


Whole-genome sequencing illuminates multifaceted targets of selection to humic substances in Eurasian perch

Mikhail Ozerov^{1,2,3}  | Kristina Noreikiene⁴ | Siim Kahar⁴ | Magnus Huss⁵ | Ari Huusko⁶ | Toomas Kõiv⁷ | Margot Sepp⁷ | María-Eugenia López¹ | Anna Gårdmark⁵ | Riho Gross⁴ | Anti Vasemägi^{1,4}

¹Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Drottningholm, Sweden

²Department of Biology, University of Turku, Turku, Finland

³Biodiversity Unit, University of Turku, Turku, Finland

⁴Chair of Aquaculture, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, Tartu, Estonia

⁵Department of Aquatic Resources, Swedish University of Agricultural Sciences, Öregrund, Sweden

⁶Natural resources Institute Finland (Luke), Paltamo, Finland

⁷Chair of Hydrobiology and Fishery, Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Tartu, Estonia

Correspondence

Anti Vasemägi, Swedish University of Agricultural Sciences, Department of Aquatic Resources, Institute of Freshwater Research, Stångholmsvägen 2, Drottningholm 17893, Sweden.
Email: anti.vasemagi@slu.se

Funding information

European Regional Development Fund and the program Mobilitas Plus, Grant/Award Number: MOBJD344; Ella ja Georg Ehrnroothin Säätiö; Swedish University of Agricultural Sciences; Vetenskapsrådet, Grant/Award Number: 2020-03916; Eesti Teadusagentuur, Grant/Award Number: PRG852 and PUTJD954

Handling Editor: Philine Feulner

Abstract

Extreme environments are inhospitable to the majority of species, but some organisms are able to survive in such hostile conditions due to evolutionary adaptations. For example, modern bony fishes have colonized various aquatic environments, including perpetually dark, hypoxic, hypersaline and toxic habitats. Eurasian perch (*Perca fluviatilis*) is among the few fish species of northern latitudes that is able to live in very acidic humic lakes. Such lakes represent almost “nocturnal” environments; they contain high levels of dissolved organic matter, which in addition to creating a challenging visual environment, also affects a large number of other habitat parameters and biotic interactions. To reveal the genomic targets of humic-associated selection, we performed whole-genome sequencing of perch originating from 16 humic and 16 clear-water lakes in northern Europe. We identified over 800,000 single nucleotide polymorphisms, of which >10,000 were identified as potential candidates under selection (associated with >3000 genes) using multiple outlier approaches. Our findings suggest that adaptation to the humic environment may involve hundreds of regions scattered across the genome. Putative signals of adaptation were detected in genes and gene families with diverse functions, including organism development and ion transportation. The observed excess of variants under selection in regulatory regions highlights the importance of adaptive evolution via regulatory elements, rather than via protein sequence modification. Our study demonstrates the power of whole-genome analysis to illuminate the multifaceted nature of humic adaptation and provides the foundation for further investigation of causal mutations underlying phenotypic traits of ecological and evolutionary importance.

KEYWORDS

Candidate gene, DOC, fish, humic adaptation, SNP, water colour

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Most organisms cannot live in extreme environments, but some are able to thrive in such challenging habitats due to evolutionary changes driven by natural selection. This has fuelled a long-standing interest in how organisms cope with and adapt to extreme conditions (Riesch et al., 2015). Modern bony fishes belong to the most species-rich clade of vertebrates, consisting of more than 30,000 described species (Nelson, 2006), many of which have colonized the most extreme types of aquatic environments, including caves and deserts, and deep-sea, high-altitude, hypoxic, temporary, hypersaline and toxic habitats. In recent years, fish living in extreme habitats have become emerging model species for studying adaptation and the predictability of evolution (Riesch et al., 2014; Tobler & Plath, 2011). For example, genome-wide studies are starting to reveal the molecular mechanisms of adaptation linked to various abiotic factors, such as salinity (Dalongeville et al., 2018; Garcia-Elfring et al., 2021), light (Marques et al., 2017), toxins (Pfenninger et al., 2014; Reid et al., 2016), and acidic (Haenel et al., 2019) and alkaline environments (Tong & Li, 2020; Xu et al., 2017). Over the last century we have also witnessed the expansion of hypoxic zones coinciding with eutrophication and extreme heat waves (Altieri & Gedan, 2015; Breitbart et al., 2018), as well as acidification (Hannan & Rummer, 2018; Tagliarolo, 2019) and increase of coloration (e.g., brownification) of lakes and rivers in the northern hemisphere (Creed et al., 2018; Evans et al., 2005; Forsberg, 1992; Kritzberg et al., 2020; Roulet & Moore, 2006; Solomon et al., 2015; Vuorenmaa et al., 2006). However, the evolutionary consequences of these and other rapid human-induced environmental changes are currently not well understood (Grummer et al., 2019).

Organic matter is a complex mixture of organic molecules originating from the decay of plant and animal debris. Most aquatic organic matter occurs in the dissolved form. Humic and fulvic acids account for the majority of dissolved organic matter in many lake ecosystems (McKnight & Aiken, 1998), and their presence in water is observable as a yellowish or brownish coloration (Bricaud et al., 1981). Due to the heterogeneity of dissolved organic matter, it is usually quantified as dissolved organic carbon (DOC; Wood et al., 2011). DOC plays a significant role in freshwater ecosystems by driving the carbon and energy cycle, and also contributes significantly to greenhouse gas emissions (Battin et al., 2009; Sobek et al., 2006; Tranvik et al., 2018). Humic (also known as dystrophic) lakes containing high levels of DOC are extreme, almost “nocturnal” visual environments—both down-welling short-wavelength and almost all up- or side-welling light is absorbed, while only long-wavelength red light is able to partially penetrate the water column (Eloranta, 1978; Jones, 1992). A striking example of the evolutionary adaptation to long-wavelength visual environments has been described in three-spined stickleback (*Gasterosteus aculeatus*), where nonsynonymous mutations observed in the SWS2 opsin gene have caused a red-shift in light absorption of the visual photo pigment (Marques et al., 2017). Adaptation to extreme visual environments may also include

adjustments at the gene expression level, and/or gene duplications (Carleton et al., 2020; Musilova et al., 2019).

In addition to the visual environment, DOC affects a myriad of other habitat parameters and biotic interactions. Importantly, a high level of DOC contributes to acidification in low-alkalinity and weakly buffered waters (Hope et al., 1996; Sobek et al., 2003). Therefore, humic lakes generally have low pH levels (Arvola et al., 2010; Erlandsson et al., 2010; Keskitalo & Eloranta, 1999). Freshwater fish living in acidic water must constantly import ions from their food and environment to compensate for diffusive ion losses, primarily from the gills (Dymowska et al., 2012). Thus, it is probable that low pH represents a strong selective force and physiological challenge for teleosts in humic lakes, especially during early life stages (Parra & Baldisserotto, 2007; Rask, 1984). On the other hand, DOC may also partially protect against acidic water by altering gill membrane permeability and stimulating ion uptake (Wood et al., 2011). Acidification and increased DOC can, in turn, affect ion composition in humic lakes, where the concentration of essential ions, including calcium, is very low (Weyhenmeyer et al., 2019). Calcium plays an important role in many cellular functions, and is a critical component of body structures, such as bones, carapaces, scales and shells (Greenaway, 1985). Moreover, humic lakes usually exhibit steep thermal and oxygen stratification, leading to hypoxia in deeper areas (Bastviken et al., 2004; Kankaala et al., 2006). Hence, increased DOC can also limit whole-lake primary production and underwater higher vegetation, supporting the microbial loop and influencing resource availability and quality for consumers (Ask et al., 2009; Cole, 2009; Karlsson et al., 2009). Thus, DOC influences whole lake food-webs from primary producers to top predators and parasites (Grether et al., 2001; Tobler & Path, 2011; Kaeuffer et al., 2012; Noreikiene et al., 2020).

Total biomass production, body growth and mean body size of Eurasian perch (*Perca fluviatilis*) have been shown to differ between humic and nonhumic lakes (van Dorst et al., 2019). Therefore, selective agents in humic lakes are probably multifaceted and may include multiple factors, such as biotic, abiotic, and interactions within and between them. Yet, the evolutionary responses to extreme environments have been traditionally studied by focusing on a single key abiotic factor (e.g. Garcia-Elfring et al., 2021; Haenel et al., 2019; Marques et al., 2017; Reid et al., 2016; Xu et al., 2017). Alternatively, instead of choosing a few a priori defined traits and genes, a more comprehensive view of the main mechanisms, molecular targets and selective agents could be achieved by characterizing the signatures of divergent selection based on analyses of whole genomes (Jones et al., 2012; Miller et al., 2019).

Eurasian perch is among the few fish species of northern latitudes able to live in acidic and humic lakes, and is also widely distributed from the Iberian Peninsula to the River Kolyma in Siberia (Collette & Bănărescu, 1977). Perch are abundant in both humic and clear-water lakes, and represent an important component in freshwater food-webs; it acts as a key predator, and is also an important prey species for birds and other fishes (Diehl, 1992). Recent work has shown that perch in humic lakes play a significant role in regulating

bacterial abundance via feeding on zooplankton, which affects carbon cycling via methane efflux (Devlin et al., 2015). However, the evolutionary mechanisms and molecular processes that allow perch to inhabit these extreme conditions remain unknown. Thus, dissecting the molecular mechanisms of humic adaptation will not only help to obtain an unbiased view of the molecular targets and main mechanisms of adaptation, but may also shed light on the evolutionary changes related to brownification (Creed et al., 2018), with implications for overall food web dynamics and greenhouse gas (methane) emission rates from lakes through trophic cascades. Given that humic lakes are very common around the world (Wetzel, 2001) and up to one third of known freshwater fish species are associated with tropical peat swamps containing high levels of DOC (Ng et al., 1994; Posa et al., 2011), understanding the nature of humic adaptation has a broad significance across species and geographical regions.

In this study, we characterize shared genomic signatures of adaptation to extreme humic environments using whole-genome sequencing of perch sampled from 16 humic and 16 clear-water lakes in northern Europe. We predicted that this adaptive process may involve both well-characterized genes with known function (e.g., visual performance, maintenance of ionic balance), as well as novel targets. By using over 800,000 high-quality single nucleotide polymorphisms (SNPs) and applying genome-wide divergence and environmental association approaches, we aimed to ascertain whether adaptation occurs by major shifts of allele frequencies in a few key genes (e.g., visual opsins and ion transporter genes), or if footprints of selection are scattered across the genome and include many regions and genes with diverse functions. In addition, we tested if signatures of selection are randomly distributed across the genome

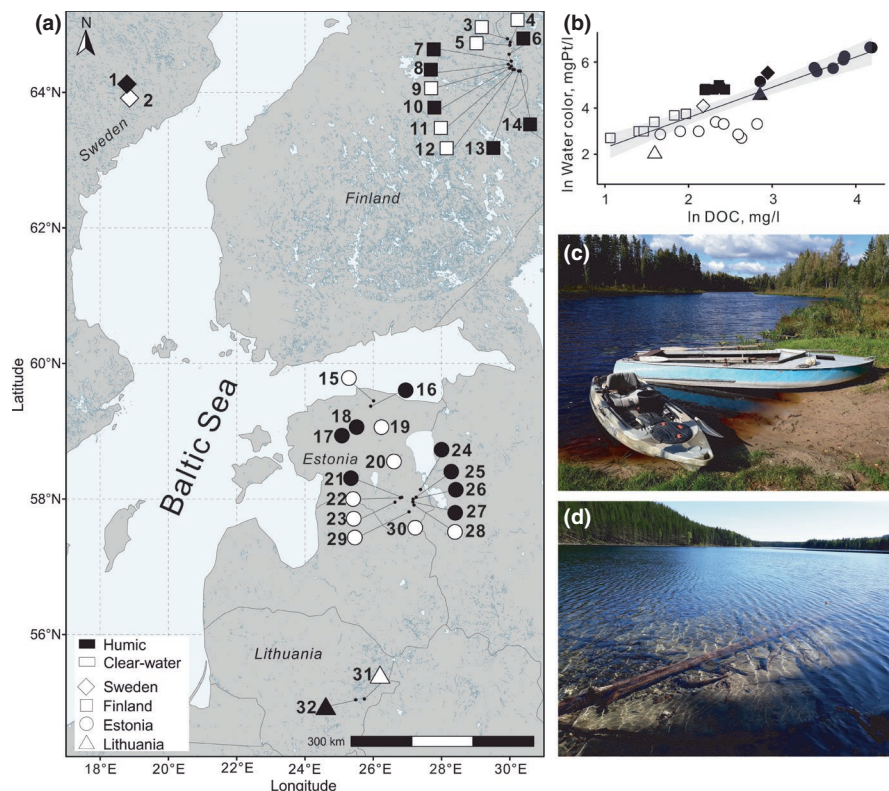
or if they are concentrated within or near coding and regulatory regions. Finally, we evaluated if genes putatively under divergent selection between humic and clear-water environments are associated with specific biological processes and molecular functions using Gene Ontology (GO) analysis.

2 | MATERIAL AND METHODS

2.1 | Biological samples

In total, 32 perch individuals were sampled from 16 humic and 16 clear-water lakes in northern Europe—a single specimen per lake, as in Jones et al. (2012) (Figure 1a; Table S1). The selection of lakes for the analysis was based on drastic differences in water colour (median_{HUMIC} = 212.5; range = 95.0–765.0 vs. median_{CLEAR} = 20.0; range = 7.5–60.0, nonparametric Mann–Whitney test $p < .001$, Figure S1) and geographical proximity of different type of lakes to increase the power of detecting loci under selection and to minimize the effect of population structuring (De Mita et al., 2013; Hoban et al., 2016; Lotterhos & Whitlock, 2015). Most fish were caught using a gill-net, beach seine or rod (Table S1). Egg ribbons were collected from two Swedish lakes (Nedre Björntjärnen and Snottertjärn) and transported to laboratory facilities (Institute of Freshwater Research, Drottningholm, Sweden). After hatching, fry were killed using a benzocaine overdose and stored in 96% ethanol without measuring fork length or body mass or determining sex. The fish from the remaining lakes were killed with a sharp blow to the head. Thereafter, fork or total length (to the nearest mm) and wet body mass (to the nearest g) were measured, sex was

FIGURE 1 (a) Map indicating sampling locations. (b) Dot plot showing the relationships between DOC and water colour values (ln-transformed) in each lake. Regression is shown as a black solid line; 95% confidence interval is shaded in grey. Photographs illustrate (c) humic and (d) clear-water habitat. Black- and white-filled points represent humic and clear-water populations, respectively. The country of origin is depicted by different shapes as shown in the key. Population ID numbers follow those given in Table S1



determined by visual examination of the gonads, and a tissue sample (pelvic fin) was placed in 96% ethanol for DNA extraction.

The requirements outlined in Annex III (Requirements for establishments and for the care and accommodation of animals) and Annex IV (Methods of killing animals) Section B point 11 of the "Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purpose" were fully met. The authors have followed the principles of the 3Rs (replacement, reduction and refinement) and have involved the minimum number of animals to produce statistically reproducible results. The collection of samples was conducted in accordance with national legislation based on permits issued by the Estonian Ministry of Environment (54/2016) and the regional ethical review board in Uppsala, Sweden (Dnr 5.8.18-03449/2017). Sampling in Finland and Lithuania was carried out using recreational fishing gear (rod and line), following the fishing rules set by the national legislations, and ethical issues related to recreational fishing and handling of caught fish recommended by the recreational fishing federations of the countries.

2.2 | Environmental data

Lake water was collected at the time of fish sampling for chemical analysis. Two measures reflecting the amount of dissolved organic matter in the water and the visual environment in lakes—DOC (mg L^{-1}) and water colour (mg Pt L^{-1})—were analysed as described in Sepp et al. (2019). The latter parameter is also influenced by iron concentrations in the water (Lei et al., 2020; Maloney et al., 2005; Weyhenmeyer et al., 2014). Clear-water lakes have been oligotrophic, alcalitrophic or developed to mesotrophic lakes since melting of the last ice sheet. The humic lakes, on the other hand, formed after the last glaciation when organic matter (carbon) from terrestrial (allochthonous) origin entered lakes from surrounding soils and vegetation (Tranvik, 2009). The increase of DOC in lakes, however, has not been linear over time since these changes are ultimately climate-driven, reflecting both climate and lake-catchment coupling processes (e.g., Rantala et al., 2015).

Lake surface area and lake shoreline length were manually estimated using satellite photos on Google maps (<https://www.google.com/maps/>; Table S1) to estimate the size of the lakes. Both DOC and water colour values were log-normalized prior to environmental association analyses. DOC and water colour were highly correlated (Pearson's $r = .96$, $p < .001$, Figure 1b), and both showed significant differences between humic and clear-water lakes (nonparametric Mann-Whitney test, both $p < .001$; Figure S1). Neither lake surface area nor shoreline distance differed between humic and clear-water lakes (Figure S1).

2.3 | DNA extraction and sequencing

Total genomic DNA (gDNA) was extracted from the tissue samples using a NucleoSpin Tissue kit (Macherey-Nagel) according to

the manufacturer's protocol. The quality of the total gDNA was assessed using a Fragment Analyzer (Advanced Analytical), and the concentration was measured using Qubit fluorometric quantification (Life Technologies). Sequencing libraries (average insert size 350 bp) were constructed from 1 μg of gDNA for each individual according to the Illumina TruSeq DNA PCR-Free Library Preparation Guide (part #15036187). A unique Illumina TruSeq indexing adapter was ligated to each gDNA sample. The samples were normalized and pooled for automated cluster preparation using an Illumina cBot station. Sixteen libraries (each corresponding to one perch sample from Estonia) were pooled and sequenced in eight lanes on an Illumina HiSeq 3000 using paired-end sequencing (2×150 -bp read length with 8-bp index), and another 16 libraries (each corresponding to one perch sample from Finland, Lithuania and Sweden) were pooled and sequenced in four lanes on an Illumina NovaSeq 6000 using paired-end sequencing (2×150 -bp read length with 8-bp index).

2.4 | SNP calling and filtering

Read quality was assessed using FASTQC version 0.11.8 (Andrews, 2017). Illumina adapters, as well as short (<60 bp) and low-quality reads (average quality score <25) were trimmed using TRIMMOMATIC version 0.36 (Bolger et al., 2014; CROP:149 HEADCROP:9 SLIDINGWINDOW:5:25 MINLEN:60) for the 16 samples sequenced on the HiSeq 3000 instrument, or FASTP version 0.20 (Chen et al., 2018) for the 16 samples sequenced on the NovaSeq 6000 instrument, with the same parameters (`-g -w 12 -r -W 5 -M 25 --trim_front1 9 --trim_front2 9 --trim_tail1 2 --trim_tail2 2 -l 60`) due to the excess of polyG tails in the latter (Arora et al., 2019; De-Kayne et al., 2021).

Filtered sequence reads of each individual were mapped to the Eurasian perch reference genome (NCBI: GCA_010015445.1) using BOWTIE2 version 2.3.5.1 (Langmead et al., 2009), applying default parameters except the modified score minimum threshold (`--score-min L,-0.3,-0.3`) and maximum fragment length for valid paired-end alignments (`-X 700`). Mean coverage (d) across all 32 perch individuals was 29.9, and varied from 24.0 to 40.6. Raw SNPs were called using two alternative pipelines. First, SAMTOOLS version 1.10 (Li, 2011) was applied on the locally realigned and sorted BAM files (`samtools mpileup -u -l -t DP,AD,INFO/AD,ADF,ADR,SP -q 20`) with the following variant calling using BCFTOOLS version 1.8 (Li, 2011). Second, the HaplotypeCaller subroutine from GATK version 4.1.4.1 (McKenna et al., 2010) with similar parameters was applied to call variants using the same BAM files (`-ERC GVCF --minimum-mapping-quality 20 -mbq 13 --indel-size-to-eliminate-in-ref-model 12 -G AS_StandardAnnotation -G StandardAnnotation`) with the following import of single-sample GVCF files into GenomicsDB using GenomicsDBImport, and final calling of consensus genotypes with GenotypeGVCFs. Further quality filtering of raw variants generated by the two pipelines was performed using VCFTOOLS version 0.1.15 (Danecek et al., 2011) retaining only the variants that met the following criteria:

(i) minimum and maximum mean sequencing depth (d) was set to 10 (as a minimum recommended by Illumina) and 66 (maximum depth = $d_{\max} + 4\sqrt{d_{\max}}$; Li, 2014), respectively; (ii) the consensus quality was ≥ 30 ; (iii) a variant had at least two copies of an allele; (iv) no missing data were allowed; (v) only bi-allelic sites were included; and (vi) a variant did not occur in repetitive genomic regions (--max-meanDP 66 --min-meanDP 10 --max-missing 1 --mac 2 --min-alleles 2 --max-alleles 2 --minQ 30 --exclude-bed Pfluv_repeats.bed).

In total, 1,078,577 and 1,074,879 SNPs were retained after applying the SAMTOOLS-BCFTOOLS and GATK pipelines, respectively. To ensure the quality and reliability of the data set, only the variants consistently called by both pipelines were retained (1,025,544 SNPs). To further exclude false heterozygous calls and regions with an excess of variable sites (such as multilocus contigs, paralogues, repetitive elements and otherwise superficially similar sequences; Hohenlohe et al., 2011; Willis et al., 2017; O'Leary et al., 2018), the additional quality control filtering was applied: number of heterozygotes per SNP ≤ 16 and MAF ≥ 0.05 , which resulted in 810,591 SNP loci in the final data set.

The identified SNPs were annotated using SnpEFF version 5.0 (Cingolani et al., 2012). The SnpEff database was generated using the Eurasian perch reference genome sequence and annotation (NCBI: GCA_010015445.1).

2.5 | Population diversity and structure

Observed heterozygosity per individual was calculated using VCFTOOLS version 0.1.15 (Danecek et al., 2011). The difference in observed heterozygosity between humic and clear-water lakes was tested using a nonparametric Mann-Whitney test (wilcox.test function in the basic R-package stats). The same package was used to examine linear relationships of observed heterozygosity and DOC, water colour, lake surface area and lake shoreline length with the cor.test function. The R package ADEGENET version 2.1.3 (Jombart, 2008; Jombart & Ahmed, 2011) was used to convert SNP data into a genind object. Overall population genetic structure was examined by applying principal component analysis (PCA) using the dudi.pca function of the ADE4 version 1.7-16 R-package (Dray & Dufour, 2007) on three sets of SNP loci: (i) all 810,591 SNPs; (ii) 256,880 intergenic SNPs located exclusively in the intergenic regions and assumed as putatively neutral (if a given SNP was linked to multiple annotations, it was not included to the intergenic SNP list; File S1); and (iii) 10,245 candidate SNPs potentially under natural selection (identified based on two or three outlier tests, see details below). The per cent of variation explained by each PC axis was extracted using the FACTOREXTRA version 1.0.7 R package (Kassambara & Mundt, 2020). Pairwise genetic differentiation among the samples was estimated as pairwise mean absolute allele frequency differences (D ; Prevosti et al., 1975). Genetic relationships among the samples were explored using Nei's genetic distances (Nei, 1972) in the R package POPPR version 2.8.2 (Kamvar et al., 2014, 2015); neighbour-joining trees (Saitou & Nei,

1987) were reconstructed with 100 bootstrap replicates among loci using the R-package APE version 5.4-1 (Paradis & Schliep, 2019).

2.6 | Signatures of selection

Candidate SNPs potentially under natural selection were identified using three approaches: one population divergence method and two environmental association methods.

2.6.1 | Highly divergent loci

The genetic divergence of each SNP locus was estimated as a mean absolute allele frequency difference (δ) between humic and clear-water lakes. Candidate SNPs were identified as SNPs with a δ that was higher than 2.5 SD (standard deviations) from the mean δ (δ threshold = 0.268; empirical two-tailed $p = .012$; Miller, 1991).

2.6.2 | Environmental association analysis

Redundancy analysis (RDA)

RDA is an extension of multiple linear regression (Legendre & Legendre, 2012) that compares a matrix of multiple response variables (i.e., allele frequencies) with multiple independent predictor variables (i.e., environmental variables). We analysed allele frequencies and environmental variables of each lake with linear regressions. Further, PCA was applied to constrain the fitted values of the regressions and to produce ordination axes, which are linear combinations of the original predictor variables. RDA was performed to test the effects of DOC and water colour on the estimated allele frequencies. Although both DOC and water colour estimates were highly correlated to each other (Pearson's $r = .96$, $p < .001$), this does not influence the ability of RDA to obtain a good fit nor does it influence the predictions and precision of the predictions (Kutner et al., 2005). Furthermore, we did not aim to distinguish the effects of DOC and water colour, but to predict their influence on genetic variation as a complex. The effect of genetic and spatial structure was controlled using PC1 loadings based on intergenic SNPs (Figure S3) and geographical coordinates (latitude and longitude) of the sampled lakes. For the RDA, we used the rda function of the R package VEGAN version 2.5-7 (Oksanen et al., 2020). Candidate SNPs were identified on each of the first two ordination axes as SNPs that had a "locus score" above $\pm 2.5 SD$ from the mean score for that axis (RDA loading threshold at axis 1 = ± 0.0202 and at axis 2 = ± 0.0196 ; empirical two-tailed $p = .012$; Dalongeville et al., 2018; Forester et al., 2016; Miller, 1991; Xuereb et al., 2018).

Latent factor mixed model (LFMM)

LFMMs are factor regression models in which allele-environment correlations are estimated between each locus and each environmental variable (Frichot & François, 2015; Frichot et al., 2013).

Environmental variables are considered as fixed effects, while population structure is modelled using a number of latent factors (K) included in the model as covariates (Frichot et al., 2013). The latent factors comprise the levels of population structure due to genetic variation or shared demographic history (Frichot et al., 2013). The analysis was performed using the LFMM version 1.0 package in R, implementing a least-squares approach for latent factor estimation (Caye et al., 2019). Here, we used $K = 2$ to correct for population structure inferred using the PCA of intergenic SNPs (see Results; Figure S3). The significance of the association was tested and p -values were calibrated using the genomic control method (Devlin & Roeder, 1999). Genomic control uses a robust estimate of the variance of z -scores called "genomic inflation factor." The p -value threshold ($p \leq .012$) for LFMM was chosen (above 2.5 SD) to correspond to the two other methods (RDA and δ). However, this does not necessarily mean that all three methods have similar levels of false discovery rate in terms of detecting loci under divergent selection.

2.7 | Candidate SNPs supported by multiple methods

The final set of candidate SNPs potentially under natural selection included 10,245 candidates, which were identified using at least two methods, which is expected to reduce the number of false positive loci in the final set and control for genetic and spatial structuring (Figure 2e). Based on the empirical p -value threshold of .012 and assuming that detection of candidate loci is independent in each outlier method, we expect 445 false positives (3.4%) among the identified candidate SNPs.

To test for excess and deficiency of identified candidate SNPs in each annotation category, a chi-squared test was performed using the `STATS` version 3.6.2 package in R to compare the 10,245 candidate SNPs with all identified SNPs. Genes that were located in genomic regions with at least one candidate SNP in exon, intron or regulatory sequences, including 5K up- and downstream regions, were considered as candidate genes.

In addition, we used the `CIRCLIZE` version 0.4.12 (Gu et al., 2014) package in R to visualize the distribution of all and candidate SNPs, as well as the genes, genomic divergence and diversity across the perch genome. The density of SNPs, genes and candidate SNPs across the genome was estimated at each chromosome based on nonoverlapped 500-kb windows using the `genomicDensity` function of the `CIRCLIZE` R package. In addition, the density of candidate SNPs was estimated in 25-kb windows using the same package. Genomic regions were identified as regions with a high density of candidate SNPs if the density estimate was higher than 2.5 SD from the mean density. The linear plots representing candidate regions affected by divergent selection were visualized using `KARIOPLOTER` version 1.16.0 (Gel & Serra, 2017).

2.8 | Gene Ontology analysis

For GO analysis, we first identified perch genes that are orthologous to human and zebrafish genes using the `RENTREZ` (Winter, 2017) package in R. GO enrichment analysis of candidate genes against all orthologous gene symbols as a background was performed using a binomial test in `PANTHER` (Thomas et al., 2003). The GO terms with a false discovery rate (FDR) ≤ 0.05 were considered as significant.

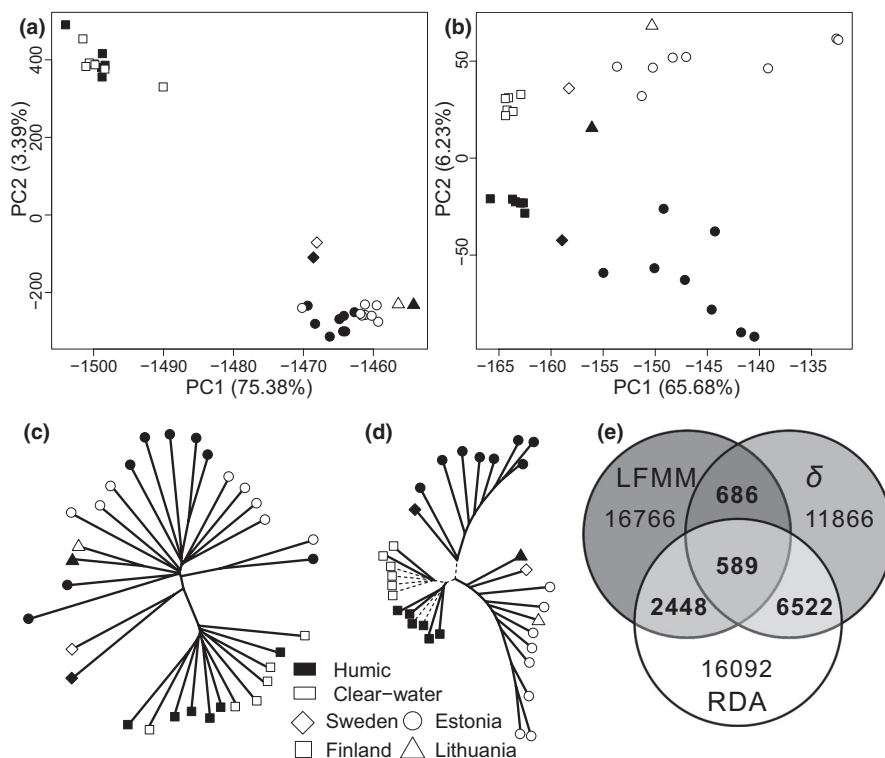


FIGURE 2 PCA summarizing the genetic structure for (a) 810,591 genome-wide SNPs and for (b) 10,245 candidate SNPs. Neighbour-joining dendrogram based on D_{Nei} genetic distances, demonstrating the genetic relationships among perch samples based on (c) 810,591 genome-wide SNPs and (d) 10,245 candidate SNPs. The branches with bootstrap value support $<80\%$ are represented as dashed lines. Humic and clear-water populations are indicated as black- and white-filled symbols, respectively. The country of origin is depicted by different shapes as shown in the key. (e) Venn diagram showing overlap among candidate SNPs revealed by three methods; the final set of candidate SNPs is highlighted in bold

3 | RESULTS

3.1 | Population genomic variation and structure

Overall genome-wide genetic diversity (estimated as observed heterozygosity, H_O) varied from 0.064 (ELOO) to 0.309 (EUJA) with a mean $H_O = 0.215$ (Table S1). Average genetic diversity among perch in humic lakes (mean $H_O = 0.195$, median $H_O = 0.221$) was slightly lower compared to that in clear-water lakes (mean $H_O = 0.236$, median $H_O = 0.240$), although the difference was not significant (Figure S2a). There was a weak, nonsignificant tendency for genetic diversity to decrease with increasing DOC and water colour, and to increase with lake size (Figure S2b–e).

Overall genome-wide genetic divergence (measured as D ; Prevosti et al., 1975) among the samples was $D = 0.309$, and ranged from $D = 0.204$ (FIMU vs. FHYV) to $D = 0.356$ (FITA vs. ELOO; Table S2). The first two PCA axes based on 810,591 SNPs explained nearly 79% of the genomic variation, and depicted a genetic structure corresponding to geographical origin of perch (Figure 2a). Similarly, the topology of the neighbour-joining tree based on D_{Nei} genetic distances reflected the main clusters identified by the PCA (Figure 2c). In general, perch samples clustered into the two main groups: (i) Finnish, including the samples from Finland, and (ii) South-Western Baltic, including perch from Sweden, Estonia and Lithuania.

3.2 | Candidate SNPs under divergent selection

The combination of genetic divergence and two environmental association analyses revealed 10,245 candidate SNPs (Figure 2e). Based on the PCA using these 10,245 candidate SNPs, 6% of the variation on the second axis was explained by humic vs. clear-water separation, whereas 65% was attributed to the geographical origin (Figure 2b). Similarly, the topology of the neighbour-joining tree based on D_{Nei} distances using 10,245 candidate SNPs showed the separation of humic and clear-water perch within the Finnish and South-Western Baltic groups (Figure 2d). The majority of candidate SNPs (8,271 SNPs; 80.7%) were polymorphic in both Finnish and South-Western Baltic groups, whereas a lower proportion of SNPs were monomorphic within Finnish (1920 SNPs; 18.7%) and South-Western Baltic fish (54 SNPs; 0.5%). Significant enrichment (chi-squared test: $\chi^2 = 11.55$ –85.50, $p = 2.3 \times 10^{-20}$ – 4.8×10^{-4}) and depletion (chi-squared test: $\chi^2 = 3.98$ –37.55, $p = 4.6 \times 10^{-2}$ – 8.9×10^{-10}) for candidate SNPs was observed in seven and 10 chromosomes, respectively (Figure 3; Table S3). Furthermore, the distribution of candidate SNPs across the chromosomes was not even; several peaks of high density were identified (Figure 3; File S2). The number of genomic regions with a high density of candidate SNPs varied from one to six (mean = 2.4) per chromosome with 500-kb windows or from 17 to 49 (mean = 31.4) per chromosome when 25-kb windows were used for density estimation. The total number of regions with a high density of candidate SNPs was 50 or 658 when using 500 or 25-kb windows, respectively (Table S4). Most of the

candidate SNPs were located in intergenic regions (32.3%) and introns (33.8%), whereas ~4% were found in exons (Table 1). However, the enrichment analysis indicated that the frequency of candidate SNPs in intergenic regions was lower compared to the rest of the genome (chi-squared test: $\chi^2 = 9.76$, $p = 1.9 \times 10^{-3}$; Table 1). On the other hand, the 5'UTR (untranslated region), 3'UTR and 5K downstream gene regions were enriched for candidate SNPs (chi-squared test: $\chi^2 = 8.88$ –5.42, $p = 2.0 \times 10^{-2}$ – 2.8×10^{-3} ; Table 1).

3.3 | Candidate genes under divergent selection

In total, 3,245 candidate genes were identified (Table S5). Since the strongest selective sweeps are expected to affect multiple adjacent SNPs, further ranking of the candidate gene list (average $\delta > 0.4$, >5 SNPs per gene) revealed 31 of the strongest candidates containing six to 38 outlier SNPs (Table S5). A large proportion of genes among these are involved in anatomical structure development (ASXL1, CDON, ECE1, FGF11, FLII, HOXB9, LRRN1, MYLIP, PLAGL2, RTN1, TDRD9, TLL3, TTN) and regulation of nervous system development (MYLIP, BHLHE40, CDON, FGF11, LRRN1, RTN1). The highest number of SNPs was observed in the MYLIP gene located on chromosome 13 (Figure 4c). This gene is involved in various biological processes, including multicellular organism development, regulation of plasma membrane-bound cell projection organization and nervous system development (Bornhauser et al., 2003; Calkin et al., 2011; Olsson et al., 1999). The functions of other strong candidate genes included regulation of voltage-gated sodium channel activity (FGF11, SLMAP), inorganic ion transmembrane transport (KCNMA1, TMEM163; Li et al., 2018) and regulation of circadian rhythm (BHLHE40; Honma et al., 2002).

We observed several 50–100-kb regions of elevated δ between humic and clear-water perch in multiple chromosomes, involving genes adjacent to the strongest candidates (Figure 4a,b,d). For example, the ~100-kb region on chromosome 5 included genes regulating anatomical structure development (ASXL1, MYT1L, PLAGL2), oxidative stress responding regulation (PLAGL2), response to oxygen-containing compound (ASXL1) and mitochondrial fusion (MFN2); the ~100-kb region on chromosome 7 contained genes regulating response to stress (RPA2, THEMIS2), lipid (CYP4B1) and glycogen (PPP1R8) metabolism; and the ~100-kb region on chromosome 15 involved genes regulating ion transport (LLGL1, SHISA9, DESI1) and cell differentiation (FLII, LLGL1). As expected, most of the regions showing elevated δ also showed a reduction in H_O among humic perch, suggesting that directional selection associated with high DOC content has reduced the genetic diversity of adjacent genomic regions.

In contrast to our previous expectations, SNPs found in visual opsin genes (orthologous to human *OPN1LW*, *OPN1MW*, *OPN1SW* and *RHO*) did not show high allele frequency differences between humic and clear-water populations (Figure S4), suggesting that the strongest selective sweeps associated with humic adaptation in perch do not involve visual opsins. However, we found a nonsynonymous candidate SNP (CM020912_7210778; T669G;

TABLE 1 Enrichment and depletion of candidate SNPs in each annotation category

Variant	All SNPs ^a	Candidate SNPs ^a	χ^2	<i>p</i>	All SNPs (%)	Candidate SNPs (%)
5K upstream	132,087	1675	0.056	.813	12.22	12.14
5'UTR	8040	127	5.415	.020	0.74	0.92
5'UTR premature start codon gain	1386	13	0.979	.322	0.13	0.09
Synonymous	25,046	305	0.607	.436	2.32	2.21
Splice region & synonymous	580	8	0.001	.974	0.05	0.06
Nonsynonymous	18,801	237	0.024	.876	1.74	1.72
Nonsynonymous & splice region	445	3	0.826	.363	0.04	0.02
Intron	358,885	4668	1.207	.272	33.20	33.83
Splice region & intron	3564	42	0.192	.661	0.33	0.30
Splice region	154	2	0.000	1.000	0.01	0.01
3'UTR	33,552	492	8.883	.003	3.10	3.57
5K downstream	130,029	1775	6.937	.008	12.03	12.86
Intergenic	368,105	4450	9.759	.002	34.05	32.25
Stop gained	145	1	0.064	.801	0.01	0.01

^aThe sum of SNPs per annotation category does not correspond to the total number of SNPs due to multiple annotations of some SNPs located in nearby (<5 K) or overlapped genes (see details in File S1).

Putative signals of adaptation were detected in genes and gene families with diverse functions. Most frequently, the candidate genes were involved in the regulation of organism development, nervous system development and calcium/potassium/sodium exchange, suggesting their possible role during early development and in maintenance of ion balance. In contrast, we did not observe strong evidence of selection on visual opsins, despite the extreme differences in visual environment between the studied lakes. The observed overrepresentation of candidate SNPs in regulatory regions (5'UTR, 3'UTR and 5K downstream gene regions) suggests that regulatory changes may play a more important role in humic adaptation compared to changes in amino acids.

4.1 | Role of regulatory regions in humic adaptation

Most candidate SNPs were detected in intergenic and intronic regions, which comprise a large part of the perch genome (Ozerov et al., 2018). At the same time, we observed a significant excess of candidate SNPs in regulatory regions (5'UTR, 3'UTR and 5K downstream gene regions), indicating their important role in humic adaptation of perch. On the other hand, the proportion of nonsynonymous mutations among candidates did not show an increase compared to the whole data set. Thus, our findings are in line with a growing body of evidence suggesting that natural selection is predominantly acting at regulatory regions (e.g., Fagny & Austerlitz, 2021; Fraser, 2013; Glaser-Schmitt & Parsch, 2018; Verta & Jones, 2019). For example, Fraser (2013) showed that local adaptations found in a subset of human populations are over 10-fold more likely to affect gene expression than to alter protein sequences. The excess of candidate SNPs in 5' and 3' UTRs observed in perch is in agreement

with this, as UTRs are among the main elements involved in the regulation of gene expression (Barrett et al., 2012). 5'UTRs, located upstream of the protein coding sequence, play an important role in translation initiation, and expression of alternative 5'UTRs allows variation in expression from a single gene and tissue-specific expression patterns (Barrett et al., 2012). 3'UTRs, located downstream of the protein coding region, impact post-transcriptional and translational processes, including mRNA localization (Andreassi & Riccio, 2009), stability (Goldstrohm & Wickens, 2008) and expression levels (Matoukova et al., 2012). In general, 3'UTRs are more polymorphic and variable in length than 5'UTRs, resulting in a greater evolutionary potential of these regulatory elements (Barrett et al., 2012; Steri et al., 2018). Indeed, the observed number of SNPs in 3'UTRs in the perch genome was nearly four times higher than that in 5'UTRs (Table 1). The important role of 3'UTRs in teleost evolution has been recently highlighted in cichlid fishes, suggesting that these regions might act as metaregulators (i.e., regulators of other mechanisms governing post-transcriptional regulation), particularly in species undergoing rapid adaptation and speciation (Xiong et al., 2018). Therefore, our results suggest that the changes in regulatory regions may play a more important role in shaping humic adaptation in perch, compared to mutations in protein sequences.

4.2 | Candidate genes and gene families involved in humic adaptation

We detected multiple signals of selection consisting of over 3,000 genes scattered across the perch genome. These candidate genes are involved in multiple biological processes and cellular functions, most significant of which are multicellular organism

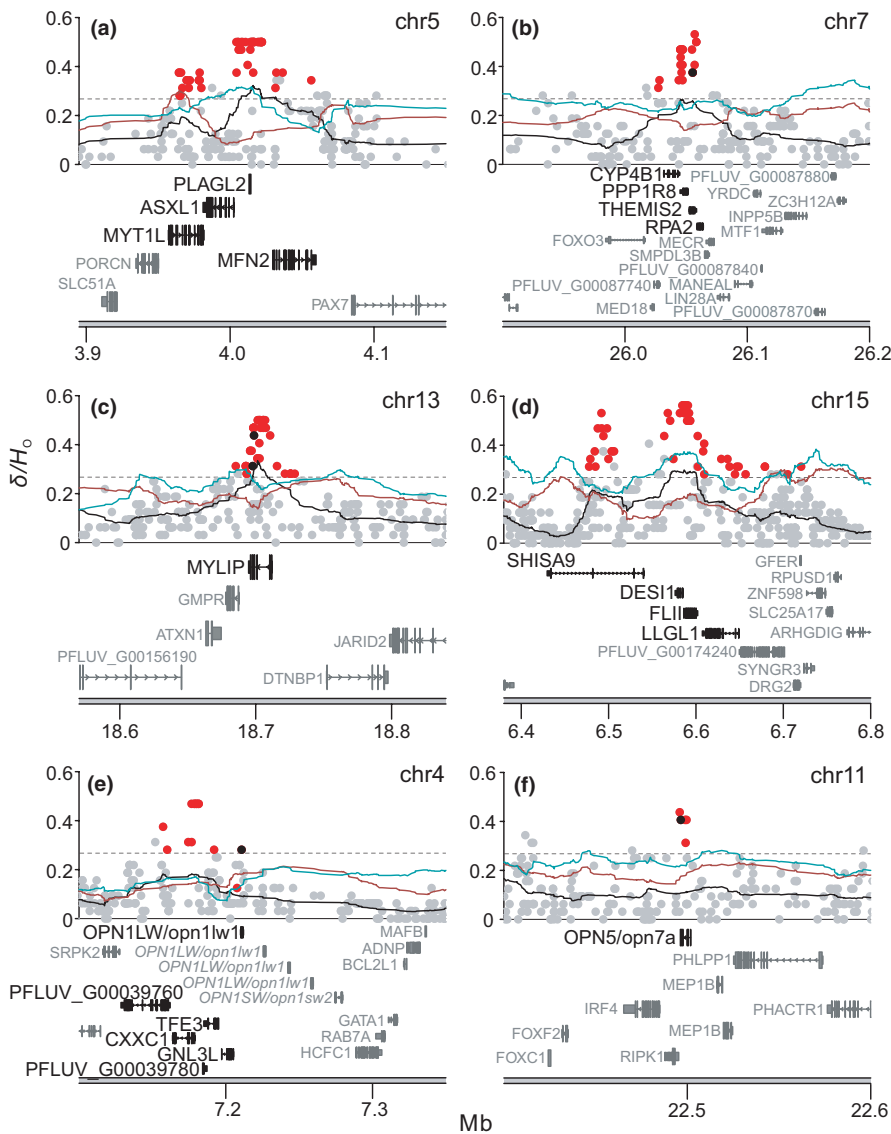


FIGURE 4 Examples of (a–d) genomic regions of high genetic divergence (δ) between humic and clear-water perch, and (e, f) opsins with nonsynonymous mutations. Candidate SNPs, nonsynonymous candidate SNPs and other SNPs are shown by red, black and grey dots, respectively. Moving averages of δ , H_0 of humic perch and H_0 of clear-water perch across 50 SNPs are shown with black, brown and cyan solid lines, respectively. Gene symbols are presented as human (and zebrafish for opsins) orthologues and perch GenBank gene IDs (PFLUV_G) for genes with unidentified functions. Candidate genes in the regions of elevated genetic divergence and opsin genes with nonsynonymous substitutions are highlighted in black, while other opsin genes at chromosome 4 are highlighted in italic. The dashed line indicates δ threshold = 0.268, which corresponds to 2.5 SD of mean δ

processes and system development, actin and calmodulin binding, and ion transfer. Several genes that showed elevated genetic divergence between humic and clear-water perch were also involved in development processes. For example, the signal of divergent selection involving more than 30 SNPs was centred around the *MYLIP* gene, essential for embryonic development (Knowlton et al., 2003). Moreover, the role of *MYLIP* in early embryonic development of zebrafish involves calcium-dependent mechanisms during gastrulation (Knowlton et al., 2003). Therefore, it is possible that the observed adaptive variation in perch may be linked to Ca^{2+} deficiency compensation during embryonic development in humic lakes. However, drawing firm conclusions about the role of *MYLIP* in perch embryogenesis and humic adaptation requires further investigation (e.g., Kurko et al., 2020).

Similar to *MYLIP*, many other genes involved in embryonic development were found in the regions of elevated genetic divergence. For example, a region on chromosome 5 includes genes shown to regulate brain and eyes (*PLAGL2*; Pendeville et al., 2006), hypothalamus (*MYT1L*; Blanchet et al., 2017), axonal and neuromuscular

(*MFN2*; Vettori et al., 2011), and neutrophil development (*ASXL1*; Fang et al., 2021) in zebrafish. The *LLGL1* gene located on chromosome 15 was found to regulate zebrafish cardiac development (Flinn et al., 2020). Other regions located on chromosomes 7 and 15 included genes involved in the immune response, such as the T-cell receptor signalling pathway (*THEMIS2*; Cheng et al., 2017; Peirce et al., 2010), inflammation and wound repair (*FLII*; Strudwick & Cowin, 2020), as well as genes involved in DNA replication and reparation (*RPA2*, *DES11*; Mer et al., 2000; Shin et al., 2012). However, more elaborate molecular and ecological studies are needed to shed light on specific phenotypic consequences and fitness effects of these and other candidate genes.

Among the candidate genes, we identified a large group of genes encoding transmembrane transportation of ions, such as solute carriers (SLCs; Hediger et al., 2004), calcium channel subunits (CACNs; Catterall et al., 2005), potassium channel subunits (KCNs; Shieh et al., 2000) and sodium channel subunits (SCNs; Yu & Catterall, 2003). There were also genes involved in energy-dependent transportation, such as ATP-binding cassette transporters (ABCs; Jones

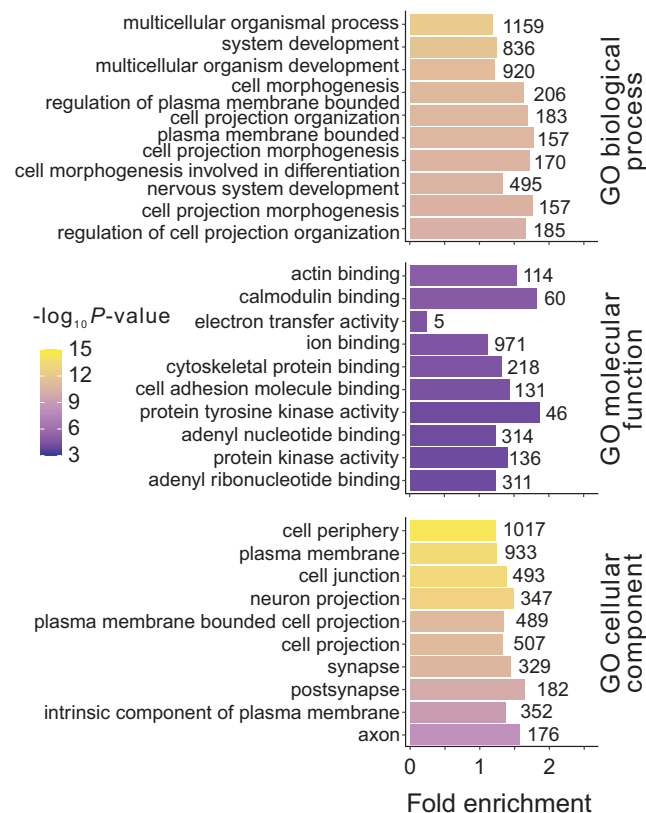


FIGURE 5 Top 10 significantly enriched gene ontology (GO) terms (FDR \leq 0.05) among the 3245 candidate genes for humic adaptation in perch. Bar length and numbers on the right represent the fold enrichment and number of enriched genes for each GO term, respectively

& George, 2004). Given the low pH and low level of calcium ions in humic lakes, we expected a major role of Na^+/H^+ , K^+ and Ca^{2+} exchangers during osmoregulation and maintenance of internal homeostasis in humic perch (Hwang et al., 2011). Accordingly, among the candidate genes were several SLC family 9 genes (*SLC9A1*, *SLC9A3R1*, *SLC9A5*, *SLC9A6* and *SLC9B2*) and one SLC family 4 gene (*SLC4A5*), which play an important role in the regulation of cellular pH via the transport of bicarbonate ions and protons (Pizzagalli et al., 2021). Moreover, selection signals were also observed in SLC family 8 (*SLC8A1* and *SLC8A2*) and family 24 (*SLC24A2*), and in calcium channel subunits alpha (*CACNA1A*, *CACNA1B*, *CACNA1C*, *CACNA1D*, *CACNA1E*, *CACNA1G*, *CACNA1I*, *CACNA2D1* and *CACNA2D4*), which play a significant role in the regulation of intracellular Ca^{2+} concentrations and Ca^{2+} influx (Pallone et al., 2012; Pizzagalli et al., 2021). Similarly, selection signals were found in several genes linked to pH maintenance via ion and acid–base equivalent exchanges, such as SCNs (*SCN1B*, *SCN3B*) and KCNs (*KCNA7*, *KCNAB1*, *KCNC3*, *KCNC4*, *KCNE1*, *KCNF1*, *KCNH2*, *KCNIP4*, *KCNJ12*, *KCNK13*, *KCNMA1*, *KCNN2*, *KCNQ1*, *KCNQ2*, *KCNQ3*, *KCNQ5*). Therefore, our results suggest that adaptation to high DOC concentrations probably includes a large group of genes involved in ion transport and balance. The above-mentioned gene families have also been associated with adaptation to acidic (Haenel et al., 2019) and alkaline (Xu et al., 2017)

environments in three-spined stickleback (*G. aculeatus*) and Amur ide (*Leuciscus waleckii*), respectively. Together, our results corroborate findings that variation in water chemistry, including DOC content, ion concentration and pH (Parra & Baldisserotto, 2007; Rask, 1984; Wood et al., 2011), is a strong selective force shaping molecular mechanisms of adaptation in teleosts.

4.3 | Opsins

Although we found several nonsynonymous substitutions in green- and red-sensitive opsins and in rhodopsin, the allele frequency differences between humic and clear-water habitats were small and could be explained by random drift alone. One potential exception was the single nonsynonymous SNP in one of the four red-sensitive opsin-like gene orthologues (*OPN1LW*) that showed a moderate allele frequency difference ($\delta = 0.28$) between humic and clear-water perch. However, as this putative selective sweep region consisted of several genes (e.g., *TFE3*, *CXXC1*) and multiple SNPs with a high allele frequency difference, it is not trivial to pinpoint the specific variant under divergent selection. Furthermore, a high allele frequency difference ($\delta = 0.45$) at *OPN1LW* between humic and clear-water perch was observed only among South-Western Baltic samples, whereas all 10 studied Finnish lakes were homozygous for one allele irrespective of visual environment. Further comparison of *P. fluviatilis* *OPN1LW* protein sequence with other teleosts revealed that Met and Ile at position 223 probably correspond to the ancestral and the derived allele, respectively, and the frequency of the derived allele was higher in clear-water lakes. Therefore, we cannot exclude the possibility that the allele frequency difference at nonsynonymous SNPs in *OPN1LW* is shaped by divergent selection occurring in clear-water, rather than in the humic environment.

We also detected a nonsynonymous substitution in the nonvisual opsin 7a (*OPN5/opn7a*) that had a large difference in allele frequency between humic and clear-water perch ($\delta = 0.41$; 99.94 percentile of all nonsynonymous SNPs). The opsin 7 family shows responsiveness at wavelengths <380 nm, which might be expected for a tissue that is exposed to direct sunlight. These genes are expressed in various tissues including the brain, digestive system, eye, heart and testis (Davies et al., 2015; Liu et al., 2020). However, as the SNP diversity adjacent to *OPN5/opn7a* was rather similar in both habitats, the role of this nonsynonymous variation in the context of adaptation to contrasting visual environments needs further investigation.

To conclude, our results suggest that amino acid changes in perch visual opsins most likely do not play a main role in adaptation to extreme visual environments. As such, these findings contrast recent reports in three-spine stickleback (Marques et al., 2017) and Baltic herring (Hill et al., 2019), where a single amino acid change was shown to have a major effect on ecological adaptation. Nevertheless, recent work based on RNA-sequencing analysis of the whole eye in Eurasian perch has shown differential expression of visual red-, green- and short-wavelength-sensitive opsins (*OPN1LW/opn1lw1*, *OPN1MW/opn1mw1* and

OPN1SW/opn1sw2, respectively) between humic and clear-water environments (Noreikiene et al., 2020), indicating that regulation of visual opsin expression still occurs in humic lakes and may represent an important mechanism of adaptation.

4.4 | Potential implications of genetic and spatial structuring on outlier detection

Genetic and spatial structuring has important consequences on the performance of outlier tests, since the true signals of selection may be harder to detect among highly differentiated populations (Hoban et al., 2016) and the effects of drift may overcome the effects of selection if selection is weak and/or the population is small (Hedrick, 2005). We aimed to mitigate the influence of random genetic drift and population stratification in our work by the following measures. First, the paired sampling strategy applied in our study (i.e., focusing on geographically proximate lakes with drastic environmental differences in humic content) allowed us to minimize the potential effects of spatial genetic structuring and maximize the number of independent population replicates. Second, the applied strategy potentially increases the power to detect common loci under weak to moderate selection compared to random sampling designs (Hoban et al., 2016; Lotterhos & Whitlock, 2015). On the other hand, analysis of a limited number of populations combined with an unequal number of samples per geographical region precluded us from carrying out detailed investigations of the population- or region-specific signatures of adaptation. For example, splitting the samples into two genetically distinct groups corresponding to the South-Western Baltic and the Finnish lakes likely reduces power and increases the false discovery rate due to a decrease of sample size, which, in turn, makes the analyses more sensitive to high between-population variance (De Mita et al., 2013). Instead, we concentrated on identification of common signatures of selection from standing genetic variation after perch colonized the studied area ~12,000–11,000 years ago (Nesbø et al., 1999). Furthermore, candidate SNPs potentially influenced by humic adaptation were called only when multiple methods supported the outlier identification, and two approaches applied in the current study accounted for global genetic (LFMM and RDA) and spatial structuring (RDA). Integration of population genetic structure, such as random factors, latent factors or covariates in the models, can improve the accuracy of both differentiation and genetic–environment association outlier tests (Frichot et al., 2013; Lotterhos & Whitlock, 2015). The lack of substantial influence of population stratification on our set of candidate SNPs was further supported by the high number of common outlier SNPs detected using both RDA (which takes genetic and geographical structuring into account) and simple allele frequency difference approaches, showing 7111 overlapping SNPs (Figure 2a). Therefore, it is unlikely that the set of candidate SNPs is driven strongly by large-scale genetic or spatial structuring. However, because most of the studied lakes originated from a geographically small region, the

relative weight of these samples (i.e., patterns observed among Estonian samples) were probably also detected as global signals, while population-specific or less common signals of selection remained undetected using our sampling design.

4.5 | Conclusions

This study revealed that humic adaptation in perch comprises a large number of regions and genes scattered across the genome. The excess of putatively adaptive variants that are found in 5'UTR, 3'UTR and 5K downstream gene regions highlights the importance of adaptive evolution via regulatory elements, rather than via amino acid changes in proteins. Putative adaptation signals were detected in genes and gene families with diverse functions, including genes involved in organism development, plasma membrane and ion transportation, underlying the multifaceted nature of humic-driven selection. Our work illustrates the power of whole genome analysis to identify the most promising candidates involved in adaptation to complex environmental conditions and provides the foundation to further investigate functions of candidate genes underlying phenotypic traits of ecological and evolutionary importance.

ACKNOWLEDGEMENTS

This study was funded by the Swedish Research Council grant 2020-03916 (to A.V.), Estonian Research Council grants PRG852 (to R.G., A.V. and K.N.), PRG1167 (to T.K.) and PUTJD954 (to M.S.), European Regional Development Fund and the programme Mobilias Plus (MOBJD344 to K.N.), the Ella and Georg Ehrnrooth foundation (to M.O.), and Swedish University of Agricultural Sciences (start-up grant to A.V.). We thank Ludvig Orsén for help with collecting samples in Sweden, Linas Ložys for providing samples from Lithuania, fishermen Kalle Bruus (Saadjärve) and August Toots (Kuulma) for help and assistance with perch sampling in Estonia, Eino Huotari and the late Arto Juntunen for participating in the field sampling trip in Eastern Finnish lakes, Centre of Evolutionary Applications (University of Turku, Finland) for help with DNA extraction and preparation for sequencing and Freed Ahmad (University of Turku, Finland) for help during the first steps of WGS analyses. We acknowledge CSC-IT Center for Science, Finland and Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), Sweden, for providing generous computational resources.

AUTHOR CONTRIBUTIONS

M.O. analysed the data and wrote the first draft of the manuscript together with A.V. M.L. analysed the data. T.K. and M.S. performed laboratory analyses of water samples. S.K., A.H., M.H., A.G., K.N. and A.V. were involved in fieldwork. A.V. and R.G. conceived the study. All the authors contributed to revisions of the draft manuscript. All the authors read and approved the final manuscript.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Raw sequence reads were deposited in the SRA (BioProject PRJNA760273; Ozerov et al. 2021a). Individual genotype data, supplementary tables and figures are available on DataDryad (<https://doi.org/10.5061/dryad.m0cfxpp4t>; Ozerov et al. 2021b). All other data are available within the Supporting Information accompanying this paper.

ORCID

Mikhail Ozerov  <https://orcid.org/0000-0002-1817-7707>

REFERENCES

- Altieri, A. H., & Gedan, K. B. (2015). Climate change and dead zones. *Global Change Biology*, 21(4), 1395–1406. <https://doi.org/10.1111/gcb.12754>
- Andreassi, C., & Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. *Trends in Cell Biology*, 19(9), 465–474. <https://doi.org/10.1016/j.tcb.2009.06.001>
- Andrews, S. (2017). *FastQC: A quality control tool for high throughput sequence data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Arora, K., Shah, M., Johnson, M., Sanghvi, R., Shelton, J., Nagulapalli, K., Oschwald, D. M., Zody, M. C., Germer, S., Jobanputra, V., Carter, J., & Robine, N. (2019). Deep whole-genome sequencing of 3 cancer cell lines on 2 sequencing platforms. *Scientific Reports*, 9(1), 19123. <https://doi.org/10.1038/s41598-019-55636-3>
- Arvola, L., Rask, M., Ruuhijärvi, J., Tulonen, T., Vuorenmäa, J., Ruoho-Airola, T., & Tulonen, J. (2010). Long-term patterns in pH and colour in small acidic boreal lakes of varying hydrological and landscape settings. *Biogeochemistry*, 101(1), 269–279. <https://doi.org/10.1007/s10533-010-9473-y>
- Ask, J., Karlsson, J., Persson, L., Ask, P., Byström, P., & Jansson, M. (2009). Terrestrial organic matter and light penetration: Effects on bacterial and primary production in lakes. *Limnology and Oceanography*, 54(6), 2034–2040. <https://doi.org/10.4319/lo.2009.54.6.2034>
- Barrett, L. W., Fletcher, S., & Wilton, S. D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences: CMLS*, 69(21), 3613–3634. <https://doi.org/10.1007/s00018-012-0990-9>
- Bastviken, D., Cole, J., Pace, M., & Tranvik, L. (2004). Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. *Global Biogeochemical Cycles*, 18(4), <https://doi.org/10.1029/2004GB002238>
- Battin, T. J., Luyssaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., & Tranvik, L. J. (2009). The boundless carbon cycle. *Nature Geoscience*, 2(9), 598–600. <https://doi.org/10.1038/ngeo618>
- Blanchet, P., Bebin, M., Bruet, S., Cooper, G. M., Thompson, M. L., Duban-Bedu, B., Gerard, B., Piton, A., Suckno, S., Deshpande, C., Clowes, V., Vogt, J., Turnpenny, P., Williamson, M. P., Alembik, Y., Glasgow, E., & McNeill, A. (2017). MYT1L mutations cause intellectual disability and variable obesity by dysregulating gene expression and development of the neuroendocrine hypothalamus. *PLoS Genetics*, 13(8), e1006957. <https://doi.org/10.1371/journal.pgen.1006957>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bornhauser, B. C., Olsson, P.-A., & Lindholm, D. (2003). MSAP Is a Novel MIR-interacting protein that enhances neurite outgrowth and increases myosin regulatory light chain*. *Journal of Biological Chemistry*, 278(37), 35412–35420. <https://doi.org/10.1074/jbc.M306271200>
- Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., Garçon, V., Gilbert, D., Gutiérrez, D., Isensee, K., Jacinto, G. S., Limburg, K. E., Montes, I., Naqvi, S. W. A., Pitcher, G. C., Rabalais, N. N., Roman, M. R., Rose, K. A., Seibel, B. A., ... Zhang, J. (2018). Declining oxygen in the global ocean and coastal waters. *Science*, 359(6371), eaam7240. <https://doi.org/10.1126/science.aam7240>
- Bricaud, A., Morel, A., & Prieur, L. (1981). Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnology and Oceanography*, 26(1), 43–53. <https://doi.org/10.4319/lo.1981.26.1.0043>
- Calkin, A. C., Goult, B. T., Zhang, L., Fairall, L., Hong, C., Schwabe, J. W., & Tontoz, P. (2011). FERM-dependent E3 ligase recognition is a conserved mechanism for targeted degradation of lipoprotein receptors. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), 20107–20112. <https://doi.org/10.1073/pnas.1111589108>
- Carleton, K. L., Escobar-Camacho, D., Stieb, S. M., Cortesi, F., & Marshall, N. J. (2020). Seeing the rainbow: mechanisms underlying spectral sensitivity in teleost fishes. *Journal of Experimental Biology*, 223(8), <https://doi.org/10.1242/jeb.193334>
- Catterall, W. A., Perez-Reyes, E., Snutch, T. P., & Striessnig, J. (2005). International union of pharmacology. XLVIII. nomenclature and structure-function relationships of voltage-gated calcium channels. *Pharmacological Reviews*, 57(4), 411–425. <https://doi.org/10.1124/pr.57.4.5>
- Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution*, 36(4), 852–860. <https://doi.org/10.1093/molbev/msz008>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Cheng, D., Deobagkar-Lele, M., Zvezdova, E., Choi, S., Uehara, S., Baup, D., Bennett, S. C., Bull, K. R., Crockford, T. L., Ferry, H., Warzecha, C., Marcellin, M., de Peredo, A. G., Lesourne, R., Anzilotti, C., Love, P. E., & Cornall, R. J. (2017). Themis2 lowers the threshold for B cell activation during positive selection. *Nature Immunology*, 18(2), 205–213. <https://doi.org/10.1038/ni.3642>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Cole, J. J. (2009). Production in pristine lakes. *Nature*, 460(7254), 463–464. <https://doi.org/10.1038/460463a>
- Collette, B. B., & Bănărescu, P. (1977). Systematics and zoogeography of the fishes of the family percidae. *Journal of the Fisheries Research Board of Canada*, 34(10), 1450–1463. <https://doi.org/10.1139/f77-209>
- Creed, I. F., Bergström, A.-K., Trick, C. G., Grimm, N. B., Hessen, D. O., Karlsson, J., Kidd, K. A., Kritzberg, E., McKnight, D. M., Freeman, E. C., Senar, O. E., Andersson, A., Ask, J., Berggren, M., Cherif, M., Giesler, R., Hotchkiss, E. R., Kortelainen, P., Palta, M. M., ... Weyhenmeyer, G. A. (2018). Global change-driven effects on dissolved organic matter composition: Implications for food webs of northern lakes. *Global Change Biology*, 24(8), 3692–3714. <https://doi.org/10.1111/gcb.14129>
- Dalongeville, A., Benestan, L., Mouillot, D., Lobreaux, S., & Manel, S. (2018). Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC Genomics*, 19(1), 217. <https://doi.org/10.1186/s12864-018-4579-z>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean,

- G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Davies, W. I. L., Tamaï, T. K., Zheng, L., Fu, J. K., Rihel, J., Foster, R. G., Whitmore, D., & Hankins, M. W. (2015). An extended family of novel vertebrate photopigments is widely expressed and displays a diversity of function. *Genome Research*, 25(11), 1666–1679. <https://doi.org/10.1101/gr.189886.115>
- De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, 22(5), 1383–1399. <https://doi.org/10.1111/mec.12182>
- De-Kayne, R., Frei, D., Greenway, R., Mendes, S. L., Retel, C., & Feulner, P. G. D. (2021). Sequencing platform shifts provide opportunities but pose challenges for combining genomic data sets. *Molecular Ecology Resources*, 21(3), 653–660. <https://doi.org/10.1111/1755-0998.13309>
- Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, 55(4), 997–1004. <https://doi.org/10.1111/j.0006-341x.1999.00997.x>
- Devlin, S. P., Saarenheimo, J., Syväranta, J., & Jones, R. I. (2015). Top consumer abundance influences lake methane efflux. *Nature Communications*, 6, 8787. <https://doi.org/10.1038/ncomms9787>
- Diehl, S. (1992). Fish predation and benthic community structure: The role of omnivory and habitat complexity. *Ecology*, 73(5), 1646–1661. <https://doi.org/10.2307/1940017>
- Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 1–20. <https://doi.org/10.18637/jss.v022.i04>
- Dymowska, A. K., Hwang, P.-P., & Goss, G. G. (2012). Structure and function of ionocytes in the freshwater fish gill. *Respiratory Physiology & Neurobiology*, 184(3), 282–292. <https://doi.org/10.1016/j.resp.2012.08.025>
- Eloranta, P. (1978). Light penetration in different types of lakes in Central Finland. *Ecography*, 1(4), 362–366. <https://doi.org/10.1111/j.1600-0587.1978.tb00971.x>
- Erlandsson, M., Cory, N., Köhler, S., & Bishop, K. (2010). Direct and indirect effects of increasing dissolved organic carbon levels on pH in lakes recovering from acidification. *Journal of Geophysical Research: Biogeosciences*, 115(G03004), 1–8. <https://doi.org/10.1029/2009JG001082>
- Evans, C. D., Monteith, D. T., & Cooper, D. M. (2005). Long-term increases in surface water dissolved organic carbon: Observations, possible causes and environmental impacts. *Environmental Pollution*, 137(1), 55–71. <https://doi.org/10.1016/j.envpol.2004.12.031>
- Fagny, M., & Austerlitz, F. (2021). Polygenic adaptation: Integrating population genetics and gene regulatory networks. *Trends in Genetics*, 37(7), 631–638. <https://doi.org/10.1016/j.tig.2021.03.005>
- Fang, X., Xu, S., Zhang, Y., Xu, J., Huang, Z., Liu, W., Wang, S., Yen, K., & Zhang, W. (2021). Asx1 C-terminal mutation perturbs neutrophil differentiation in zebrafish. *Leukemia*, 35, 2299–2310. <https://doi.org/10.1038/s41375-021-01121-8>
- Flinn, M. A., Otten, C., Brandt, Z. J., Bostrom, J. R., Kenarsary, A., Wan, T. C., Auchampach, J. A., Abdelilah-Seyfried, S., O'Meara, C. C., & Link, B. A. (2020). Lgl1 regulates zebrafish cardiac development by mediating Yap stability in cardiomyocytes. *Development*, 147(16), dev193581 <https://doi.org/10.1242/dev.193581>
- Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, 25(1), 104–120. <https://doi.org/10.1111/mec.13476>
- Forsberg, C. (1992). Will an increased greenhouse impact in Fennoscandia give rise to more humic and coloured lakes? *Hydrobiologia*, 229(1), 51–58. <https://doi.org/10.1007/BF00006990>
- Fraser, H. B. (2013). Gene expression drives local adaptation in humans. *Genome Research*, 23(7), 1089–1096. <https://doi.org/10.1101/gr.152710.112>
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. <https://doi.org/10.1111/2041-210X.12382>
- Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, 30(7), 1687–1699. <https://doi.org/10.1093/molbev/mst063>
- García-Elfring, A., Paccard, A., Thurman, T. J., Wasserman, B. A., Palkovacs, E. P., Hendry, A. P., & Barrett, R. D. H. (2021). Using seasonal genomic changes to understand historical adaptation to new environments: Parallel selection on stickleback in highly-variable estuaries. *Molecular Ecology*, 30(9), 2054–2064. <https://doi.org/10.1111/mec.15879>
- Gel, B., & Serra, E. (2017). karyoplotR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*, 33(19), 3088–3090. <https://doi.org/10.1093/bioinformatics/btx346>
- Glaser-Schmitt, A., & Parsch, J. (2018). Functional characterization of adaptive variation within a cis-regulatory element influencing *Drosophila melanogaster* growth. *PLoS Biology*, 16(1), e2004538. <https://doi.org/10.1371/journal.pbio.2004538>
- Goldstrohm, A. C., & Wickens, M. (2008). Multifunctional deadenylase complexes diversify mRNA control. *Nature Reviews Molecular Cell Biology*, 9(4), 337–344. <https://doi.org/10.1038/nrm2370>
- Greenaway, P. (1985). Calcium balance and molting in the Crustacea. *Biological Reviews of the Cambridge Philosophical Society*, 60(3), 425–454. <https://doi.org/10.1111/j.1469-185X.1985.tb00424.x>
- Grether, G. F., Millie, D. F., Bryant, M. J., Reznick, D. N., & Mayea, W. (2001). Rain forest canopy cover, resource availability, and life history evolution in guppies. *Ecology*, 82, 546–1559. <https://doi.org/10.2307/2679799>
- Grummer, J. A., Beheregaray, L. B., Bernatchez, L., Hand, B. K., Luikart, G., Narum, S. R., & Taylor, E. B. (2019). Aquatic landscape genomics and environmental effects on genetic variation. *Trends in Ecology & Evolution*, 34(7), 641–654. <https://doi.org/10.1016/j.tree.2019.02.013>
- Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize implements and enhances circular visualization in R. *Bioinformatics*, 30(19), 2811–2812. <https://doi.org/10.1093/bioinformatics/btu393>
- Haenel, Q., Roesti, M., Moser, D., MacColl, A. D. C., & Berner, D. (2019). Predictable genome-wide sorting of standing genetic variation during parallel adaptation to basic versus acidic environments in stickleback fish. *Evolution Letters*, 3(1), 28–42. <https://doi.org/10.1002/evl3.99>
- Hannan, K. D., & Rummer, J. L. (2018). Aquatic acidification: a mechanism underpinning maintained oxygen transport and performance in fish experiencing elevated carbon dioxide conditions. *Journal of Experimental Biology*, 221(5), jeb154559 <https://doi.org/10.1242/jeb.154559>
- Hediger, M. A., Romero, M. F., Peng, J.-B., Rolfs, A., Takanaga, H., & Bruford, E. A. (2004). The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Archiv*, 447(5), 465–468. <https://doi.org/10.1007/s00424-003-1192-y>
- Hedrick, P. W. (2005). *Genetics of Populations*, 3rd ed. (675 pp). Sinauer & Associates.
- Hill, J., Enbody, E. D., Pettersson, M. E., Sprehn, C. G., Bekkevold, D., Folkvord, A., Laikre, L., Kleinau, G., Scheerer, P., & Andersson, L. (2019). Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proceedings of the National Academy of Sciences*, 116(37), 18473–18478. <https://doi.org/10.1073/pnas.1908332116>
- Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolín, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M.

- C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist*, 188(4), 379–397. <https://doi.org/10.1086/688018>
- Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and west-slope cutthroat trout. *Molecular Ecology Resources*, 11, 117–122. <https://doi.org/10.1111/j.1755-0998.2010.02967.x>
- Honma, S., Kawamoto, T., Takagi, Y., Fujimoto, K., Sato, F., Noshiro, M., Kato, Y., & Honma, K.-I. (2002). Dec1 and Dec2 are regulators of the mammalian molecular clock. *Nature*, 419(6909), 841–844. <https://doi.org/10.1038/nature01123>
- Hope, D., Kratz, T. K., & Riera, J. L. (1996). Relationship between and dissolved organic carbon in northern Wisconsin lakes. *Journal of Environmental Quality*, 25(6), 1442–1445. <https://doi.org/10.2134/jeq1996.00472425002500060039x>
- Hwang, P.-P., Lee, T.-H., & Lin, L.-Y. (2011). Ion regulation in fish gills: Recent progress in the cellular and molecular mechanisms. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 301(1), R28–R47. <https://doi.org/10.1152/ajpregu.00047.2011>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., ... Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55–61. <https://doi.org/10.1038/nature10944>
- Jones, P. M., & George, A. M. (2004). The ABC transporter structure and mechanism: Perspectives on recent research. *Cellular and Molecular Life Sciences: CMLS*, 61(6), 682–699. <https://doi.org/10.1007/s00018-003-3336-9>
- Jones, R. I. (1992). The influence of humic substances on lacustrine planktonic food-chains. *Hydrobiologia*, 229(1), 73–91. <https://doi.org/10.1007/BF0006992>
- Kaeuffer, R., Peichel, C. L., Bolnick, D. I., & Hendry, A. P. (2012). Parallel and nonparallel aspects of ecological, phenotypic and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution*, 66, 402–418. <https://doi.org/10.1111/j.1558-5646.2011.01440.x>
- Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6, 208. <https://doi.org/10.3389/fgene.2015.00208>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Kankaala, P., Huotari, J., Peltomaa, E., Saloranta, T., & Ojala, A. (2006). Methanotrophic activity in relation to methane efflux and total heterotrophic bacterial production in a stratified, humic, boreal lake. *Limnology and Oceanography*, 51(2), 1195–1204. <https://doi.org/10.4319/lo.2006.51.2.1195>
- Karlsson, J., Byström, P., Ask, J., Ask, P., Persson, L., & Jansson, M. (2009). Light limitation of nutrient-poor lake ecosystems. *Nature*, 460(7254), 506–509. <https://doi.org/10.1038/nature08179>
- Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. Retrieved from <https://cran.r-project.org/web/packages/factoextra/index.html>
- Keskitalo, J., & Eloranta, P. (1999). *Limnology of Humic Waters*. Backhuys, 284 pp.
- Knowlton, M. N., Chan, B. M. C., & Kelly, G. M. (2003). The zebrafish band 4.1 member Mir is involved in cell movements associated with gastrulation. *Developmental Biology*, 264(2), 407–429. <https://doi.org/10.1016/j.ydbio.2003.09.001>
- Kritzbeg, E. S., Hasselquist, E. M., Škerlep, M., Löfgren, S., Olsson, O., Stadmark, J., Valinia, S., Hansson, L.-A., & Laudon, H. (2020). Browning of freshwaters: Consequences to ecosystem services, underlying drivers, and potential mitigation measures. *Ambio*, 49(2), 375–390. <https://doi.org/10.1007/s13280-019-01227-5>
- Kurko, J., Debes, P. V., House, A. H., Aykanat, T., Erkinaro, J., & Primmer, C. R. (2020). Transcription profiles of age-at-maturity-associated genes suggest cell fate commitment regulation as a key factor in the Atlantic salmon maturation process. *G3: Genes|genomes|genetics*, 10(1), 235. <https://doi.org/10.1534/g3.119.400882>
- Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*, 5th ed. McGraw-Hill Irwin.
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Legendre, P., & Legendre, L. (2012). *Numerical Ecology*, 3rd ed. (990 pp). Elsevier.
- Lei, L., Thompson, J. A., & McDonald, L. M. (2020). Assessment of dissolved organic carbon and iron effects on water color between a forest and pasture-dominated fine-scale catchment in a Central Appalachian region, West Virginia. *Environmental Science and Pollution Research*, 27(23), 29464–29474. <https://doi.org/10.1007/s11356-020-09251-9>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, 30(20), 2843–2851. <https://doi.org/10.1093/bioinformatics/btu356>
- Li, X., Poschmann, S., Chen, Q., Fazeli, W., Oundjian, N. J., Snoeijs-Schouwenaars, F. M., Fricke, O., Kamsteeg, E.-J., Willemsen, M., & Wang, Q. K. (2018). De novo BK channel variant causes epilepsy by affecting voltage gating but not Ca(2+) sensitivity. *European Journal of Human Genetics*, 26(2), 220–229. <https://doi.org/10.1038/s41433-017-0073-3>
- Liu, Y., Zhang, W., Du, X., Liu, Y., Qu, J., Liu, X., & Zhang, Q. (2020). Genome-wide identification of nonvisual opsin family reveals amplification of RPE-retinal G protein receptor gene (RGR) and offers novel insights into functions of RGR(s) in Paralichthys olivaceus (Paralichthyidae, Teleostei). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 334(1), 25–36. <https://doi.org/10.1002/jez.b.22914>
- Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, 24(5), 1031–1046. <https://doi.org/10.1111/mec.13100>
- Maloney, K. O., Morris, D. P., Moses, C. O., & Osburn, C. L. (2005). The role of iron and dissolved organic carbon in the absorption of ultraviolet radiation in humic lake water. *Biogeochemistry*, 75(3), 393–407. <https://doi.org/10.1007/s10533-005-1675-3>
- Marques, D. A., Taylor, J. S., Jones, F. C., Di Palma, F., Kingsley, D. M., & Reimchen, T. E. (2017). Convergent evolution of SWS2 opsin facilitates adaptive radiation of threespine stickleback into different light environments. *PLoS Biology*, 15(4), e2001627. <https://doi.org/10.1371/journal.pbio.2001627>
- Matoulkova, E., Michalova, E., Vojtesek, B., & Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of

- protein expression in mammalian cells. *RNA Biology*, 9(5), 563–576. <https://doi.org/10.4161/rna.20231>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McKnight, D. M., & Aiken, G. R. (1998). Sources and Age of Aquatic Humus. In D. O. Hessen, & L. J. Tranvik (Eds.), *Aquatic Humic Substances: Ecology and Biogeochemistry* (pp. 9–39). Springer, Berlin Heidelberg.
- Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C. J., Edwards, A. M., & Chazin, W. J. (2000). Structural basis for the recognition of DNA repair proteins UNG2, XPA, and RAD52 by replication factor RPA. *Cell*, 103(3), 449–456. [https://doi.org/10.1016/S0092-8674\(00\)00136-7](https://doi.org/10.1016/S0092-8674(00)00136-7)
- Miller, J. (1991). Reaction time analysis with outlier exclusion: bias varies with sample size. *The Quarterly Journal of Experimental Psychology Section A*, 43(4), 907–912. <https://doi.org/10.1080/14640749108400962>
- Miller, S. E., Roesti, M., & Schluter, D. (2019). A single interacting species leads to widespread parallel evolution of the stickleback genome. *Current Biology*, 29(3), 530–537.e536. <https://doi.org/10.1016/j.cub.2018.12.044>
- Musilova, Z., Cortesi, F., Matschiner, M., Davies, W. I. L., Patel, J. S., Stieb, S. M., de Busserolles, F., Malmström, M., Tørresen, O. K., Brown, C. J., Mountford, J. K., Hanel, R., Stenkamp, D. L., Jakobsen, K. S., Carleton, K. L., Jentoft, S., Marshall, J., & Salzburger, W. (2019). Vision using multiple distinct rod opsins in deep-sea fishes. *Science*, 364(6440), 588. <https://doi.org/10.1126/science.aav4632>
- Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, 106(949), 283–292. <https://doi.org/10.1086/282771>
- Nelson, J. S. (2006). *Fishes of the world*. John Wiley.
- Nesbø, C. L., Fosshem, T., Vøllestad, L. A., & Jakobsen, K. S. (1999). Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Molecular Ecology*, 8(9), 1387–1404. <https://doi.org/10.1046/j.1365-294x.1999.00699.x>
- Ng, P. K. L., Tay, J. B., & Lim, K. K. P. (1994). Diversity and conservation of blackwater fishes in Peninsular Malaysia, particularly in the North Selangor peat swamp forest. *Hydrobiologia*, 285(1), 203–218. <https://doi.org/10.1007/BF00005667>
- Noreikiene, K., Ozerov, M., Ahmad, F., Kõiv, T., Kahar, S., Gross, R., Sepp, M., Pellizzone, A., Vesterinen, E. J., Kisand, V., & Vasemägi, A. (2020). Humic-acid-driven escape from eye parasites revealed by RNA-seq and target-specific metabarcoding. *Parasites & Vectors*, 13(1), 433. <https://doi.org/10.1186/s13071-020-04306-9>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Wagner, H. (2020). *vegan: Community Ecology Package. R package version 2.5-7*. CRAN R-project. Retrieved from <https://CRAN.R-project.org/package=vegan>
- O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, 27(16), 3193–3206. <https://doi.org/10.1111/mec.14792>
- Olsson, P. A., Korhonen, L., Mercer, E. A., & Lindholm, D. (1999). MIR is a novel ERM-like protein that interacts with myosin regulatory light chain and inhibits neurite outgrowth. *Journal of Biological Chemistry*, 274(51), 36288–36292. <https://doi.org/10.1074/jbc.274.51.36288>
- Ozerov, M. Y., Ahmad, F., Gross, R., Pukk, L., Kahar, S., Kisand, V., & Vasemägi, A. (2018). Highly continuous genome assembly of Eurasian perch (*Perca fluviatilis*) using linked-read sequencing. *G3: Genes|genomes|genetics*, 8(12), 3737–3743. <https://doi.org/10.1534/g3.118.200768>
- Ozerov, M. Y., Noreikiene, K., Kahar, S., Huss, M., Huusko, A., Kõiv, T., Sepp, M., López, M. - E., Gårdmark, A., Gross, R., & Vasemägi, A. (2021a). Genomics of humic adaptation in Eurasian perch (*Perca fluviatilis*); raw DNA-seq reads; NCBI; BioProject: PRJNA760273. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA760273>
- Ozerov, M. Y., Noreikiene, K., Kahar, S., Huss, M., Huusko, Kõiv, T., Gross, R. & Vasemagi, A. (2021b). Genomics of humic adaptation in Eurasian perch (*Perca fluviatilis*), SNP genotypes of 32 perch individuals. Dryad; <https://doi.org/10.5061/dryad.m0cfxpp4t>
- Pallone, T. L., Khurana, S., & Cao, C. (2012). Voltage-Gated Calcium Channels: Structure and Function (CACNA). In S. Choi (Ed.), *Encyclopedia of Signaling Molecules* (pp. 1984–1992). Springer, New York.
- Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Parra, J. E. G., & Baldissarotto, B. (2007). Effect of water pH and hardness on survival and growth of freshwater teleosts. In B. Baldissarotto (Ed.), *Fish Osmoregulation*. CRC Press.
- Peirce, M. J., Brook, M., Morrice, N., Snelgrove, R., Begum, S., Lanfrancotti, A., Notley, C., Hussell, T., Cope, A. P., & Wait, R. (2010). Themis2/ICB1 is a signaling scaffold that selectively regulates macrophage Toll-like receptor signaling and cytokine production. *PLoS One*, 5(7), e11465. <https://doi.org/10.1371/journal.pone.0011465>
- Pendeville, H., Peers, B., Kas, K., & Voz, M. L. (2006). Cloning and embryonic expression of zebrafish PLAG genes. *Gene Expression Patterns*, 6(3), 267–276. <https://doi.org/10.1016/j.modgep.2005.08.001>
- Pfenninger, M., Lerp, H., Tobler, M., Passow, C., Kelley, J. L., Funke, E., Greshake, B., Erkoc, U. K., Berberich, T., & Plath, M. (2014). Parallel evolution of cox genes in H2S-tolerant fish as key adaptation to a toxic environment. *Nature Communications*, 5(1), 3873. <https://doi.org/10.1038/ncomms4873>
- Pizzagalli, M. D., Bensimon, A., & Superti-Furga, G. (2021). A guide to plasma membrane solute carrier proteins. *The FEBS Journal*, 288(9), 2784–2835. <https://doi.org/10.1111/febs.15531>
- Posa, M. R. C., Wijedasa, L. S., & Corlett, R. T. (2011). Biodiversity and conservation of tropical peat swamp forests. *BioScience*, 61(1), 49–57. <https://doi.org/10.1525/bio.2011.61.1.10>
- Prevosti, A., Ocaña, J., & Alonso, G. (1975). Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. *Theoretical and Applied Genetics*, 45(6), 231–241. <https://doi.org/10.1007/BF00831894>
- Rantala, M. V., Luoto, T. P., & Nevalainen, L. (2015). Late Holocene changes in the humic state of a boreal lake and their associations with organic matter transport and climate dynamics. *Biogeochemistry*, 123(1), 63–82. <https://doi.org/10.1007/s10533-014-0053-4>
- Rask, M. (1984). The effect of low pH on perch, *Perca fluviatilis* L. II. The effect of acid stress on different development stages of perch. *Annales Zoologici Fennici*, 21(1), 9–13.
- Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., Karchner, S. I., Hahn, M. E., Nacci, D., Oleksiak, M. F., Crawford, D. L., & Whitehead, A. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, 354(6317), 1305–1308. <https://doi.org/10.1126/science.aah4993>
- Riesch, R., Plath, M., Schlupp, I., Tobler, M., & Langerhans, B. (2014). Colonisation of toxic environments drives predictable life-history evolution in livebearing fishes (Poeciliidae). *Ecology Letters*, 17(1), 65–71. <https://doi.org/10.1111/ele.12209>
- Riesch, R., Tobler, M., & Plath, M. (Eds.) (2015). *Extremophile Fishes. Ecology, Evolution, and Physiology of Teleosts in Extreme Environments*. Springer (326 pp.). <https://doi.org/10.1007/978-3-319-13362-1>
- Roulet, N., & Moore, T. R. (2006). Browning the waters. *Nature*, 444(7117), 283–284. <https://doi.org/10.1038/444283a>
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and*

- Evolution*, 4(4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sepp, M., Kõiv, T., Nõges, P., & Nõges, T. (2019). The role of catchment soils and land cover on dissolved organic matter (DOM) properties in temperate lakes. *Journal of Hydrology*, 570, 281–291. <https://doi.org/10.1016/j.jhydrol.2019.01.012>
- Shieh, C. C., Coghlan, M., Sullivan, J. P., & Gopalakrishnan, M. (2000). Potassium channels: Molecular defects, diseases, and therapeutic opportunities. *Pharmacological Reviews*, 52(4), 557–594.
- Shin, E. J., Shin, H. M., Nam, E., Kim, W. S., Kim, J.-H., Oh, B.-H., & Yun, Y. (2012). DeSUMOylating isopeptidase: A second class of SUMO protease. *EMBO Reports*, 13(4), 339–346. <https://doi.org/10.1038/embor.2012.3>
- Sobek, S., Algesten, G., Bergström, A.-K., Jansson, M., & Tranvik, L. J. (2003). The catchment and climate regulation of pCO₂ in boreal lakes. *Global Change Biology*, 9(4), 630–641. <https://doi.org/10.1046/j.1365-2486.2003.00619.x>
- Sobek, S., Söderbäck, B., Karlsson, S., Andersson, E., & Brunberg, A. K. (2006). A carbon budget of a small humic lake: An example of the importance of lakes for organic matter cycling in boreal catchments. *Ambio*, 35(8), 469–475. <https://doi.org/10.1579/0044-7447>
- Solomon, C. T., Jones, S. E., Weidel, B. C., Buffam, I., Fork, M. L., Karlsson, J., Larsen, S., Lennon, J. T., Read, J. S., Sadro, S., & Saros, J. E. (2015). Ecosystem consequences of changing inputs of terrestrial dissolved organic matter to lakes: current knowledge and future challenges. *Ecosystems*, 18(3), 376–389. <https://doi.org/10.1007/s10021-015-9848-y>
- Steri, M., Idda, M. L., Whalen, M. B., & Orrù, V. (2018). Genetic variants in mRNA untranslated regions. *Wiley Interdisciplinary Reviews. RNA*, 9(4), e1474. <https://doi.org/10.1002/wrna.1474>
- Strudwick, X. L., & Cowin, A. J. (2020). Multifunctional roles of the actin-binding protein flightless I in inflammation, cancer and wound healing. *Frontiers in Cell and Developmental Biology*, 8(1394), <https://doi.org/10.3389/fcell.2020.603508>
- Tagliaro, M. (2019). Acidification in aquatic systems. In B. Fath (Ed.), *Encyclopedia of Ecology*, 2nd ed. (pp. 6–13). Elsevier.
- Thomas, P. D., Kejarawal, A., Campbell, M. J., Mi, H., Diemer, K., Guo, N., & Doremieux, O. (2003). PANTHER: A browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research*, 31(1), 334–341. <https://doi.org/10.1093/nar/gkg115>
- Tobler, M., & Plath, M. (2011). Living in extreme environments. In J. P. Evans, A. Pilastro, & I. Schlupp (Eds.), *Ecology and Evolution of Poeciliid Fishes* (pp. 120–127). University of Chicago Press.
- Tong, C., & Li, M. (2020). Genomic signature of accelerated evolution in a saline-alkaline lake-dwelling Schizothoracine fish. *International Journal of Biological Macromolecules*, 149, 341–347. <https://doi.org/10.1016/j.ijbiomac.2020.01.207>
- Tranvik, L. J. (2009). Dystrophy. In T. Mehner (Eds), *Encyclopaedia of Inland Waters* (pp. 405–410). Elsevier. <https://doi.org/10.1016/B978-012370626-3.00202-7>
- Tranvik, L. J., Cole, J. J., & Prairie, Y. T. (2018). The study of carbon in inland waters—from isolated ecosystems to players in the global carbon cycle. *Limnology and Oceanography Letters*, 3(3), 41–48. <https://doi.org/10.1002/lol2.10068>
- van Dorst, R. M., Gårdmark, A., Svanbäck, R., Beier, U., Weyhenmeyer, G. A., & Huss, M. (2019). Warmer and browner waters decrease fish biomass production. *Global Change Biology*, 25(4), 1395–1408. <https://doi.org/10.1111/gcb.14551>
- Verta, J.-P., & Jones, F. C. (2019). Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *eLife*, 8, e43785. <https://doi.org/10.7554/eLife.43785>
- Vettori, A., Bergamin, G., Moro, E., Vazza, G., Polo, G., Tiso, N., Argenton, F., & Mostacciolo, M. L. (2011). Developmental defects and neuromuscular alterations due to mitofusin 2 gene (MFN2) silencing in zebrafish: a new model for Charcot-Marie-Tooth type 2A neuropathy. *Neuromuscular Disorders*, 21(1), 58–67. <https://doi.org/10.1016/j.nmd.2010.09.002>
- Vuorenmaa, J., Forsius, M., & Mannio, J. (2006). Increasing trends of total organic carbon concentrations in small forest lakes in Finland from 1987 to 2003. *Science of the Total Environment*, 365(1–3), 47–65. <https://doi.org/10.1016/j.scitotenv.2006.02.038>
- Wetzel, R. G. (2001). *Limnology. Lake and River Ecosystems*. 3rd edn. Academic Press, 1006 pp. <https://doi.org/10.1016/C2009-0-02112-6>
- Weyhenmeyer, G. A., Hartmann, J., Hessen, D. O., Kopáček, J., Hejzlar, J., Jacquet, S., Hamilton, S. K., Verburg, P., Leach, T. H., Schmid, M., Flaim, G., Nõges, T., Nõges, P., Wentzky, V. C., Rogora, M., Rusak, J. A., Kosten, S., Paterson, A. M., Teubner, K., ... Zechmeister, T. (2019). Widespread diminishing anthropogenic effects on calcium in freshwaters. *Scientific Reports*, 9(1), 10450. <https://doi.org/10.1038/s41598-019-46838-w>
- Weyhenmeyer, G. A., Prairie, Y. T., & Tranvik, L. J. (2014). Browning of boreal freshwaters coupled to carbon-iron interactions along the aquatic continuum. *PLoS One*, 9(2), e88104. <https://doi.org/10.1371/journal.pone.0088104>
- Winter, D. J. (2017). rentrez: An R package for the NCBI eUtils API. *The R Journal*, 9(2), 520–526. <https://doi.org/10.32614/RJ-2017-058>
- Willis, S. C., Hollenbeck, C. M., Puritz, J. B., Gold, J. R., & Portnoy, D. S. (2017). Haplotyping RAD loci: an efficient method to filter paralogs and account for physical linkage. *Molecular Ecology Resources*, 17(5), 955–965. <https://doi.org/10.1111/1755-0998.12647>
- Wood, C. M., Al-Reasi, H. A., & Smith, D. S. (2011). The two faces of DOC. *Aquatic Toxicology*, 105(3–4), 3–8. <https://doi.org/10.1016/j.aquatox.2011.03.007>
- Xiong, P., Hulse, C. D., Meyer, A., & Franchini, P. (2018). Evolutionary divergence of 3' UTRs in cichlid fishes. *BMC Genomics*, 19(1), 433. <https://doi.org/10.1186/s12864-018-4821-8>
- Xu, J., Li, J. T., Jiang, Y., Peng, W., Yao, Z., Chen, B., & Xu, P. (2017). Genomic basis of adaptive evolution: The survival of Amur Ide (*Leuciscus waleckii*) in an extremely alkaline environment. *Molecular Biology and Evolution*, 34(1), 145–159. <https://doi.org/10.1093/molbev/msw230>
- Xuereb, A., Kimber, C. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M. J. (2018). Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. *Molecular Ecology*, 27(24), 5035–5048. <https://doi.org/10.1111/mec.14942>
- Yu, F. H., & Catterall, W. A. (2003). Overview of the voltage-gated sodium channel family. *Genome Biology*, 4(3), 207. <https://doi.org/10.1186/gb-2003-4-3-207>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ozerov, M., Noreikiene, K., Kahar, S., Huss, M., Huusko, A., Kõiv, T., Sepp, M., López, M.-E., Gårdmark, A., Gross, R., & Vasemägi, A. (2022). Whole-genome sequencing illuminates multifaceted targets of selection to humic substances in Eurasian perch. *Molecular Ecology*, 31, 2367–2383. <https://doi.org/10.1111/mec.16409>