1 **One substance to rule them all and in the darkness bind them: whole-genome**

2 **sequencing illuminates multifaceted targets of humic adaptation in Eurasian**

3 **perch**

4

5 ***Running title***: *Genomics of humic adaptation in perch*

6

7 Mikhail Ozerov[1,2,3], Kristina Noreikiene[4], Siim Kahar[4], Magnus Huss[5], Ari Huusko[6], Toomas Kõiv[7],

8 Margot Sepp[7], María López[1], Anna Gårdmark[5], Riho Gross[4], Anti Vasemägi[1,4,*]

9

10 [1]Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of

11 Agricultural Sciences, 17893, Drottningholm, Sweden

12 [2]Department of Biology, University of Turku, 20014, Turku, Finland

13 [3]Biodiversity Unit, University of Turku, 20014, Turku, Finland

14 [4]Chair of Aquaculture, Institute of Veterinary Medicine and Animal Sciences, Estonian University of

15 Life Sciences, Kreutzwaldi 46, 51006, Tartu, Estonia

16 [5]Swedish University of Agricultural Sciences, Department of Aquatic Resources, 74242, Öregrund,

17 Sweden

18 [6]Natural resources Institute Finland (Luke), 88300 Paltamo, Finland

19 [7]Chair of Hydrobiology and Fishery, Institute of Agricultural and Environmental Sciences, Estonian

20 University of Life Sciences, Kreutzwaldi 5, 51006, Tartu, Estonia

21

22 ****Corresponding author***: Anti Vasemägi, Swedish University of Agricultural Sciences,

23 Department of Aquatic Resources, Institute of Freshwater Research, Stångholmsvägen 2,

24 Drottningholm 17893, Sweden. Tel: + 46 10 478 4277; e-mail: anti.vasemagi@slu.se

25 **ABSTRACT**

26   Extreme environments are inhospitable to the majority of species, but some organisms are able to

27 survive in such hostile conditions due to evolutionary adaptations. For example, modern bony fishes

28 have colonized various aquatic environments, including perpetually dark, hypoxic, hypersaline and toxic

29 habitats. Eurasian perch (*Perca fluviatilis*) is among the few fish species of northern latitudes that is able

30 to live in extremely acidic humic lakes. Such lakes represent almost "nocturnal" environments; they

31 contain high levels of dissolved organic matter, which in addition to creating a challenging visual

32 environment, also affects a large number of other habitat parameters and biotic interactions. To reveal

33 the genomic targets of humic-associated selection, we performed whole-genome sequencing of perch

34 originating from 16 humic and 16 clear-water lakes in northern Europe. We identified over 800,000

35 SNPs, of which >10,000 were identified as potential candidates under selection (associated with >3,000

36 genes) using multiple outlier approaches. Our findings suggest that adaptation to the humic environment

37 involves hundreds of regions scattered across the genome. Putative signals of adaptation were detected

38 in genes and gene families with diverse functions, including organism development and ion

39 transportation. The observed excess of variants under selection in regulatory regions highlights the

40 importance of adaptive evolution via regulatory elements, rather than via protein sequence modification.

41 Our study demonstrates the power of whole-genome analysis to illuminate multifaceted nature of humic

42 adaptation and highlights the next challenge moving from high-throughput outlier identification towards

43 functional validation of causal mutations underlying phenotypic traits of ecological and evolutionary

44 importance.

45

46 *Keywords*: fish, humic adaptation, SNP, candidate, DOC, water color

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 fueled a long-standing interest in how organisms cope with and adapt to extreme conditions (Riesch, Tobler, & Plath, 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, deep sea, high altitude, hypoxic, temporary, hypersaline and toxic habitats. During recent years, fish living in extreme habitats have become emerging model species for studying adaptation and predictability of evolution (Riesch, Plath, Schlupp, Tobler, & Langerhans, 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, Benestan, Mouillot, Lobreaux, & Manel, 2018; Garcia-Elfring et al., 2021), light (Marques et al., 2017), toxins (Pfenninger et al., 2014; Reid et al., 2016), acidic (Haenel, Roesti, Moser, MacColl, & Berner, 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; Breitburg 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, Monteith, & Cooper, 2005; Forsberg, 1992; Kritzberg et al., 2020; Roulet & Moore, 2006; Solomon et al., 2015; Vuorenmaa, Forsius, & Mannio, 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, Morel, & Prieur, 1981). Due to the heterogeneity of dissolved organic matter, it is usually quantified as dissolved organic carbon (DOC; Wood & Salzberg, 2014). DOC plays a significant role in freshwater ecosystems by

74 driving the carbon and energy cycle, and also contributes significantly to greenhouse gas emissions

75 (Battin et al., 2009; Sobek, Söderbäck, Karlsson, Andersson, & Brunberg, 2006; Tranvik, Cole, &

76 Prairie, 2018). Humic (also known as dystrophic) lakes containing high levels of DOC are extreme,

77 almost "nocturnal" visual environments – both down-welling short wavelength and almost all up- or

78 side-welling light is absorbed, while only long wavelength red light is able to partially penetrate the

79 water column (Eloranta, 1978; Jones, 1992). A striking example of the evolutionary adaptation to long

80 wavelength visual environments has been described in three-spined stickleback (*Gasterosteus*

81 *aculeatus*), where non-synonymous mutations observed in the SWS2 opsin gene have caused a red-shift

82 in light absorption of the visual photo pigment (Marques et al., 2017). Adaptation to extreme visual

83 environments may also include adjustments at the gene expression level, and/or gene duplications

84 (Carleton, Escobar-Camacho, Stieb, Cortesi, & Marshall, 2020; Musilova et al., 2019).

85 In addition to the visual environment, DOC affects a myriad of other habitat parameters and biotic

86 interactions. Importantly, a high level of DOC contributes to acidification in low alkalinity and weakly

87 buffered waters (Hope, Kratz, & Riera, 1996; Sobek, Algesten, Bergström, Jansson, & Tranvik, 2003).

88 Therefore, humic lakes generally have low pH levels (Arvola et al., 2010; Erlandsson, Cory, Köhler, &

89 Bishop, 2010; Keskitalo & Eloranta, 1999). Freshwater fish living in acidic water must constantly import

90 ions from their food and environment to compensate for diffusive ion losses, primarily from the gills

91 (Dymowska, Hwang, & Goss, 2012). Thus, it is likely that low pH represents a strong selective force

92 and physiological challenge for teleosts in humic lakes, especially during early life stages (Parra &

93 Baldisserotto, 2007; Rask, 1984). On the other hand, DOC may also partially protect against acidic water

94 by altering gill membrane permeability and stimulating ion uptake (Wood, Al-Reasi, & Smith, 2011).

95 Acidification and increased DOC can, in turn, affect ion composition in humic lakes, where the

96 concentration of essential ions, including calcium, is very low (Weyhenmeyer et al., 2019). Calcium

97 plays an important role in many cellular functions, and is a critical component of body structures, such

98 as bones, carapaces, scales and shells (Greenaway, 1985). Moreover, humic lakes usually exhibit steep

99 thermal and oxygen stratification, leading to hypoxia in deeper areas (Bastviken, Cole, Pace, & Tranvik,

100 2004; Kankaala, Huotari, Peltomaa, Saloranta, & Ojala, 2006). Hence, increased DOC can also limit the

whole-lake primary production and underwater higher vegetation, supporting the microbial loop and influencing the 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 and Path 2011; Kaeuffer et al. 2012; Noreikiene, et al. 2020).

Total biomass production, body growth, as well as mean body size of Eurasian perch (*Perca fluviatilis*) have been shown to differ between humic and non-humic lakes (van Dorst et al., 2019). Therefore, selective agents in humic lakes are likely 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 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, Roesti, & Schluter, 2019).

Eurasian perch (*P. fluviatilis*) 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 is abundant in both humic and clear-water lakes, and represents 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, Saarenheimo, Syväranta, & Jones, 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)

128 and up to one third of known freshwater fish species are associated with tropical peat swamps containing

129 high levels of DOC (Ng, Tay, & Lim, 1994; Posa, Wijedasa, & Corlett, 2011), understanding the nature

130 of humic adaptation has a broad significance across species and geographic regions.

131   In this study, we characterize genomic signatures of adaptation to extreme humic environments

132 using whole-genome sequencing of perch sampled from 16 humic and 16 clear-water lakes in northern

133 Europe. We predicted that this adaptive process may involve both well-characterized genes with known

134 function (e.g. visual performance, maintenance of ionic balance), as well as novel targets. By using over

135 800,000 high quality SNPs and applying genome-wide divergence and environmental association

136 approaches, we aimed to ascertain whether adaptation occurs by major shifts of allele frequencies in a

137 few key genes (e.g. visual opsins and ion transporter genes), or if footprints of selection are scattered

138 across the genome and include many regions and genes with diverse functions. In addition, we tested if

139 signatures of selection are randomly distributed across the genome or if they are concentrated within or

140 nearby coding and regulatory regions. Finally, we evaluated if genes putatively under divergent selection

141 between humic and clear-water environments are associated with specific biological processes and

142 molecular functions using Gene Ontology (GO) analysis.

143

144 **MATERIAL AND METHODS**

145 **Biological samples**

146   In total, 32 perch individuals were sampled from 16 humic and 16 clear-water lakes in northern

147 Europe (a single specimen per lake, as in Jones et al. (2012); Fig. 1a, Table S1). The selection of lakes

148 for the analysis was based on drastic differences in water coloration and geographic proximity of

149 different type of lakes to increase the power of detecting loci under selection (De Mita et al., 2013;

150 Lotterhos & Whitlock, 2015). Most fish were caught using a gill-net, beach seine or rod (Table S1). Egg

151 ribbons were collected from two Swedish lakes (Nedre Björntjärnen and Snottertjärn) and transported

152 to laboratory facilities (Institute of Freshwater Research, Drottningholm, Sweden). After hatching, fry

153 were euthanized using a benzocaine overdose and stored in 96% ethanol without measuring fork length,

154 body mass or determining sex. The fish from the remaining lakes were sacrificed with a sharp blow to

155   the head. Thereafter, fork or total length (to the nearest mm) and wet body mass (to the nearest g) were

156   measured, sex was determined by visual examination of the gonads, and a tissue sample (pelvic fin) was

157   placed in 96% ethanol for DNA extraction.

158   The requirements outlined in the Annex III (Requirements for establishments and for the care and

159   accommodation of animals) and Annex IV (Methods of killing animals) Section B point 11 of the

160   "Directive 2010/63/EU of the European parliament and of the council of 22 September 2010 on the

161   protection of animals used for scientific purpose" were fully met. The authors have followed the

162   principles of the 3Rs (Replacement, Reduction and Refinement) and have involved the minimum

163   number of animals to produce statistically reproducible results. The collection of samples were

164   conducted in accordance with national legislation based on permits issued by Estonian Ministry of

165   Environment (54/2016) and the regional ethical review board in Uppsala, Sweden (Dnr 5.8.18-

166   03449/2017). Sampling in Finland and Lithuania was carried out using recreational fishing gear (rod

167   and line), following the fishing rules set by the national legislations, and ethical issues related to

168   recreational fishing and handling of caught fish recommended by the recreational fishing federations of

169   the countries.

170

171   **Environmental data**

172   Lake water was collected at the time of fish sampling for chemical analysis. Two measures

173   reflecting the amount of dissolved organic matter in the water – dissolved organic carbon (DOC; mg/l)

174   and water color (mg Pt/l) – were analyzed as described in Sepp, Kõiv, Nõges, and Nõges (2019). Lake

175   surface area and lake shoreline length were manually estimated using satellite photos on Google maps

176   (https://www.google.com/maps/; Table S1) to estimate the size of the lakes. Both DOC and water color

177   values were log-normalized prior to environmental association analyses. DOC and water color were

178   highly correlated (Pearson's r = 0.96, $P < 0.001$, Fig.1b), and both showed significant differences

179   between humic and clear-water lakes (Mann-Whitney test, both $P < 0.001$). Neither lake surface area

180   nor shoreline distance differed between humic and clear-water lakes (Fig. S1).

181

DNA extraction and sequencing

Total genomic DNA (gDNA) was extracted from the tissue samples using *NucleoSpin® Tissue kit* (*Macherey-Nagel*, Dueren, Germany) according to the manufacturer's protocol. The quality of the total gDNA was assessed using Fragment Analyzer (Advanced Analytical), and the concentration was measured using Qubit® Fluorometric Quantitation (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 the automated cluster preparation using Illumina cBot station. 16 libraries (each corresponding to one perch sample from Estonia) were pooled and sequenced in 8 lanes on 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 4 lanes on Illumina NovaSeq 6000 using paired-end sequencing (2×150 bp read length with 8 bp index).

SNP calling and filtering

Read quality was assessed using FastQC v.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 v.0.36 (Bolger, Lohse, & Usadel, 2014; CROP:149 HEADCROP:9 SLIDINGWINDOW:5:25 MINLEN:60) for the 16 samples sequenced on the HiSeq 3000 instrument, or fastp v.0.20 (Chen, Zhou, Chen, & Gu, 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.

Filtered sequence reads of each individual were mapped to the Eurasian perch reference genome (NCBI: GCA_010015445.1) using Bowtie2 v.2.3.5.1 (Langmead, Trapnell, Pop, & Salzberg, 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

208 perch individuals was 29.9, and varied from 24.0 to 40.6. Raw SNPs were called using two alternative

209 pipelines. First, SAMtools v.1.10 (Li, 2011) was applied on the locally realigned and sorted BAM files

210 (samtools mpileup -uIg -t DP,AD,INFO/AD,ADF,ADR,SP -q 20) with the following variant calling

211 using bcftools v.1.8 (Li, 2011). Second, HaplotypeCaller subroutine from GATK v.4.1.4.1 (McKenna

212 et al., 2010) with similar parameters was applied to call variants using the same BAM files (-ERC GVCF

213 --minimum-mapping-quality 20 -mbq 13 --indel-size-to-eliminate-in-ref-model 12 -G

214 AS_StandardAnnotation -G StandardAnnotation) with the following import of single-sample GVCFs

215 into GenomicsDB using GenomicsDBImport, and final calling of consensus genotypes with

216 GenotypeGVCFs. Further quality filtering of raw variants generated by the two pipelines was performed

217 using VCFtools v.0.1.15 (Danecek et al., 2011) retaining only the variants that met the following criteria:

218 (i) minimum and maximum mean sequencing depth (d) was set to 10 (as a minimum recommended by

219 Illumina) and 66 (max depth $= d_{max} + 4\sqrt{d_{max}}$; Li (2014)), respectively; (ii) the consensus quality was

220 ≥ 30; (iii) a variant had at least two copies of an allele; (iv) no missing data were allowed; (v) only bi-

221 allelic sites were included; (vi) a variant did not occur in repetitive genomic regions (--max-meanDP 66

222 --min-meanDP 10 --max-missing 1 --mac 2 --min-alleles 2 --max-alleles 2 --minQ 30 --exclude-bed

223 Pfluv_repeats.bed).

224 In total, 1,078,577 and 1,074,879 SNPs were retained after samtools-bcftools and GATK pipelines,

225 respectively. To ensure the quality and reliability of the dataset, only the variants consistently called by

226 both pipelines were retained (1,025,544 SNPs). To further exclude false heterozygous calls and regions

227 with an excess of variable sites, the additional quality control filtering was applied: number of

228 heterozygotes per SNP ≤ 16, MAF ≥ 0.05, which resulted in 810,591 SNP loci in the final dataset.

229 The identified SNPs were annotated using SnpEff v.5.0 (Cingolani et al., 2012). The SnpEff

230 database was generated using the Eurasian perch reference genome sequence and annotation (NCBI:

231 GCA_010015445.1).

232

233 **Population diversity and structure**

234   Observed heterozygosity per individual was calculated using VCFtools v.0.1.15 (Danecek et al.,

235   2011). The R-package adegenet v.2.1.3 (Jombart, 2008; Jombart & Ahmed, 2011) was used to convert

236   SNP-data into a genind object. Overall population genetic structure was examined by applying principal

237   component analysis (PCA) using the dudi.pca function of the ade4 v.1.7-16 R-package (Dray & Dufour,

238   2007) on three sets of SNP loci: a) all 810,591 SNPs; b) putatively neutral 256,880 SNPs located

239   exclusively in the intergenic regions; and c) 10,245 candidate SNPs potentially under natural selection.

240   The percent of variation explained by each PC axis was extracted using the factorextra v.1.0.7 R-package

241   (Kassambara & Mundt, 2020). Pairwise genetic differentiation among the samples was estimated as

242   pairwise mean absolute allele frequency differences (D (Prevosti, Ocaña, & Alonso, 1975)). Genetic

243   relationships among the samples were explored using Nei's genetic distances (Nei, 1972) in the R-

244   package poppr v.2.8.2 (Kamvar, Brooks, & Grünwald, 2015; Kamvar, Tabima, & Grünwald, 2014);

245   neighbor-joining trees (Saitou & Nei, 1987) were constructed with 100 bootstrap replicates among loci

246   using the R-package ape v.5.4-1 (Paradis & Schliep, 2019).

247

248   **Signatures of selection**

249   Candidate SNPs potentially under natural selection were identified using three approaches: one

250   population divergence and two environmental association methods.

251

252   *Highly divergent loci*

253   The genetic divergence of each SNP locus was estimated as a mean absolute allele frequency

254   difference (δ) between humic and clear-water lakes. Candidate SNPs were identified as SNPs with a δ

255   that was higher than 2.5 SD (standard deviations) from the mean δ (δ threshold = 0.268; empirical two-

256   tailed P-value = 0.012; Miller, 1991).

257

258   *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 analyzed allele frequencies and environmental variables of each lake with linear regressions. Further, principal component analysis 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 color on the estimated allele frequencies. The effect of genetic and spatial structure was controlled using PC1 loadings based on putatively neutral intergenic SNPs (Fig. S3) and geographic coordinates (latitude and longitude) of the sampled lakes. For the RDA, we used the rda function of the R package vegan v.2.5-7 (Oksanen et al., 2020). Candidate SNPs were identified on each of the first three ordination axes as SNPs that had a 'locus score' ± 2.5 SD from the mean score for that axis (RDA loading threshold at axis $1 = \pm 0.0202$ and at axis $2 = \pm 0.0196$; empirical two-tailed P-value = 0.012; Dalongeville et al., 2018; Forester, Jones, Joost, Landguth, & Lasky, 2016; Miller, 1991; Xuereb, Kimber, Curtis, Bernatchez, & Fortin, 2018).

Latent factor mixed model (LFMM)

Latent Factor Mixed Models (LFMM) are factor regression models in which allele-environment correlations are estimated between each locus and each environmental variable (Frichot & François, 2015; Frichot, Schoville, Bouchard, & François, 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 v.1.0 package in R, implementing a least-squares approach for latent factor estimation (Caye, Jumentier, Lepeule, & François, 2019). Here, we used $K = 2$ to correct for population structure inferred using the PCA of putatively neutral intergenic SNPs (see results, Fig. S3). The significance of the association was tested and P-values were calibrated using the

286 genomic control method (Devlin & Roeder, 1999). Genomic control uses a robust estimate of the

287 variance of z-scores called "genomic inflation factor". The *P*-value threshold ($P \leq 0.012$) that showed

288 significant association with DOC or/and water color was chosen to correspond to the two other methods

289 used to identify putative signatures of selection.

290

Candidate SNPs supported by multiple methods

292   The final set of candidate SNPs potentially under natural selection included 10,245 candidates,

293 which were identified using at least two methods (Fig. 2e). Based on the empirical *P*-value threshold of

294 0.012 and assuming that detection of candidate loci is independent in each outlier method, we expect

295 445 false positives (3.4%) among the identified candidate SNPs.

296   To test for excess and deficiency of identified candidate SNPs in each annotation category, a chi-

297 squared test was performed using stats v.3.6.2 package in R to compare the 10,245 candidate SNPs with

298 all identified SNPs. Genes that were located in genomic regions with at least one candidate SNP in exon,

299 intron or regulatory sequences, including 5K up- and downstream regions, were considered as candidate

300 genes.

301   In addition, we used the circlize v.0.4.12 (Gu, Gu, Eils, Schlesner, & Brors, 2014) package in R to

302 visualize the distribution of all and candidate SNPs, as well as the genes, genomic divergence and

303 diversity across the perch genome. The density of SNPs, genes and candidate SNPs across the genome

304 was estimated at each chromosome based on half overlapped 500 Kb windows using the

305 genomicDensity function of the circlize R package. In addition, the density of candidate SNPs was

306 estimated in 25 Kb windows using the same package. Genomic regions were identified as regions with

307 a high density of candidate SNPs if the density estimate was higher than 2.5 SD from the mean density.

308

Gene ontology (GO) analysis

310   In order to carry out GO analysis, we first identified perch genes that are orthologous to human and

311 zebrafish genes using the rentrez (Winter, 2017) package in R. GO enrichment analysis of candidate

312 genes against all orthologous gene symbols as a background was performed using a binomial test in

313   Panther (Thomas et al., 2003). The GO terms with a false discovery rate (FDR) ≤ 0.05 were considered

314   as significant.

315

316   **RESULTS**

317   **Population genomic variation and structure**

318   Overall genome-wide genetic diversity (estimated as observed heterozygosity, H_O) varied from

319   0.064 (ELOO) to 0.309 (EUIA) with a mean $H_O = 0.215$ (Table S1). Average genetic diversity among

320   perch in humic lakes (mean $H_O = 0.195$, median $H_O = 0.221$) was slightly lower compared to that in

321   clear-water lakes (mean $H_O = 0.236$, median $H_O = 0.240$), although the difference was non-significant

322   (Fig. S2a). There was a weak, non-significant tendency for genetic diversity to decrease with increasing

323   DOC and water color, and increase with lake size (Fig. S2b-e).

324   Overall genome-wide genetic divergence (measured as D; Prevosti et al., 1975) among the samples

325   was $D = 0.309$, and ranged from $D = 0.204$ (FIMU vs. FHYV) to $D = 0.356$ (FITA vs. ELOO; Table

326   S2). The first two PCA axes based on 810,591 SNPs explained nearly 79% of genomic variation, and

327   depicted a genetic structure corresponding to geographic origin of perch (Fig. 2a). Similarly, the

328   topology of the neighbour-joining tree based on D_{Nei} genetic distances reflected the main clusters

329   identified by the PCA (Fig 2c). In general, perch samples clustered into the two main groups: (a) North-

330   Eastern Baltic, including the samples from Finland, and (b) South-Western Baltic, including perch from

331   Sweden, Estonia and Lithuania.

332

333   **Candidate SNPs under divergent selection**

334   The combination of genetic divergence and two environmental association analyses revealed

335   10,245 candidate SNPs (Fig. 2e). Based on the PCA using these 10,245 candidate SNPs, 6% of the

336   variation on the second axis was explained by humic vs. clear-water separation, whereas 65% was still

337   attributed to the geographic origin (Fig. 2b). Similarly, the topology of the neighbor-joining tree based

338   on D_{Nei} distances using 10,245 candidate SNPs showed the separation of humic and clear-water perch

339   within North-Eastern and South-Western Baltic groups (Fig 2d). Significant enrichment (chi-square test:

340    $\chi^2 = 11.55$–85.50, $P = 2.3\times10^{-20}$–4.8×10^{-4}) and depletion (chi-square test: $\chi^2 = 3.98$–37.55, $P = 4.6\times10^{-2}$

341   ^{2}–8.9×10^{-10}) for candidate SNPs was observed in seven and ten chromosomes, respectively (Fig. 2, Table

342   S3). Furthermore, the distribution of candidate SNPs across the chromosomes was not even; several

343   peaks of high density were identified (Fig. 3). The number of genomic regions with a high density of

344   candidate SNPs varied from 1 to 6 (mean = 2.5) or from 15 to 41 (mean = 27.4) per chromosome when

345   using 500 Kb or 25 Kb windows, respectively. The total number of regions with a high density of

346   candidate SNPs was 50 or 658 when using 500 Kb or 25 Kb windows, respectively (Table S4). Most of

347   the candidate SNPs were located in intergenic regions (32.3%) and introns (33.8%), whereas ~ 4% were

348   found in exons (Table 1). The 5'UTR, 3'UTR and 5K downstream gene regions were enriched for

349   candidate SNPs (chi-square test: $\chi^2 = 8.88$–5.42, $P = 2.0\times10^{-2}$–2.8×10^{-3}; Table 1). In contrast, intergenic

350   regions were depleted for candidate SNPs (chi-square test: $\chi^2 = 9.76$, $P = 1.9\times10^{-3}$; Table 1).

351

352   **Candidate genes under divergent selection**

353   In total, 10,245 candidate SNPs were located within or nearby (5 Kb up- or downstream) 3,245

354   candidate genes (Table S5). Since the strongest selective sweeps are expected to affect multiple adjacent

355   SNPs, further ranking of the candidate gene list (average $\delta > 0.4$, > 6 SNPs per gene) revealed 31 of the

356   strongest candidates (Table S5). A large proportion of genes among these are involved in anatomical

357   structure development (*ASXL1*, *CDON*, *ECE1*, *FGF11*, *FLII*, *HOXB9*, *LRRN1*, *MYLIP*, *PLAGL2*, *RTN1*,

358   *TDRD9*, *TTLL3*, *TTN*) and regulation of nervous system development (*MYLIP, BHLHE40, CDON*,

359   *FGF11*, *LRRN1* and *RTN1*). The highest number of SNPs was observed in the *MYLIP* gene located on

360   chromosome 13 (Fig. 4c). This gene is involved in various biological processes, including multicellular

361   organism development, regulation of plasma membrane bounded cell projection organization, nervous

362   system development, etc. (Bornhauser, Olsson, & Lindholm, 2003; Calkin et al., 2011; Olsson,

363   Korhonen, Mercer, & Lindholm, 1999). The functions of other strong candidate genes included

364   regulation of voltage-gated sodium channel activity (*FGF11*, *SLMAP*), inorganic ion transmembrane

365   transport (*KCNMA1*, *TMEM163*; Li et al., 2018) and regulation of circadian rhythm *(BHLHE40;* Honma

366   et al., 2002*).*

367   We observed several 50–100 Kb regions of elevated δ between humic and clear-water perch in

368 multiple chromosomes, involving genes adjacent to the strongest candidates (Fig. 4a, b, d). For example,

369 the ca. 100 Kb region on chromosome 5 included genes regulating anatomical structure development

370 (*ASXL1*, *MYT1L*, *PLAGL2*), oxidative stress responding regulation (*PLAGL2*), response to oxygen-

371 containing compound (*ASXL1*) and mitochondrial fusion (*MFN2*); the ca. 60 Kb region on chromosome

372 7 contained genes regulating response to stress (*RPA2*, *THEMIS2*), lipid (*CYP4B1*) and glycogen

373 (*PPP1R8*) metabolism; the ca. 100 Kb region on chromosome 15 involved genes regulating ion transport

374 (*LLGL1*, *SHISA9*, *DESI1*) and cell differentiation (*FLII*, *LLGL1*). As expected, most of the regions

375 showing elevated δ also showed a reduction in H_O among humic perch, suggesting that directional

376 selection associated with high DOC content has reduced the genetic diversity of adjacent genomic

377 regions.

378   In contrast to our prior expectations, SNPs found in visual opsin genes (orthologous to human

379 *OPN1LW*, *OPN1MW*, *OPN1SW*, *RHO*) did not show high allele frequency differences between humic

380 and clear-water populations (Fig. S4), suggesting that the strongest selective sweeps associated with

381 humic adaptation in perch do not involve visual opsins. However, we found a non-synonymous

382 candidate SNP in one of the four red-sensitive opsin-like gene orthologues *(OPN1LW)* showing

383 moderate allele frequency difference between humic and clear-water perch ($AF_{HUMIC} = 0.53$ vs. AF_{CLEAR}

384 $= 0.25$; Fig. 4e). Another non-synonymous mutation exhibiting high allele frequency difference

385 ($AF_{HUMIC} = 0.19$ vs. $AF_{CLEAR} = 0.59$) was observed in a non-visual *opsin 7: group member a* gene

386 (*OPN5*/*opn7a*, $\delta = 0.41$; Fig. 4f). The other non-synonymous mutations in visual or non-visual opsin

387 genes did not show high allele frequency differences or significant associations with environmental

388 parameters (Supplementary file 1).

389

390 **Gene ontology analysis**

391   The overall assessment of the gene functions using GO analyses showed that candidate genes were

392 enriched for 134 GO biological process terms, 98 GO biological component terms and 10 GO molecular

393 function terms (FDR ≤ 0.05; Fig. 5; Table S6). The top 10 most significant GO terms included regulation

394 of organism development and nervous system development (biological process); actin and calmodulin

395 binding (molecular function); and cell junction, including nervous tissues (cellular component). A group

396 of genes among those involved in calmodulin binding regulate calcium (calcium-transporting *ATP*-ases)

397 and potassium (*KCN* genes) transport, as well as sodium-calcium exchange (*SLC8* gene family, Table

398 S6), and play an important role in osmoregulation (Hwang, Lee, & Lin, 2011; Pallone, Khurana, & Cao,

399 2012; Pizzagalli, Bensimon, & Superti-Furga, 2021).

400

401 **DISCUSSION**

402   Based on analysis of 32 whole genomes, we discovered that footprints of selection associated with

403 humic environments comprises hundreds of regions scattered across the Eurasian perch genome.

404 Putative signals of adaptation were detected in genes and gene families with diverse functions. Most

405 frequently, the candidate genes were involved in the regulation of organism development, nervous

406 system development and calcium/potassium/sodium exchange, highlighting their role during early

407 development and in ion balance maintenance. In contrast, we did not observe strong evidence of

408 selection on visual opsins, despite the extreme differences in visual environment between the studied

409 lakes. The observed overrepresentation of candidate SNPs in regulatory (5'UTR, 3'UTR and 5K

410 downstream gene regions) genomic regions and the high number of candidates in intergenic and intronic

411 regions suggest that humic adaptation mainly occurs via regulatory changes rather than changes in amino

412 acids.

413

414 **Role of regulatory regions in humic adaptation**

415   Most candidate SNPs were detected in intergenic and intronic regions, which comprise a large part

416 of the perch genome (Ozerov et al., 2018). At the same time, we observed a significant excess of

417 candidate SNPs in regulatory regions (5'UTR, 3'UTR and 5K downstream gene regions), indicating their

418 important role in humic adaptation of perch. On the other hand, the proportion of non-synonymous

419 mutations among candidates did not show an increase compared to the whole dataset. Thus, our findings

420 are in line with a growing body of evidence suggesting that natural selection is predominantly acting at

421   regulatory regions (e.g. Fagny & Austerlitz, 2021; Fraser, 2013; Glaser-Schmitt & Parsch, 2018; Verta

422   & Jones, 2019). For example, Fraser (2013) showed that local adaptations found in a subset of human

423   populations are over 10-fold more likely to affect gene expression than to alter protein sequences. The

424   excess of candidate SNPs in 5' and 3' untranslated regions observed in perch is in agreement with this,

425   as UTRs are among the main elements involved in the regulation of gene expression (Barrett, Fletcher,

426   & Wilton, 2012). 5'UTRs, located upstream of the protein coding sequence, play an important role in

427   translation initiation, and expression of alternative 5'UTRs allows variation in expression from a single

428   gene and tissue-specific expression patterns (Barrett et al., 2012). 3'UTRs, located downstream of the

429   protein coding region, impact post-transcriptional and translational processes, including mRNA

430   localization (Andreassi & Riccio, 2009), stability (Goldstrohm & Wickens, 2008) and expression levels

431   (Matoulkova, Michalova, Vojtesek, & Hrstka, 2012). In general, 3'UTRs are more polymorphic and

432   variable in length compared to 5'UTRs, resulting in a greater evolutionary potential of these regulatory

433   elements (Barrett et al., 2012; Steri, Idda, Whalen, & Orrù, 2018). Indeed, the observed number of SNPs

434   in 3'UTR regions in the perch genome was nearly four times higher compared to that in 5'UTRs (Table

435   1). The important role of 3'UTRs in teleost evolution has been recently highlighted in cichlid fishes,

436   suggesting that these regions might act as meta-regulators (i.e., regulators of other mechanisms

437   governing post-transcriptional regulation), particularly in species undergoing rapid adaptation and

438   speciation (Xiong, Hulsey, Meyer, & Franchini, 2018). Therefore, our results suggest that this may also

439   be the case for humic adaptation in perch, where selection predominantly occurs via regulatory changes.

440

441   **Candidate genes and gene families involved in humic adaptation**

442   We detected multiple signals of selection consisting of over 3,000 genes scattered across the perch

443   genome. These candidate genes are involved in multiple biological processes and cellular functions,

444   most significant of which are multicellular organism processes and system development, actin and

445   calmodulin binding and ion transfer. Several genes that showed elevated genetic divergence between

446   humic and clear-water perch were also involved in development processes. For example, the signal of

447   divergent selection involving more than 30 SNPs was centered around the *MYLIP* gene, essential for

448 embryonic development (Knowlton, Chan, & Kelly, 2003). Moreover, the role of *MYLIP* in early

449 embryonic development of zebrafish involves calcium-dependent mechanisms during gastrulation

450 (Knowlton et al., 2003). Therefore, it is possible that the observed adaptive variation in perch is linked

451 to Ca^{++} deficiency compensation during embryonic development in humic lakes. However, drawing firm

452 conclusions about the role of *MYLIP* in perch embryogenesis and humic adaptation requires further

453 investigation (e.g. Kurko et al., 2020).

454   Similar to *MYLIP*, many other genes involved in embryonic development were found in the regions

455 of elevated genetic divergence. For example, a region on chromosome 5 includes genes shown to

456 regulate brain and eyes (*PLAGL2*; Pendeville, Peers, Kas, & Voz, 2006), hypothalamus (*MYT1L*;

457 Blanchet et al., 2017), axonal and neuromuscular (*MFN2*; Vettori et al., 2011), and neutrophil

458 development (*ASXL1*; Fang et al., 2021) in zebrafish. The *LLGL1* gene located on chromosome 15 was

459 found to regulate zebrafish cardiac development (Flinn et al., 2020). Other regions located on

460 chromosomes 7 and 15 included genes involved in immune response, such as the T-cell receptor

461 signaling pathway (*THEMIS2*; Cheng et al., 2017; Peirce et al., 2010), inflammation and wound repair

462 (*FLII*; Strudwick & Cowin, 2020), as well as genes involved in DNA replication and reparation (*RPA2*,

463 *DESI1*; Mer et al., 2000; Shin et al., 2012). However, more elaborate molecular and ecological studies

464 are needed to shed light on specific phenotypic consequences and fitness effects of these and other

465 candidate genes.

466   Among the candidate genes, we identified a large group of genes encoding transmembrane

467 transportation of ions, such as solute carriers (*SLC*s; Hediger et al., 2004), calcium channel subunits

468 (*CACN*s; Catterall, Perez-Reyes, Snutch, & Striessnig, 2005), potassium channel subunits (*KCN*s; Shieh,

469 Coghlan, Sullivan, & Gopalakrishnan, 2000) and sodium channel subunits (*SCN*s; Yu & Catterall,

470 2003). There were also genes involved in energy-dependent transportation, such as ATP-binding

471 cassette transporters (*ABC*s; Jones & George, 2004). Given the low pH and low level of calcium ions in

472 humic lakes, we expected a major role of Na^+/H^+, K^+ and Ca^{++} exchangers during osmoregulation and

473 internal homeostasis maintenance in humic perch (Hwang et al., 2011). Accordingly, among the

474 candidate genes were several *SLC* family 9 genes (*SLC9A1*, *SLC9A3R1*, *SLC9A5*, *SLC9A6* and *SLC9B2*)

475 and one *SLC* family 4 gene (*SLC4A5*), which play an important role in regulation of cellular pH via the

476 transport of bicarbonate ions and protons (Pizzagalli et al., 2021). Moreover, selection signals were also

477 observed in *SLC* family 8 (*SLC8A1* and *SLC8A2*) and 24 (*SLC24A2*), and in calcium channel subunits

478 alpha (*CACNA1A*, *CACNA1B*, *CACNA1C*, *CACNA1D*, *CACNA1E*, *CACNA1G*, *CACNA1I*, *CACNA2D1*

479 and *CACNA2D4*), which play a significant role in the regulation of intracellular Ca^{++} concentrations and

480 Ca^{++} influx (Pallone et al., 2012; Pizzagalli et al., 2021). Similarly, selection signals were found in

481 several genes linked to pH maintenance via ion and acid-base equivalent exchanges, such as *SCN*s

482 (*SCN1B*, *SCN3B*) and *KCN*s (*KCNA7*, *KCNAB1*, *KCNC3*, *KCNC4*, *KCNE1*, *KCNF1*, *KCNH2*, *KCNIP4*,

483 *KCNJ12*, *KCNK13*, *KCNMA1*, *KCNN2*, *KCNQ1*, *KCNQ2*, *KCNQ3*, *KCNQ5*). Therefore, our results

484 strongly indicate that adaptation to high DOC concentrations includes a large group of genes involved

485 in ion transport and balance. The abovementioned gene families have also been associated with

486 adaptation to acidic (Haenel et al., 2019) and alkaline (Xu et al., 2017) environments in three-spined

487 stickleback (*G. aculeatus*) and Amur ide (*Leuciscus waleckii*), respectively. Taken together, our results

488 corroborate findings that variation in water chemistry, including DOC content, ion concentration and

489 pH (Parra & Baldisserotto, 2007; Rask, 1984; Wood et al., 2011), is a strong selective force shaping

490 molecular mechanisms of adaptation in teleosts.

491

492 **Opsins**

493 Although we found several non-synonymous substitutions in green- and red-sensitive opsins and

494 in rhodopsin, the allele frequency differences between humic and clear-water habitats were small and

495 could be explained by random drift alone. One potential exception was the single non-synonymous SNP

496 in one of the four red-sensitive opsin-like gene orthologues *(OPN1LW)* that showed moderate allele

497 frequency differences between humic and clear-water perch. However, as this putative selective sweep

498 region consisted of several genes (e.g. *TFE3*, *CXXC1*) and multiple SNPs with high allele frequency

499 differences, it is not trivial to pinpoint the specific variant under divergent selection. Furthermore, based

500 on relatively small differences of genetic diversity among humic and clear-water perch in this region,

501 we cannot exclude the possibility that selection has actually occurred in clear-water environments. Thus,

502   in contrast to findings in three-spine stickleback (Marques et al., 2017) and Baltic herring (Hill et al.,

503   2019), our results suggest that amino acid changes in perch visual opsins most likely do not play a main

504   role in adaptation to extreme visual environments. Nevertheless, recent work based on RNA-seq analysis

505   of the whole eye in Eurasian perch have shown differential expression of visual red- green- and short

506   wavelength sensitive opsins (*OPN1LW/opn1lw1*, *OPN1MW/opn1mw1* and *OPN1SW/opn1sw2*,

507   respectively) between humic and clear-water environments (Noreikiene et al., 2020), indicating that

508   regulation of visual opsin expression still occurs in humic lakes.

509   We also detected a non-synonymous substitution in the non-visual *opsin 7a* (*OPN5/opn7a)* that had

510   a large allele frequency difference between humic and clear-water perch ($\delta = 0.41$; 99.94 percentile of

511   all non-synonymous SNPs). The opsin 7 family shows responsiveness at wavelengths < 380 nm, which

512   might be expected for a tissue that is exposed to direct sunlight. These genes are expressed in various

513   tissues including the brain, digestive system, eye, heart and testis (Davies et al., 2015; Liu et al., 2020).

514   However, as the SNP diversity adjacent to *OPN5/opn7a* was rather similar in both habitats, the role of

515   this non-synonymous variation in the context of adaptation to contrasting visual environments needs

516   further investigation.

517

518   **Conclusions**

519   Our findings demonstrate that humic adaptation in perch comprises a large number of regions and

520   genes scattered across the genome. The excess of putatively adaptive variants that are found in 5'UTR,

521   3'UTR and 5K downstream gene regions highlights the importance of adaptive evolution via

522   regulatory elements, rather than via amino acid modifications in proteins. Putative adaptation signals

523   were detected in genes and gene families with diverse functions, including genes involved in organism

524   development, plasma membrane and ion transportation, underlying the multifaceted nature of humic-

525   driven selection. Our study demonstrates the power of whole genome analysis to identify the most

526   promising candidates involved in adaptation to complex environmental conditions, but also highlights

527   the challenge of moving from high-throughput outlier identification towards functional

528   characterization of candidate genes underlying phenotypic traits of ecological and evolutionary

529   importance.

543 **REFERENCES**
544 Altieri, A. H., & Gedan, K. B. (2015). Climate change and dead zones. *Global Change Biology, 21*(4),
545   1395-1406. doi:https://doi.org/10.1111/gcb.12754
546 Andreassi, C., & Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends.
547   *Trends in Cell Biology, 19*(9), 465-474. doi:10.1016/j.tcb.2009.06.001
548 Andrews, S. (2017). FastQC: a quality control tool for high throughput sequence data.
549   *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.*
550 Arvola, L., Rask, M., Ruuhijärvi, J., Tulonen, T., Vuorenmaa, J., Ruoho-Airola, T., & Tulonen, J.
551   (2010). Long-term patterns in pH and colour in small acidic boreal lakes of varying
552   hydrological and landscape settings. *Biogeochemistry, 101*(1), 269-279. doi:10.1007/s10533-
553   010-9473-y
554 Ask, J., Karlsson, J., Persson, L., Ask, P., Byström, P., & Jansson, M. (2009). Terrestrial organic
555   matter and light penetration: Effects on bacterial and primary production in lakes. *Limnology*
556   *and Oceanography, 54*(6), 2034-2040. doi:https://doi.org/10.4319/lo.2009.54.6.2034
557 Barrett, L. W., Fletcher, S., & Wilton, S. D. (2012). Regulation of eukaryotic gene expression by the
558   untranslated gene regions and other non-coding elements. *Cellular and molecular life sciences*
559   *: CMLS, 69*(21), 3613-3634. doi:10.1007/s00018-012-0990-9
560 Bastviken, D., Cole, J., Pace, M., & Tranvik, L. (2004). Methane emissions from lakes: Dependence
561   of lake characteristics, two regional assessments, and a global estimate. *Global*
562   *Biogeochemical Cycles, 18*(4). doi:https://doi.org/10.1029/2004GB002238
563 Battin, T. J., Luyssaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., & Tranvik, L. J. (2009).
564   The boundless carbon cycle. *Nature Geoscience, 2*(9), 598-600. doi:10.1038/ngeo618
565 Blanchet, P., Bebin, M., Bruet, S., Cooper, G. M., Thompson, M. L., Duban-Bedu, B., . . . McNeill, A.
566   (2017). MYT1L mutations cause intellectual disability and variable obesity by dysregulating
567   gene expression and development of the neuroendocrine hypothalamus. *PLoS Genet, 13*(8),
568   e1006957. doi:10.1371/journal.pgen.1006957
569 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
570   sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170
571 Bornhauser, B. C., Olsson, P.-A., & Lindholm, D. (2003). MSAP Is a Novel MIR-interacting Protein
572   That Enhances Neurite Outgrowth and Increases Myosin Regulatory Light Chain*. *Journal of*
573   *Biological Chemistry, 278*(37), 35412-35420. doi:https://doi.org/10.1074/jbc.M306271200
574 Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., . . . Zhang, J.
575   (2018). Declining oxygen in the global ocean and coastal waters. *Science, 359*(6371).
576   doi:10.1126/science.aam7240
577 Bricaud, A., Morel, A., & Prieur, L. (1981). Absorption by dissolved organic matter of the sea (yellow
578   substance) in the UV and visible domains1. *Limnology and Oceanography, 26*(1), 43-53.
579   doi:https://doi.org/10.4319/lo.1981.26.1.0043
580 Calkin, A. C., Goult, B. T., Zhang, L., Fairall, L., Hong, C., Schwabe, J. W., & Tontonoz, P. (2011).
581   FERM-dependent E3 ligase recognition is a conserved mechanism for targeted degradation of
582   lipoprotein receptors. *Proceedings of the National Academy of Sciences of the United States of*
583   *America, 108*(50), 20107-20112. doi:10.1073/pnas.1111589108
584 Carleton, K. L., Escobar-Camacho, D., Stieb, S. M., Cortesi, F., & Marshall, N. J. (2020). Seeing the
585   rainbow: mechanisms underlying spectral sensitivity in teleost fishes. *Journal of Experimental*
586   *Biology, 223*(8). doi:10.1242/jeb.193334
587 Catterall, W. A., Perez-Reyes, E., Snutch, T. P., & Striessnig, J. (2005). International Union of
588   Pharmacology. XLVIII. Nomenclature and structure-function relationships of voltage-gated
589   calcium channels. *Pharmacological Reviews, 57*(4), 411-425. doi:10.1124/pr.57.4.5
590 Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and Accurate Inference of
591   Gene-Environment Associations in Genome-Wide Studies. *Molecular Biology and Evolution,*
592   *36*(4), 852-860. doi:10.1093/molbev/msz008
593 Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor.
594   *Bioinformatics, 34*(17), i884-i890. doi:10.1093/bioinformatics/bty560

Cheng, D., Deobagkar-Lele, M., Zvezdova, E., Choi, S., Uehara, S., Baup, D., . . . Cornall, R. J. (2017). Themis2 lowers the threshold for B cell activation during positive selection. *Nature Immunology, 18*(2), 205-213. doi:10.1038/ni.3642

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly, 6*(2), 80-92. doi:10.4161/fly.19695

Cole, J. J. (2009). Production in pristine lakes. *Nature, 460*(7254), 463-464. doi: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. doi:10.1139/f77-209

Creed, I. F., Bergström, A.-K., Trick, C. G., Grimm, N. B., Hessen, D. O., Karlsson, J., . . . 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. doi: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. doi:10.1186/s12864-018-4579-z

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330

Davies, W. I., Tamai, T. K., Zheng, L., Fu, J. K., Rihel, J., Foster, R. G., . . . 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. doi: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. doi:https://doi.org/10.1111/mec.12182

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics, 55*(4), 997-1004. doi: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. doi:10.1038/ncomms9787

Diehl, S. (1992). Fish Predation and Benthic Community Structure: The Role of Omnivory and Habitat Complexity. *Ecology, 73*(5), 1646-1661. doi:10.2307/1940017

Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *2007, 22*(4), 20. doi: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. doi: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. doi: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*(G3). doi: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. doi: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. doi:https://doi.org/10.1016/j.tig.2021.03.005

Fang, X., Xu, S. e., Zhang, Y., Xu, J., Huang, Z., Liu, W., . . . Zhang, W. (2021). Asxl1 C-terminal mutation perturbs neutrophil differentiation in zebrafish. *Leukemia*. doi:10.1038/s41375-021-01121-8

Flinn, M. A., Otten, C., Brandt, Z. J., Bostrom, J. R., Kenarsary, A., Wan, T. C., . . . Link, B. A. (2020). Llgl1 regulates zebrafish cardiac development by mediating Yap stability in cardiomyocytes. *Development, 147*(16). doi: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. doi: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. doi:10.1007/BF00006990

Fraser, H. B. (2013). Gene expression drives local adaptation in humans. *Genome Research, 23*(7), 1089-1096. doi: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. doi: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. doi:10.1093/molbev/mst063

Garcia-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. doi:10.1111/mec.15879

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. doi: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. doi:10.1038/nrm2370

Greenaway, P. (1985). Calcium Balance and Molting in the Crustacea. *Biological Reviews of the Cambridge Philosophical Society, 60*(3), 425-454. doi:DOI 10.1111/j.1469-185X.1985.tb00424.x

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. doi: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. doi: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. doi: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). doi: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. doi:10.1007/s00424-003-1192-y

Hill, J., Enbody, E. D., Pettersson, M. E., Sprehn, C. G., Bekkevold, D., Folkvord, A., . . . Andersson, L. (2019). Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. *Proceedings of the National Academy of Sciences, 116*(37), 18473. doi:10.1073/pnas.1908332116

Honma, S., Kawamoto, T., Takagi, Y., Fujimoto, K., Sato, F., Noshiro, M., . . . Honma, K. (2002). Dec1 and Dec2 are regulators of the mammalian molecular clock. *Nature, 419*(6909), 841-844. doi: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. doi: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. doi:10.1152/ajpregu.00047.2011

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*(11), 1403-1405. doi: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. doi:10.1093/bioinformatics/btr521

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., . . . Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature, 484*(7392), 55-61. doi:http://www.nature.com/nature/journal/v484/n7392/abs/nature10944.html#supplementary-information

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. doi: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. doi:Doi 10.1007/Bf00006992

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*, 208-208. doi: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. doi: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. doi: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. doi: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). Backhuys.

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. doi:https://doi.org/10.1016/j.ydbio.2003.09.001

Kritzberg, E. S., Hasselquist, E. M., Škerlep, M., Löfgren, S., Olsson, O., Stadmark, J., . . . Laudon, H. (2020). Browning of freshwaters: Consequences to ecosystem services, underlying drivers, and potential mitigation measures. *Ambio, 49*(2), 375-390. doi: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. doi:10.1534/g3.119.400882

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. doi:10.1186/gb-2009-10-3-r25

Legendre, P., & Legendre, L. (2012). *Numerical Ecology* (3rd ed.): Elsevier.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21), 2987-2993. doi:10.1093/bioinformatics/btr509

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics, 30*(20), 2843-2851. doi:10.1093/bioinformatics/btu356

Li, X., Poschmann, S., Chen, Q., Fazeli, W., Oundjian, N. J., Snoeijen-Schouwenaars, F. 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. doi:10.1038/s41431-017-0073-3

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. doi: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. doi:https://doi.org/10.1111/mec.13100

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. doi: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 Biol, 9*(5), 563-576. doi:10.4161/rna.20231

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research, 20*(9), 1297-1303. doi: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). Berlin, Heidelberg: Springer Berlin Heidelberg.

Mer, G., Bochkarev, A., Gupta, R., Bochkareva, E., Frappier, L., Ingles, C. J., . . . 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. doi:https://doi.org/10.1016/S0092-8674(00)00136-7

Miller, J. (1991). Reaction time analysis with outlier exclusion: bias varies with sample size. *Q J Exp Psychol A, 43*(4), 907-912. doi: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. doi: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., . . . Salzburger, W. (2019). Vision using multiple distinct rod opsins in deep-sea fishes. *Science, 364*(6440), 588. doi:10.1126/science.aav4632

Nei, M. (1972). Genetic Distance between Populations. *Am Nat, 106*(949), 283-292. Retrieved from http://www.jstor.org/stable/2459777

Nelson, J. S. (2006). *Fishes of the world.* Hoboken, N.J.: John Wiley.

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. doi:10.1007/BF00005667

Noreikiene, K., Ozerov, M., Ahmad, F., Kõiv, T., Kahar, S., Gross, R., . . . Vasemägi, A. (2020). Humic-acid-driven escape from eye parasites revealed by RNA-seq and target-specific metabarcoding. *Parasites & Vectors, 13*(1), 433. doi:10.1186/s13071-020-04306-9

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, 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

802 Olsson, P. A., Korhonen, L., Mercer, E. A., & Lindholm, D. (1999). MIR is a novel ERM-like protein
803   that interacts with myosin regulatory light chain and inhibits neurite outgrowth. *Journal of*
804   *Biological Chemistry, 274*(51), 36288-36292. doi:10.1074/jbc.274.51.36288
805 Ozerov, M. Y., Ahmad, F., Gross, R., Pukk, L., Kahar, S., Kisand, V., & Vasemagi, A. (2018). Highly
806   Continuous Genome Assembly of Eurasian Perch (Perca fluviatilis) Using Linked-Read
807   Sequencing. G3-Genes Genomes Genetics, 8(12), 3737-3743. doi:10.1534/g3.118.200768
808 [dataset] Ozerov, M. Y., Noreikiene, K., Kahar, S., Huss, M., Huusko, Kõiv, T., . . . , Vasemagi, A.
809   (2021). Genomics of humic adaptation in Eurasian perch (Perca fluviatilis), SNP genotypes of
810   32 perch individuals; Dryad; DOI: XXXX/XXXX
811 [dataset] Ozerov, M. Y., Noreikiene, K., Kahar, S., Huss, M., Huusko, Kõiv, T., . . . , Vasemagi, A.
812   (2021). Genomics of humic adaptation in Eurasian perch (Perca fluviatilis); raw DNA-seq
813   reads; NCBI; BioProject: PRJNA760273
814 Pallone, T. L., Khurana, S., & Cao, C. (2012). Voltage-Gated Calcium Channels: Structure and
815   Function (CACNA). In S. Choi (Ed.), *Encyclopedia of Signaling Molecules* (pp. 1984-1992).
816   New York, NY: Springer New York.
817 Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary
818   analyses in R. *Bioinformatics, 35*(3), 526-528. doi:10.1093/bioinformatics/bty633
819 Parra, J. E. G., & Baldisserotto, B. (2007). Effect of water pH and hardness on survival and growth of
820   freshwater teleosts. In B. Baldisserotto (Ed.), *Fish Osmoregulation*: CRC Press.
821 Peirce, M. J., Brook, M., Morrice, N., Snelgrove, R., Begum, S., Lanfrancotti, A., . . . Wait, R. (2010).
822   Themis2/ICB1 Is a Signaling Scaffold That Selectively Regulates Macrophage Toll-Like
823   Receptor Signaling and Cytokine Production. *PLoS One, 5*(7), e11465.
824   doi:10.1371/journal.pone.0011465
825 Pendeville, H., Peers, B., Kas, K., & Voz, M. L. (2006). Cloning and embryonic expression of
826   zebrafish PLAG genes. *Gene Expression Patterns, 6*(3), 267-276.
827   doi:https://doi.org/10.1016/j.modgep.2005.08.001
828 Pfenninger, M., Lerp, H., Tobler, M., Passow, C., Kelley, J. L., Funke, E., . . . Plath, M. (2014).
829   Parallel evolution of cox genes in H2S-tolerant fish as key adaptation to a toxic environment.
830   *Nature Communications, 5*(1), 3873. doi:10.1038/ncomms4873
831 Pizzagalli, M. D., Bensimon, A., & Superti-Furga, G. (2021). A guide to plasma membrane solute
832   carrier proteins. *The FEBS Journal, 288*(9), 2784-2835. doi:https://doi.org/10.1111/febs.15531
833 Posa, M. R. C., Wijedasa, L. S., & Corlett, R. T. (2011). Biodiversity and Conservation of Tropical
834   Peat Swamp Forests. *Bioscience, 61*(1), 49-57. doi:10.1525/bio.2011.61.1.10
835 Prevosti, A., Ocaña, J., & Alonso, G. (1975). Distances between populations ofDrosophila subobscura,
836   based on chromosome arrangement frequencies. *Theoretical and Applied Genetics, 45*(6),
837   231-241. doi:10.1007/BF00831894
838 Rask, M. (1984). The effect of low pH on perch, Perca fluviatilis L. II. The effect of acid stress on
839   different development stages of perch. *Annales Zoologici Fennici, 21*(1), 9-13. Retrieved from
840   http://www.jstor.org/stable/23734093
841 Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., . . .
842   Whitehead, A. (2016). The genomic landscape of rapid repeated evolutionary adaptation to
843   toxic pollution in wild fish. *Science, 354*(6317), 1305-1308. doi:10.1126/science.aah4993
844 Riesch, R., Plath, M., Schlupp, I., Tobler, M., & Langerhans, B. (2014). Colonisation of toxic
845   environments drives predictable life-history evolution in livebearing fishes (Poeciliidae).
846   *Ecology Letters, 17*(1), 65-71. doi:https://doi.org/10.1111/ele.12209
847 Riesch, R., Tobler, M., & Plath, M. (2015). *Extremophile Fishes*.
848 Roulet, N., & Moore, T. R. (2006). Browning the waters. *Nature, 444*(7117), 283-284.
849   doi:10.1038/444283a
850 Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing
851   phylogenetic trees. *Molecular Biology and Evolution, 4*.
852 Sepp, M., Kõiv, T., Nõges, P., & Nõges, T. (2019). The role of catchment soils and land cover on
853   dissolved organic matter (DOM) properties in temperate lakes. *Journal of Hydrology, 570*,
854   281-291. doi: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. doi: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. doi: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. doi:10.1579/0044-7447

Solomon, C. T., Jones, S. E., Weidel, B. C., Buffam, I., Fork, M. L., Karlsson, J., . . . 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. doi: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-e1474. doi: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). doi:10.3389/fcell.2020.603508

Tagliarolo, M. (2019). Acidification in Aquatic Systems. In B. Fath (Ed.), *Encyclopedia of Ecology (Second Edition)* (pp. 6-13). Oxford: Elsevier.

Thomas, P. D., Kejariwal, 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. doi: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). Chicago: 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. doi:10.1016/j.ijbiomac.2020.01.207

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. doi: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. doi: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. doi:10.7554/eLife.43785

Vettori, A., Bergamin, G., Moro, E., Vazza, G., Polo, G., Tiso, N., . . . Mostacciuolo, 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. *Neuromuscul Disord, 21*(1), 58-67. doi: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. doi:10.1016/j.scitotenv.2006.02.038

Wetzel, R. G. (2001). *Limnology. Lake and River Ecosystems. 3rd Edition*. San Diego: Academic Press.

Weyhenmeyer, G. A., Hartmann, J., Hessen, D. O., Kopáček, J., Hejzlar, J., Jacquet, S., . . . Zechmeister, T. (2019). Widespread diminishing anthropogenic effects on calcium in freshwaters. *Scientific Reports, 9*(1), 10450. doi:10.1038/s41598-019-46838-w

908   Winter, D. J. (2017). rentrez: An R package for the NCBI eUtils API. *The R Journal, 9*(2), 520-526.

909   Wood, C. M., Al-Reasi, H. A., & Smith, D. S. (2011). The two faces of DOC. *Aquatic Toxicology,*
910   *105*(3, Supplement), 3-8. doi:https://doi.org/10.1016/j.aquatox.2011.03.007

911   Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using
912   exact alignments. *Genome Biology, 15*(3), R46. doi:10.1186/gb-2014-15-3-r46

913   Xiong, P., Hulsey, C. D., Meyer, A., & Franchini, P. (2018). Evolutionary divergence of 3' UTRs in
914   cichlid fishes. *BMC Genomics, 19*(1), 433. doi:10.1186/s12864-018-4821-8

915   Xu, J., Li, J. T., Jiang, Y., Peng, W., Yao, Z., Chen, B., . . . Xu, P. (2017). Genomic Basis of Adaptive
916   Evolution: The Survival of Amur Ide (Leuciscus waleckii) in an Extremely Alkaline
917   Environment. *Molecular Biology and Evolution, 34*(1), 145-159. doi:10.1093/molbev/msw230

918   Xuereb, A., Kimber, C. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M. J. (2018). Putatively
919   adaptive genetic variation in the giant California sea cucumber (Parastichopus californicus) as
920   revealed by environmental association analysis of restriction-site associated DNA sequencing
921   data. *Molecular Ecology, 27*(24), 5035-5048. doi:10.1111/mec.14942

922   Yu, F. H., & Catterall, W. A. (2003). Overview of the voltage-gated sodium channel family. *Genome*
923   *Biology, 4*(3), 207. doi:10.1186/gb-2003-4-3-207

924 **DATA ACCESSIBILITY**

925   Raw sequence reads were deposited in the SRA (BioProject PRJNA760273). Individual genotype

926   data are available on DataDryad (DOI:XXXX/XXXX).

927

928 **AUTHOR CONTRIBUTIONS**

929   MO analyzed the data and wrote the first draft of the manuscript together with AV. ML analyzed

930   the data. TK and MS performed laboratory analyses of water samples. SK, AH, MH, AG, KN and AV

931   were involved in fieldwork. AV and RG conceived the study. All the authors contributed to revisions of

932   the draft manuscript. All the authors read and approved the final manuscript.

TABLES

934

935 **Table 1.** Enrichment and depletion of candidate SNPs in each annotation category.

936

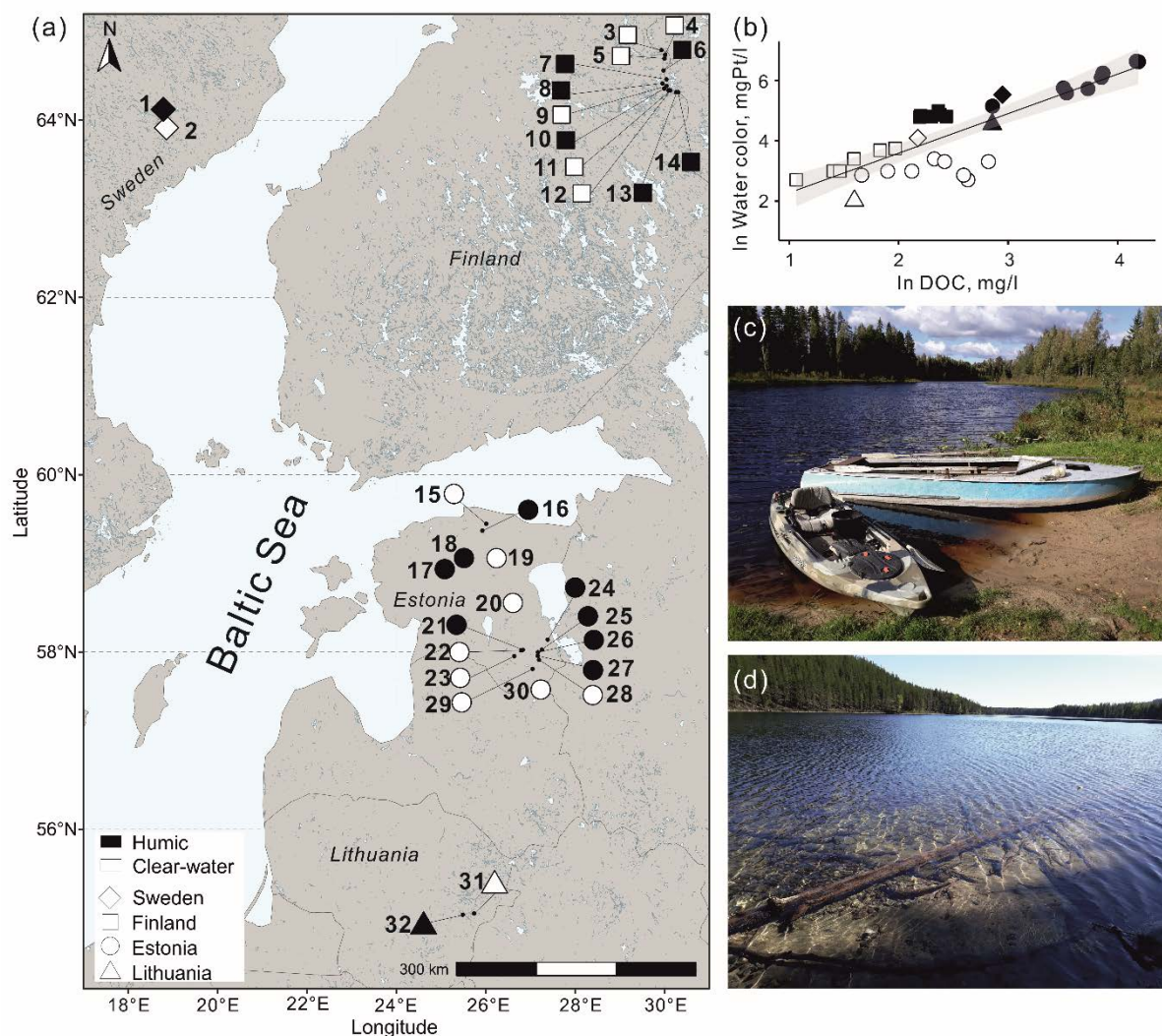| Variant	All SNPs	Candidate SNPs	χ^2	*P*	All SNPs, %	Candidate SNPs, %
5K upstream	132087	1675	0.056	0.813	12.22	12.14
5'UTR	8040	127	5.415	**0.020**	0.74	0.92
5'UTR premature start codon gain	1386	13	0.979	0.322	0.13	0.09
Synonymous	25046	305	0.607	0.436	2.32	2.21
splice region & synonymous	580	8	0.001	0.974	0.05	0.06
non-synonymous	18801	237	0.024	0.876	1.74	1.72
non-synonymous & splice region	445	3	0.826	0.363	0.04	0.02
Intron	358885	4668	1.207	0.272	33.20	33.83
splice region & intron	3564	42	0.192	0.661	0.33	0.30
splice region	154	2	0.000	1.000	0.01	0.01
3'UTR	33552	492	8.883	**0.003**	3.10	3.57
5K downstream	130029	1775	6.937	**0.008**	12.03	12.86
intergenic	368105	4450	9.759	**0.002**	34.05	32.25
stop gained	145	1	0.064	0.801	0.01	0.01

937

FIGURES

Fig. 1. (a) Map indicating sampling locations. (b) Dot plot showing the relationships between DOC and water color values (ln-transformed) in each lake. Regression is shown as 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 legend. Population ID numbers follow those given in Table S1.
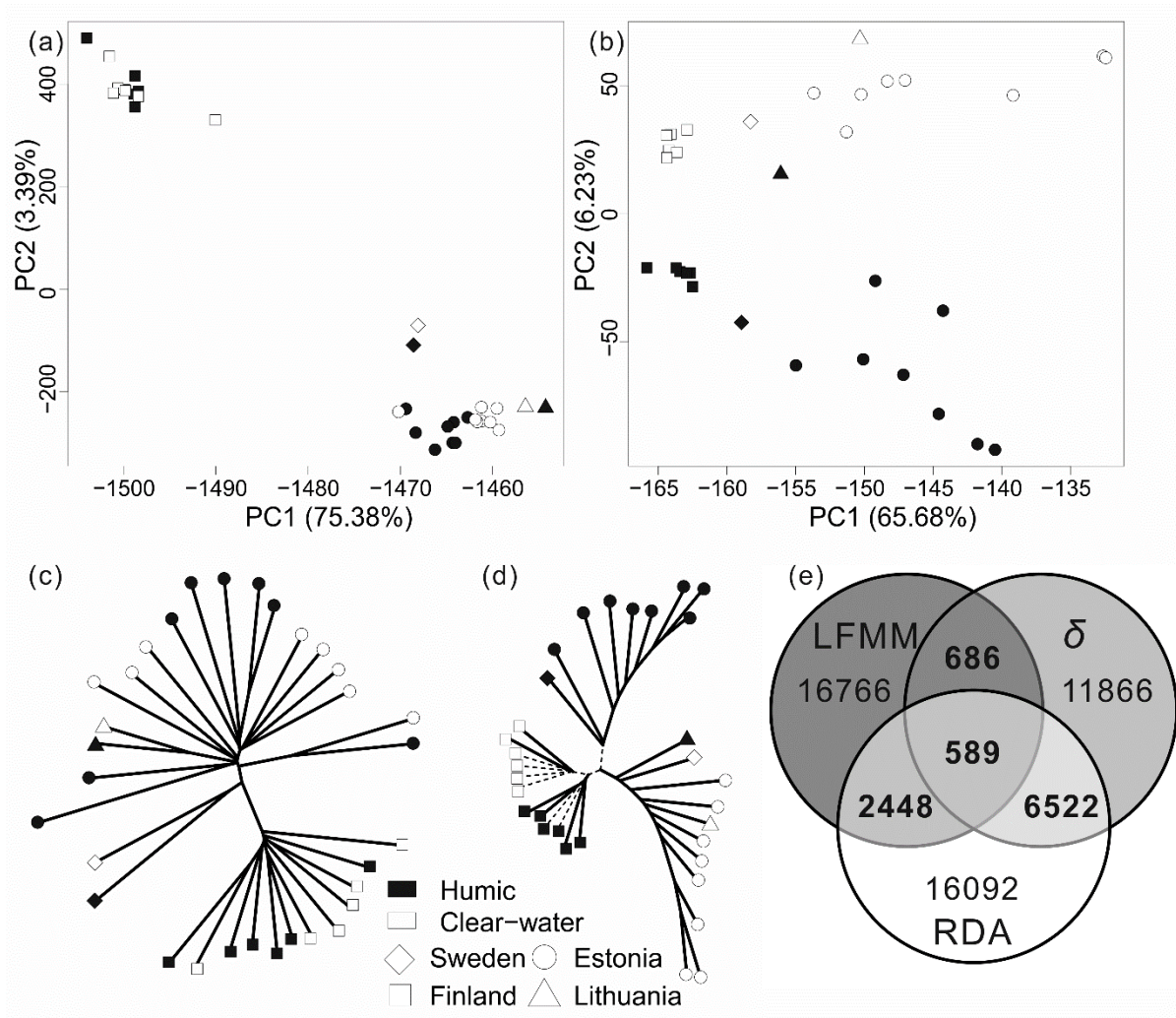
948
949
950 **Fig. 2.** PCA summarizing the genetic structure for (a) 810,591 genome-wide SNPs and for (b) 10,245
951 candidate SNPs. Neighbor-joining dendrogram based on D_{Nei} genetic distances, demonstrating the
952 genetic relationships among perch samples based on (c) 810,591 genome-wide SNPs and (d) 10,245
953 candidate SNPs. The branches with bootstrap value support < 80% are represented as dashed lines.
954 Humic and clear-water populations are indicated as black- and white-filled symbols, respectively. The
955 country of origin is depicted by different shapes as shown in the legend. (e) Venn diagram showing
956 overlap among candidate SNPs revealed by three methods; the final set of candidate SNPs is highlighted
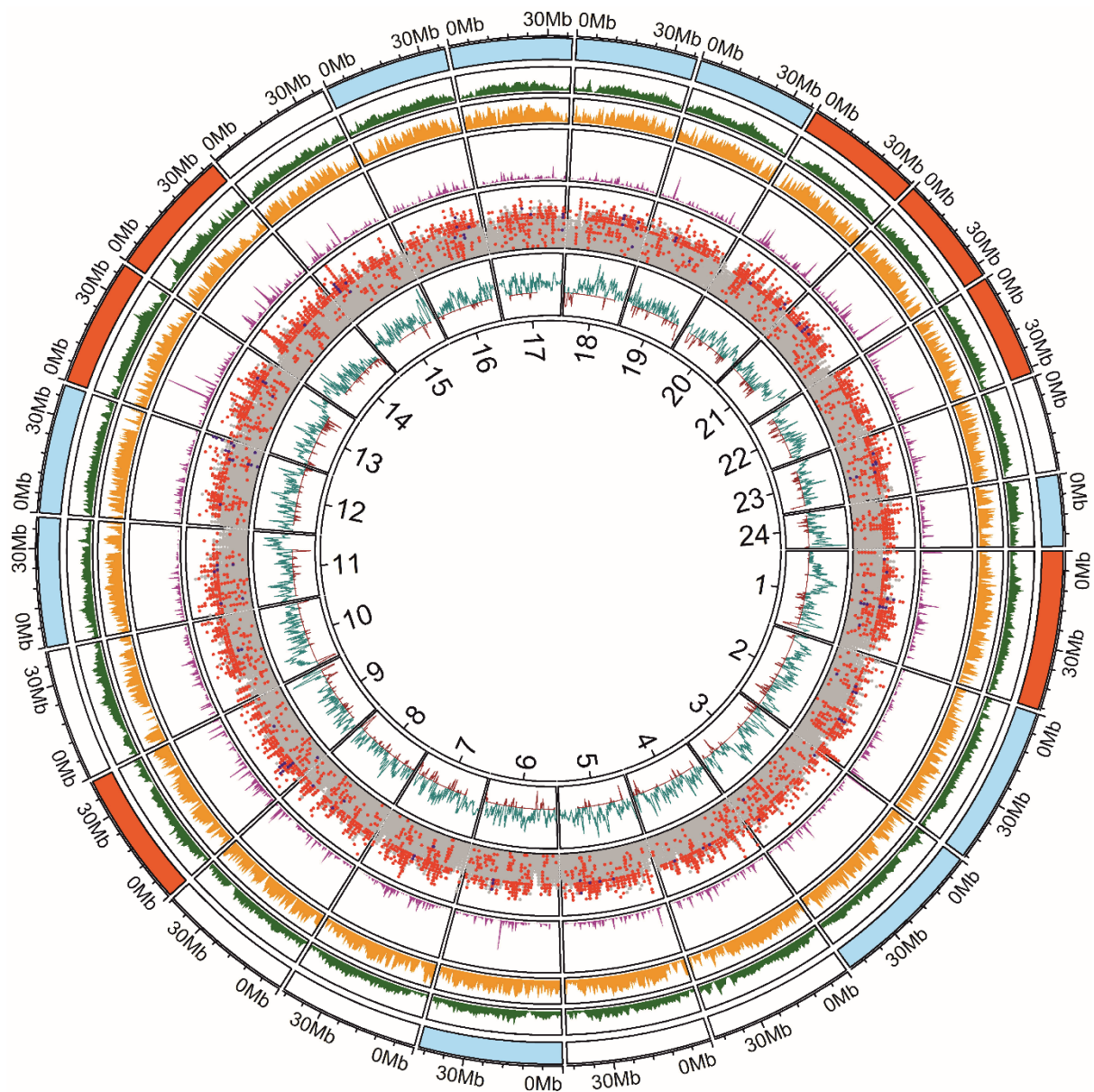957 in bold.

Fig. 3. Circos plot showing the distribution of SNPs, genes, genomic divergence and diversity in humic and clear-water perch (from 32 lakes) across 24 chromosomes. Chromosomes with a significant enrichment or depletion of candidate SNPs are highlighted in red and light blue, respectively. The circles from outer to inner show all SNP density (green), gene density (orange), candidate SNP density (magenta), genetic divergence (δ) and difference of genetic diversity (H_O) between clear-water and humic perch. Genomic densities were calculated using window sizes of 0.5 Mb. Candidate SNPs, SNPs related to calmodulin binding genes and other SNPs on the δ plots are shown as red, blue and grey dots, respectively. An excess or deficiency of H_O in humic perch is shown as brown and light blue lines, respectively. The observed heterozygosity difference between humic and clear-water perch was estimated as the difference between moving averages of H_O across 500 SNPs.
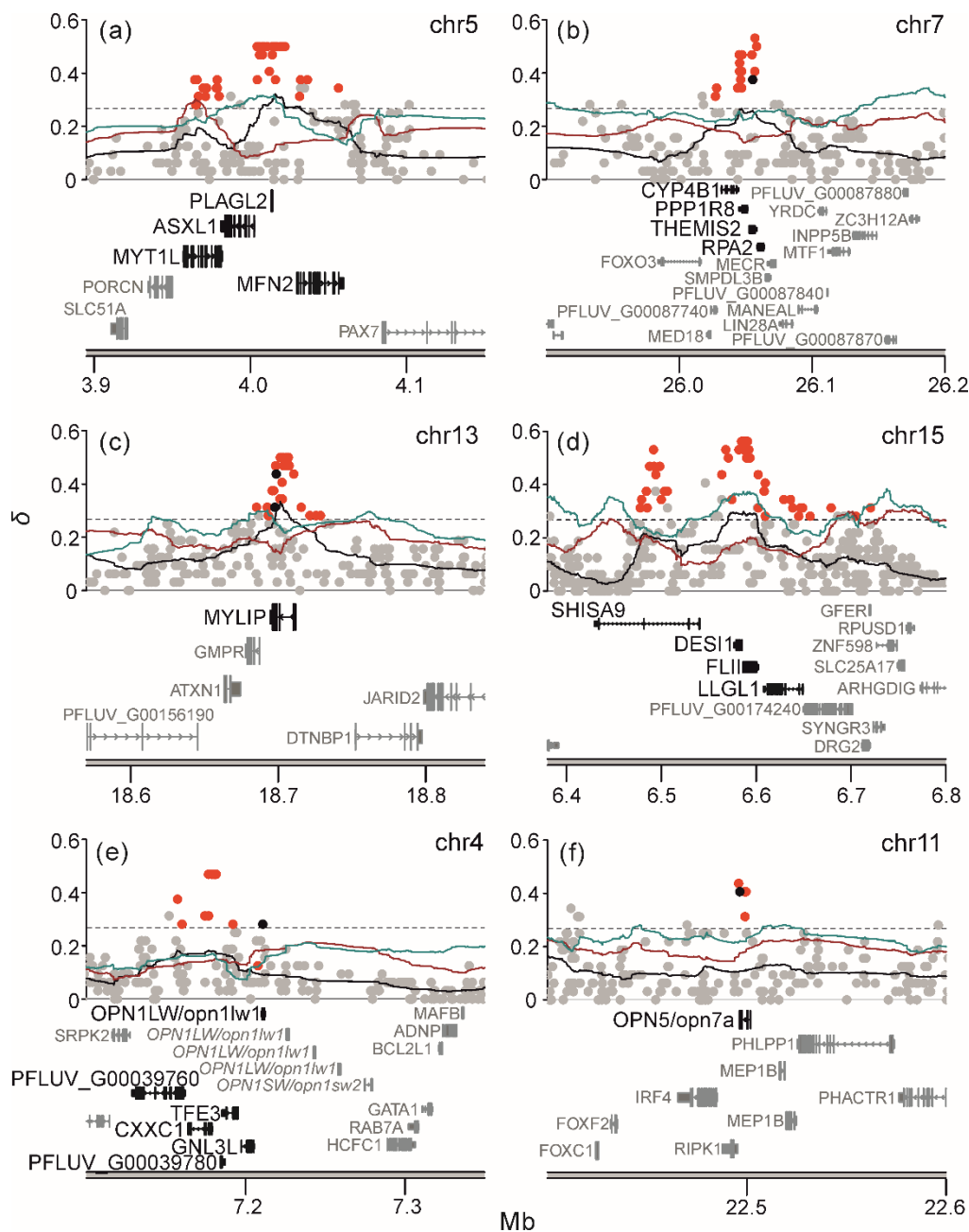
Fig. 4. Examples of (a-d) genomic regions of high genetic divergence (δ) between humic and clear-water perch, and (e, f) opsins with non-synonymous mutations. Candidate SNPs, non-synonymous candidate SNPs and other SNPs are shown as red, black and grey dots, respectively. Moving average of δ and H_O of humic and clear-water perch across 50 SNPs are shown as 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 non-synonymous substitutions are highlighted in black, while other opsin genes at chromosome 4 are highlighted in italic. Dashed line indicates δ threshold = 0.268, which corresponds to 2.5 SD of mean δ.
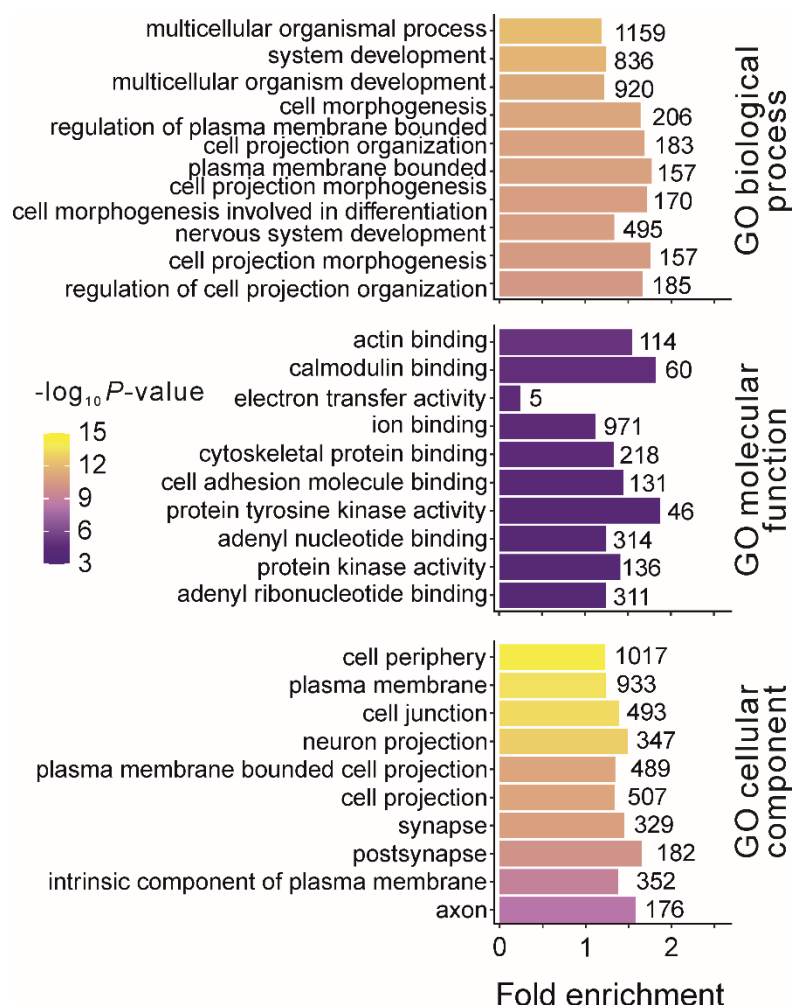
Fig. 5. Top ten significantly enriched gene ontology (GO) terms (FDR ≤ 0.05) among the 3,245 candidate genes for humic adaptation in perch. Bar length and numbers on the right represent the fold enrichment and number of enriched genes for each GO term, respectively.

SUPPLEMENTAL INFORMATION

992

993 **Supplementary file 1.** Genome-wide SNP annotation information, LFMM, RDA and δ scores, and
994 frequency of alternative alleles among humic and clear-water perch.

995

996 **Fig. S1.** Box-plots showing the level of dissolved organic matter content (DOC) and water color (ln-
997 transformed) and lake size (area and shoreline distance, ln-transformed) between humic and clear-water
998 lakes (*P*-values of non-parametric Mann-Whitney test are presented). Horizontal line, rectangle, and
999 whiskers indicate the median, 25th and 75th quartiles, and the non-outlier range, respectively.

1000

1001 **Fig. S2.** (a) Box-plot showing the level of observed heterozygosity (H_O) between perch samples from
1002 humic and clear-water lakes (non-parametric Mann-Whitney test $P = 0.11$). Horizontal line, rectangle,
1003 and whiskers indicate the median, 25th and 75th quartiles, and the non-outlier range, respectively. Dot
1004 plots showing the relationships between observed heterozygosity and (b) DOC, (c) water color, (d) area,
1005 and (e) shoreline length estimates (ln-transformed) in each lake. Humic and clear-water populations are
1006 indicated as black- and white-filled points, respectively. The country of origin is depicted by different
1007 point shapes as shown in the legend.

1008

1009 **Fig. S3.** PCA summarizing the genetic structure for 256,880 putatively neutral intergenic SNPs. Humic
1010 and clear-water populations are indicated as black- and white-filled points, respectively. The country of
1011 origin is depicted by different point shapes as shown in the legend.

1012

1013 **Fig. S4.** (a-d) Genetic divergence (δ) between humic and clear-water perch in the genomic regions
1014 containing visual opsin genes. Candidate SNPs, non-synonymous candidate SNPs and other SNPs are
1015 shown as red, black and grey dots, respectively. Moving average of δ and H_O of humic and clear-water
1016 perch across 50 SNPs are shown as black, brown and cyan solid lines, respectively. Gene symbols are
1017 presented as human (and zebrafish for opsins) orthologues and perch GenBank gene IDs (PFLUV_G)
1018 for genes with unidentified functions. Opsin genes are highlighted in black. Dashed line shows δ
1019 threshold = 0.268, which corresponds to 2.5 SD of mean δ.

1020

1021 **Table S1.** Sample index; ID; lake name; location (country, longitude and latitude); ecotype (humic vs
1022 clear-water); date of sampling; sex; fork/total length; body mass; sampling method; lake size: surface
1023 area and shoreline length; values reflecting the amount of dissolved organic matter in the lake water:
1024 DOC (dissolved organic carbon) and water color; observed hetetozygosity (H_O).

1025

1026 **Table S2.** Overall genome-wide genetic divergence among perch samples measured as Nei's genetic
1027 distances (Nei 1972; above diagonal) and pairwise mean absolute allele frequency differences (Prevosti
1028 et al., 1975; below diagonal).

1029

1030 **Table S3.** Enrichment and depletion of candidate SNPs at each chromosome.

1031

1032 **Table S4.** Number of genomic regions with a high density of candidate SNPs.

1033

1034 **Table S5.** Number of SNPs per candidate gene (Gene ID) and their human (HS gene) and zebrafish (DR
1035 gene) orthologues. The strongest candidates (mean $δ ≥ 0.40$ and number of SNPs per gene $≥ 6$) are
1036 highlighted in bold.

1037

1038 **Table S6.** List of gene ontology (GO) terms with significant (FDR $≤ 0.05$) enrichment among the
1039 candidate genes compared to all annotated genes in the perch genome (N_{ref}: number of genes in the
1040 reference list associated with a specific GO term; n_{test}: number of genes in the test list associated with a
1041 specific GO term; expected: number of genes expected in the test list for this GO term, based on the
1042 reference list).