identify any known gene for the loci.

Accession	bp	LG	Gene	Gene ID	Selection
NC_025971.3	706411	LG04	vesicle-associated membrane protein 7	LOC105007731	diversifying
NC_025971.3	706570	LG04	vesicle-associated membrane protein 7	LOC105007731	diversifying
NC_025971.3	23540277	LG04	midline 2, transcript variant X3	mid2	diversifying
NC_025972.3	7499179	LG05	-	-	diversifying
NC_025974.3	28719593	LG07	opioid-binding protein/cell adhesion molecule	LOC105027132	diversifying
NC_025974.3	35934280	LG07	-	-	diversifying
NC_025975.3	5354935	LG08	ras-related protein Rab-6B	LOC105029536	diversifying
NC_025976.3	1897444	LG09	-	-	diversifying
NC_025977.3	24790442	LG10	zinc finger protein 436-like	LOC105008401	diversifying
NC_025978.3	12984145	LG11	ATP citrate lyase/ATP-citrate synthase	acly	diversifying
NC_025978.3	20568105	LG11	-	-	diversifying
NC_025979.3	3809024	LG12	probable G-protein coupled receptor 61	LOC105013497	diversifying
NC_025985.3	13795914	LG18	adhesion G protein-coupled receptor G6	adgrg6	diversifying
NC_025986.3	35565803	LG19	unconventional myosin-Ixa	LOC105018477	diversifying
NC_025988.3	15770522	LG21	-	-	diversifying
NW_017859398.1	1628416	unplaced	-	-	diversifying
NW_017859555.1	12358	unplaced	uncharacterized	LOC109615586	diversifying

Accession: genomic accession number on NCBI. bp: base position. LG: linkage group ('unplaced' indicates unplaced scaffolds).

Table 3. Annotated outlier SNPs for BayeScEnv ($q < 0.05$) with environmental variables midrange salinity and latitude. '-' indicates that the annotation did not identify any known gene for the loci.

Accession	bp	LG	Gene	Gene ID	Se
NC_025971.3	706411	LG04	vesicle-associated membrane protein7	LOC105007731	div
NC_025971.3	706570	LG04	vesicle-associated membrane protein7	LOC105007731	div
NC_025973.3	28351759	LG06	G-protein coupled receptor family C group 5 memberC	LOC105027411	bal
NC_025984.3	22297644	LG17	-	-	bal
NC_025984.3	26497648	LG17	store-operated calcium entry regulator STIMATE	LOC105017311	bal
NC_025985.3	13795914	LG18	adhesion G protein-coupled receptor G6	adgrg6	div
NC_025986.3	35565803	LG19	unconventional myosin-Ixa	LOC105018477	div
NC_025990.3	6202544	LG23	-	-	bal
NW_017858932.1	215052	unplaced	potassium voltage-gated channel subfamily A member10	LOC105008161	div
NW_017859115.1	496559	unplaced	-	-	div
NW_017859115.1	784614	unplaced	-	-	bal
NW_017859554.1	82291	unplaced	zinc finger protein OZF-like/zinc finger protein 260-like	LOC105008854	bal
NW_017859578.1	1052911	unplaced	-	-	bal

Accession: genomic accession number on NCBI. bp: base position. LG: linkage group ('unplaced' indicates unplaced scaffolds).

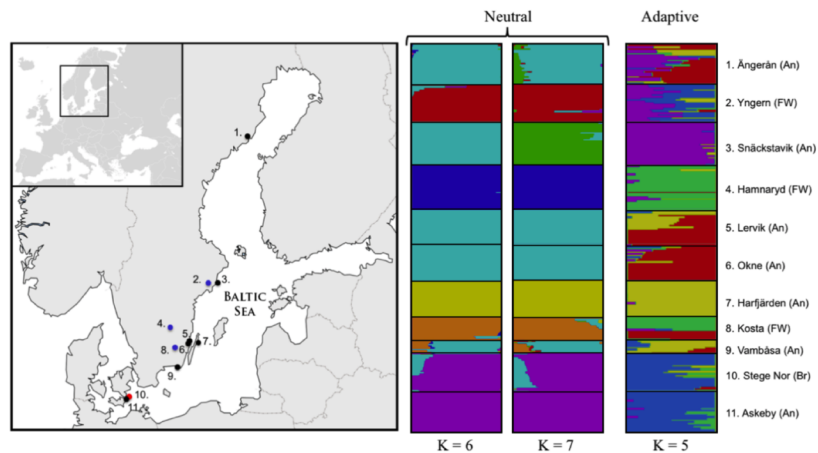


Figure 1. Map of study area, and genetic structuring among the pike, *Esox lucius*, populations. The map shows the locations and spawning ecotypes of the study populations (blue: freshwater, black: anadromous, and red: resident). It was created in Adobe Photoshop CC, v. 2015.0.1, and is a combined and modified version of two base maps (one of Scandinavia, and one of Sweden) that are available from Wikimedia Commons under the non-restrictive creative commons license. The distruct plots show the genetic structuring among the populations for the most likely number of genetic clusters (K) suggested by fastSTRUCTURE. The fastSTRUCTURE analyses were run with two different datasets (one 'neutral' and one 'adaptive') with data for 234 individuals.

Hosted file

image2.emf available at <https://authorea.com/users/362684/articles/483757-neutral-and-adaptive-differentiation-in-pike-esox-lucius-populations-from-contrasting-environments>

Figure 2. Pairwise F_{ST} -values between pike, *Esox lucius*, populations (ordered based on latitude). $P < 0.0001$ for all comparisons.

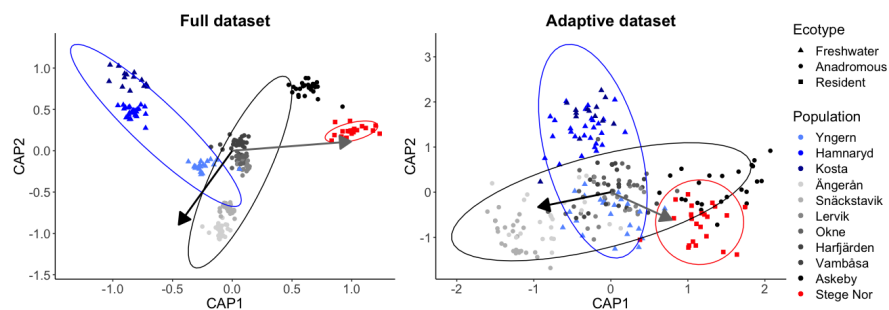


Figure 3. Visualization of distance based redundancy analyses (db-RDAs) of pike, *Esox lucius*, based on a) the full dataset, and b) a subset adaptive dataset consisting of the 17 loci identified in the outlier analyses as putatively under selection. Shape and color of dots indicate ecotype (blue pyramids: freshwater, grey circles: anadromous, and red squares: resident), and different shades within colors indicate populations in latitudinal order from lower (pale) to higher (dark). The black arrows indicate the effect of latitude ($P < 0.001$), and grey arrows the effect of midrange salinity ($P < 0.001$).

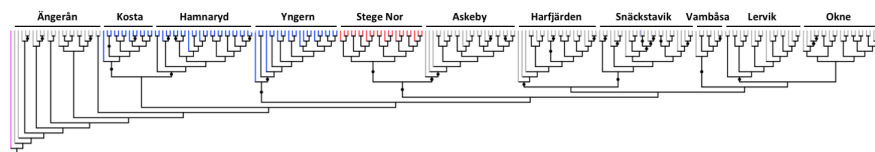


Figure 4. RAxML maximum likelihood (ML) tree of pike, *Esox lucius*, phylogeny based on the full dataset (created with iTOL, edited in photoshop Adobe Photoshop CC, v. 2015.0.1). Colors indicate ecotype (blue: freshwater, grey: anadromous, and red: resident). Rooted with the reference genome (indicated in pink) from NCBI (Rondeau et al., 2014). Black filled circles indicate nodes with bootstrap values ([?]/70).

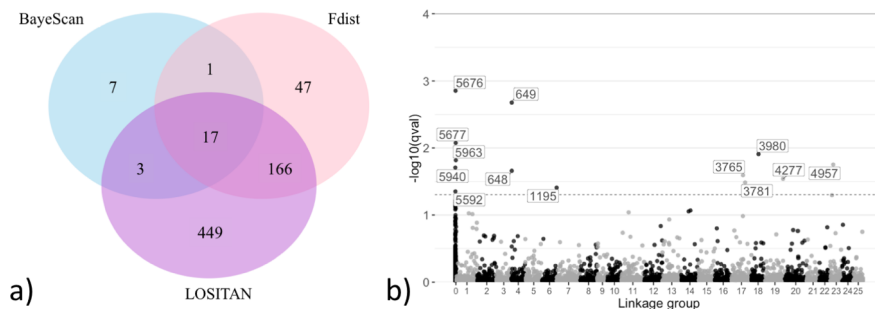


Figure 5. Loci identified as putatively under selection in the outlier analyses of pike, *Esox lucius*. a) Venn-diagrams that shows the number of loci identified by the three different software (BayeScan, Fdist and LOSITAN). BayeScan q -value < 0.05, and Fdist and LOSITAN P < 0.01. b) Results for BayeScEnv with midrange salinity and latitude introduced as environmental variables, significance threshold q -value < 0.05 (indicated by the dashed line).