1   **Rapid niche expansion by selection on functional genomic variation after**

2   **ecosystem recovery**

3

4

5

6   Arne Jacobs[1†], Madeleine Carruthers[1†], Reiner Eckmann[2], Elizabeth Yohannes[2], Colin E.

7   Adams[1,3], Jasminca Behrmann-Godel[2*], & Kathryn R. Elmer[1*]

8

9   **Affiliation:**

10   [1]Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical,

11   Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8QQ, Scotland, UK

12   [2]Limnological Institute, University of Konstanz, Konstanz, 78457 Germany

13   [3]Scottish Centre for Ecology and the Natural Environment, University of Glasgow,

14   Rowardennan, G63 0AW, Scotland, UK

15   *Corresponding

16   † These authors contributed equally to this work

17

18

19 **Introduction.**

20 It is well recognised that environmental degradation caused by human activities can result in

21 dramatic losses of species and diversity. However, comparatively little is known about the

22 ability of biodiversity to re-emerge following ecosystem recovery. Here we show that a

23 European whitefish subspecies, the gangfisch (*Coregonus lavaretus macrophthalamus*),

24 rapidly increased its ecologically functional diversity following the restoration of Lake

25 Constance after anthropogenic eutrophication. In fewer than 10 generations, gangfisch

26 evolved a greater range of gill raker number to utilise a broader ecological niche. A sparse

27 genetic architecture underlies this variation in gill raker number. Several co-expressed gene

28 modules and genes showing signals of positive selection were associated with gill raker

29 number and body shape. These were enriched for biological pathways related to trophic niche

30 expansion in fishes. Our findings demonstrate the potential of functional diversity to expand

31 following habitat restoration, given a fortuitous combination of genetic architecture, genetic

32 diversity, and selection.

33

34   **Main text.**

35

36   Anthropogenic activities are one of the major drivers of environmental change, often with

37   critical impacts on ecosystem health[1,2]. Environmental degradation caused by human

38   disturbance can be extremely challenging for organisms, resulting in rapid evolutionary and

39   ecosystem changes[3], interbreeding and hybridisation, extirpation, or even extinction[2,4–6]. For

40   example, hybridisation across closely related species can result in a loss of species richness

41   ('speciation reversal') and associated functional diversity in adaptive traits[4,5,7]. In some cases,

42   polluted environments can be improved by remediation efforts, but the potential for diversity

43   to re-emerge following ecosystem restoration remains a major unanswered issue in

44   evolutionary and conservation biology, with important implications for policy and practical

45   efforts[8,9]. Theoretical models have predicted that diversity can recover but only under

46   particular genetic, ecological and evolutionary circumstances[10]. These models suggested that

47   the probability of functional diversity to re-emerge from a collapsed species group (e.g.

48   'hybrid swarm') is higher when environmental disturbance is temporary (as longer lasting

49   disturbances, even weak ones, risk the loss of adaptive genetic diversity), and when

50   functional trait diversity has a simple genetic basis and is linked to assortative mating or

51   mating-related traits[10,11]. For example, it has been shown that fish often mate with individuals

52   that are similar[12,13]; thus, if functional traits are linked to e.g. body shape, divergence in a

53   functional trait can lead to morphological divergence and further on to genetic divergence via

54   trait-based assortative mating[10]. However, re-emergence of functional diversity following a

55   species collapse has not been observed nor empirically tested to date.

56   In many aquatic ecosystems, high levels of nutrient pollution from agriculture and

57   sewage during the mid-20th century caused cultural eutrophication. The algae blooms and

58   consequent negative impacts on water quality caused the collapse of natural trophic and

59   habitat gradients in many freshwater lakes. In the Alpine lakes, this resulted in the dramatic

60   loss of whitefish ecological and functional diversity, highlighted by the average reduction in

61   the range of gill raker numbers by 14% across lakes[4]. For example, in one of the largest

62   lakes, Lake Constance, the benthic specialists (kilch and sandfelchen) went extinct or

63   collapsed, with extensive hybridisation among those and the remaining pelagic (blaufelchen

64   and gangfisch) and littoral-benthic (weissfelchen) subspecies[4]. This depleted functional

65   diversity, evidenced as a 28% reduction in the range of gill raker number across whitefish

66   species[4,14] (Fig. 1a, Supplementary Table 1). As a result, individuals with low gill raker

67   numbers disappeared from Lake Constance during and after eutrophication[4,15]. However,

68   concerted effort from the 1980s to reduce polluting nutrient inputs enabled Lake Constance to

69   revert rapidly to its pristine oligotrophic state[16] (Fig. 1a). This provides an excellent system to

70   study biodiversity recovery after ecosystem remediation, as long-term data are available for

71   the fish and lake conditions, natural recolonisation from other whitefish populations is

72   unlikely, and, in contrast to other alpine lakes, Constance has not been subjected to external

73   stocking[15,17]. Here, we explored the variation in ecological, morphological, genomic, and

74   gene expression traits to assess the material underpinning a rapid re-emergence of functional

75   diversity in whitefish.

76

77   **Results and Discussion.**

78

79   **Functional and eco-morphological divergence.** To examine contemporary phenotypic

80   diversity and its relationship to habitat use and trophic niche in whitefish, we focused on the

81   gangfisch because it is the only subspecies that has modified its spawning depth since

82   ecosystem recovery[12]. Research proposed that gangfisch may be ecologically expanding

83   because spawning depth is strongly associated with trophic ecology in European

4

84 whitefish[15,18]. Thus, we assessed gill rakers and body shape across spawning depths (shallow:

85 2-5m, intermediate: 20-25m, and deep: below 50m) at multiple sites in the lake

86 (Supplementary Figs. 1 and 2a). Gill rakers are important trophic traits in many postglacial

87 fish species; many, dense and long rakers are used to filter-feed zooplankton while few,

88 sparse and short rakers are associated with benthos feeding, and gill raker number is well

89 documented to be genetically based[4,19,20]. Indeed, in line with expectations we found gill

90 raker number (GRN) to be significantly structured by depth (Fig. 1b; $F_{2,260}=17.27$, $P<0.001$),

91 with shallow-spawning individuals having fewer gill rakers ($GRN_{shallow}=33.9 \pm 4.7$ standard

92 error of the mean (SEM)) than deep-spawning individuals ($GRN_{deep}=37.0 \pm 2.8$ SEM). In

93 addition, we detected a significant effect of depth on gill raker length (Supplementary Fig. 2;

94 GRL: $F_{2,265}=4.824$, $P=0.0088$), and number and length were positively correlated, with

95 individuals with fewer gill rakers also having shorter gill rakers (Supplementary Fig. 3a;

96 GRL: $R^2=0.028$; $F_{1,267}=8.76$, $P=0.0034$).

97   Individuals spawning at different depths also differed in body shape (MANCOVA:

98 Pillai's $Trace_{Depth}=0.722$, df$=(1,66)$, $P<0.001$), an important eco-morphological trait

99 (Supplementary Fig. 2); shallow-spawners were deeper bodied and had a sub-terminal mouth

100 compared to deep-spawning individuals, which had more slender bodies and terminal mouths

101 (Fig. 1d; Supplementary Fig. 2b,c). Body shape divergence with depth was mostly explained

102 by PC2 (Supplementary Fig. 2c) and variation in body shape was correlated to GRN (Fig. 1f;

103 BS-PC2: $R^2=0.056$; $F_{1,266}=16.76$, $P<0.001$; Supplementary Fig. 3b,c), suggesting its

104 association with trophic ecology in gangfisch. Additionally, shallow-spawning gangfisch

105 were longer (standard length $SL_{shallow}=28.6 \pm 0.7$ cm vs $SL_{deep}=25.7 \pm 0.6$ cm) and heavier

106 (body $weight_{shallow}=327.5 \pm 12.9$ g vs body $weight_{deep}=225.2 \pm 5.6$ g) than deep-spawning

107 individuals (SL: $F_{2,261}=34.69$, $P<0.001$; body weight: $F_{2,261}=38.49$, $P<0.001$) (Supplementary

108 Fig. 2d), evidencing that multiple phenotypic traits contribute to the gradient along the depth

109 axis. Although divergence in body shape and standard length were correlated with GRN (Fig.

110 1, Supplementary Fig. 3), body length was only (negatively) correlated with GRN from age

111 three onwards (back-calculated from scales; Supplementary Fig. 3e), indicating increased

112 growth of putatively benthic whitefish later in ontogeny. However, the observed patterns

113 were heterogeneous across sampling sites, as indicated by the effect of site and depth by site

114 interaction on variation in body shape (MANCOVA: Pillai's Trace$_{site}$=1.243, df=(2,132),

115 P<0.001; Pillai's Trace$_{site*depth}$=1.054, df=(2,132), P<0.001). Individuals from Immenstaad

116 showed in general less divergence along the depth gradient (Supplementary results). In

117 comparison to the other extant whitefish subspecies, gangfisch were intermediate in body

118 shape to weissfelchen and blaufelchen, with shallow-spawning gangfisch being more similar

119 to benthic-littoral weissfelchen and deep-spawning gangfisch being similar to pelagic

120 blaufelchen (Fig. 1d).

121 To test if phenotypic diversity is correlated with trophic and habitat diversity in

122 gangfisch, and therefore reflects functional utility, we assessed the correlation of gill raker

123 number with ecological indicators: stable isotope signature and parasite infection. We found

124 that the diversity associated with spawning depth was related to trophic and ecological

125 differences (Fig.1; Supplementary Figs. 2,3). Specifically, gill raker number was correlated to

126 carbon stable isotope signatures (Fig. 1e; R^2=0.07751, P<0.001, df=147), indicating a more

127 littoral-benthic diet in individuals with fewer gill rakers[21]. Fewer gill rakers were also

128 associated with higher infection by *Diplostomum* eye flukes (Fig. 1c; Wilcoxon rank sum

129 test, P<0.001)*,* trematode parasites that infect snails in the shallow zone of the lake. Fish-

130 infective stages hatch from the snails and therefore indicate littoral-benthic habitat use[22].

131 Overall, these observed differences in morphology and functional traits with spawning depth

132 are typical of benthic-pelagic divergence in European whitefish trophic specialists along the

133   depth gradient[19,23]. Thus, we focused on variation in the key functional trait of gill raker

134   number as a proxy of overall trophic and eco-morphological diversity.

135

136   **Rapid trait expansion.** Identifying how functional diversity changes in response to rapid

137   environmental change is important for understanding its drivers. We compared gill raker

138   number (GRN) between the samples collected in 2014 for this study and samples collected in

139   1990[24]. This demonstrated that the functional trait diversity arose rapidly, as it was not

140   present in gangfisch at the end of the eutrophic era in 1990 (Fig. 1a, 2a; range in GRN:

141   1990=30-42[24], 2014=24-44). The range in GRN increased drastically over this 24-year period

142   (corresponding to about eight whitefish generations), with a notable decrease in the minimum

143   number from 30 to 24. The mean number of gill rakers decreased by 0.3423 standard

144   deviations (SD; haldane numerator) over this period, equivalent to an evolutionary rate of -

145   0.0428 haldanes (Fig. 2)[25]. This is comparable to rates of phenotypic evolution in several

146   examples of rapid evolution, such as change in gill raker number in marine-freshwater

147   sticklebacks or hindlimb length in anole lizards, and lies within the top 12% of reported rates

148   of rapid phenotypic evolution[6] (Fig. 2b). The shift in functional phenotypic diversity

149   coincides temporally with the re-oligotrophication of Lake Constance. Thus, we propose that

150   restored habitat gradients drove the emergence of functional diversity (GRN) in gangfisch by

151   selection. This rapid expansion in gangfisch partially, but not completely, recovers the

152   historical functional diversity that existed across the whitefish species complex in Lake

153   Constance (pre-eutrophication GRN range across species: 17-43)[14,15] (Fig. 1a, Supplementary

154   Table S1).

155

156   **Evolutionary history of introgression.** To examine the evolutionary genetic changes

157   underlying the observed rapid diversification, we analysed 13,589 genome-wide single

158   nucleotide polymorphisms (SNPs) within and across whitefish subspecies in Lake Constance.

159   We found two genetic clusters, with blaufelchen and weissfelchen at extremes, and gangfisch

160   showing an intermediate gradient of ancestry (Figs. 3b; Supplementary Fig. 4a,b), suggesting

161   strong historic admixture and recent segregation in gangfisch[4]. Indeed, we detected multiple

162   signals consistent with historical genetic introgression from blaufelchen and weissfelchen

163   into gangfisch, including the intermediacy of gangfisch in genetic ordination plots

164   (Supplementary Fig. 4a-d), the divergence of gangfisch along the eigenvector distinguishing

165   subspecies (Supplementary Fig. 4c,d), and support from introgression tests (Fig. 3a, *f3*

166   statistics=-0.00108362, z-score=-7.967).

167   To reconstruct the history of divergence and introgression between all extant

168   subspecies, we used a demographic modelling approach based on the joint site frequency

169   spectrum, built from the full SNP dataset and a dataset excluding differentiated SNPs (i.e.

170   putatively neutral, Supplementary Fig. 4b). Based on the full dataset, we found the most

171   likely demographic history to be one of strong and recent introgression from blaufelchen

172   (mean admixture proportion=0.78) and weissfelchen (mean admixture proportion=0.45) into

173   gangfisch (Fig. 3b, Supplementary Table 3), leading to the formation of a gangfisch hybrid

174   swarm[4]. Demographic analysis based on the neutral SNP dataset suggested a simple

175   isolation-with-migration model to be more likely, although models including introgression

176   were also highly supported (Supplementary Table 4). The weaker signal of introgression in

177   the neutral SNP dataset suggests that loci which introgressed into gangfisch from other

178   sympatric whitefish species are associated with its diversification. Thus, introgressive

179   hybridisation across species in Lake Constance, including those that are now extinct[4],

180   potentially provided the genetic material needed for the rapid eco-morphological

181   diversification in gangfisch after ecosystem recovery[26].

182

183 **Genetic basis of functional and eco-morphological traits.** To assess the genetic

184 underpinnings of eco-morphological changes, we first tested the correlation of functional

185 diversity and genetic ancestry. Gill raker number (GRN) in gangfisch was highly correlated

186 with genetic ancestry (Fig. 4a; Spearman test, r_{GRN}=-0.6007, P<0.001), suggesting genetic

187 isolation by ecological adaptation[27]. Sparsely-rakered gangfisch ('benthic gangfisch') were in

188 general genetically more similar to the benthic-littoral weissfelchen, whereas densely-rakered

189 gangfisch ('pelagic gangfisch') were genetically closer to pelagic blaufelchen (Fig. 4a,

190 Supplementary Fig. 4a-d).

191   To further assess the genetic basis of functional diversity (variation in GRN) in

192 gangfisch, we investigated its genetic architecture. Polygenic models suggested that GRN is

193 controlled by a sparse genetic architecture, with most of the phenotypic variance (phenotypic

194 variance explained by all SNPs: PVE_{GRN}=0.9991) being explained by a small expected

195 number (predicted to be between 68 and 117 (95% CI)) of large effect loci (phenotypic

196 variance explained by sparse effect loci: PGE_{GRN}=0.9718) (Supplementary Table 5). In

197 comparison, we found that body shape, a highly polygenic trait in fishes[28] that was also

198 correlated with genetic ancestry (rho_{BS-PC2}=0.334, P<0.001), had a more polygenic

199 architecture, with only a small proportion of phenotypic variance explained by large effect

200 loci (Supplementary Table 5; PVE_{BS-PC2}=0.8636; PGE_{BS-PC2}=0.2296). These results suggest

201 differences in the underlying architecture of these traits, with the large proportion of variance

202 explained (PVE) likely reflecting the high heritability of these traits and the sparse genetic

203 architecture in the case of GRN (PGE_{GRN}>0.9), with similarly high values reported for beak

204 size in Darwin's finches (PVE=0.947, PGE=0.775)[29] or dorsal brightness in deer mice

205 (PVE=0.69, PGE=0.83)[30]. Statistical difficulties involved in estimating trait architectures or

206 correlations with population structure could inflate these values. However, for this study the

207 large difference in PGE for GRN compared to body shape (97% vs 23%) is particularly

208 informative and highlights the different architectures underlying these traits. Such a simple

209 genetic architecture may facilitate the rapid expansion in gill raker number[11].

210 Next, we used genome-wide approaches to identify the loci underlying phenotypic

211 variation. Association analyses using linear-mixed models (*GEMMA*) identified six SNPs

212 associated with variation in GRN, whereas a latent-factor linear mixed model (*LFMM*)

213 identified a set of 99 associated SNPs (Fig. 4; Supplementary Table 6). All SNPs identified

214 using *GEMMA* were also identified in *LFMM*. Individuals homozygous for these shared loci

215 were on opposite ends of the gill raker spectrum and heterozygous individuals intermediate

216 (Fig. 4d; Supplementary Fig. 5a). Blaufelchen and weissfelchen were at either extreme of

217 'pelagic' allele frequencies (allele associated with higher GRN), which in gangfisch varied

218 with spawning depth (Fig. 4b). Thus, gill raker number is likely controlled by a common

219 genetic basis across species. Furthermore, using the same approaches, we identified 56 SNPs

220 (two using *GEMMA*, 54 in *LFMM*) associated with body shape in gangfisch (Fig. 4c;

221 Supplementary Table 6). Although these SNPs did not overlap across analyses, one SNP

222 located in the *MTHFD1L* gene associated with body shape was also strongly associated with

223 GRN (Locus_56834; Fig. 4c; Supplementary Table 6).

224 Thus, we suggest that a sparse genetic architecture shared across ecologically distinct

225 whitefish subspecies underlies the rapid expansion in gill raker number. Genetic linkage or

226 pleiotropy are potential mechanisms for explaining this rapid and correlated diversification of

227 body shape and gill raker number.

228

229 **Genetic differentiation.** To understand if gill raker number (GRN) expansion was driven by

230 natural selection, we tested patterns of differentiation and diversity across the genome. Using

231 *pcadapt*, we found signatures of selection at 19 SNPs, including two of the six shared SNPs

232 that had been associated with gill raker number (Fig. 4f, Supplementary Fig. 5b;

233   Supplementary Table 6). Forty of the 99 SNPs associated with gill raker number in both

234   genome-wide association analyses, including those identified with *pcadapt*, were genetically

235   differentiated (ZFst>4; Supplementary Fig. 6a) between benthic and pelagic genetic clusters

236   in gangfisch (Fig. 4a) and showed reduced diversity in one of the two genetic clusters

237   (Supplementary Fig. 6b,c; Supplementary Table 6). This is reflective of a response to positive

238   selection[31], as genetic differentiation is less affected by genomic features, such as low

239   recombination regions, at early stages of divergence[32,33], and supports a role of selection

240   driving diversification in gill raker number. Only three of the 56 body shape-associated SNPs

241   were significantly differentiated between the genetic clusters in gangfisch, suggesting that

242   selection is acting less dramatically on body shape (Supplementary Fig. 5c). However, the

243   shared SNP associated with GRN and body shape (Locus_56834) showed strong

244   differentiation. Patterns of genetic differentiation were correlated across subspecies

245   comparisons (Spearman's rho=0.11; P<0.001; Methods)), with 20 out of 165 SNPs

246   differentiated in gangfisch also differentiated (ZFst>4) between blaufelchen and weissfelchen

247   (Supplementary Fig. 6f). This suggests that genetically differentiated regions, including those

248   associated with phenotypic divergence (e.g. shared Locus_56834) in gangfisch, are also

249   involved in the divergence between weissfelchen and blaufelchen, either due to adaptive

250   introgression or shared responses to selection. However, the majority of outlier SNPs were

251   private to gangfisch, indicating a unique response to selection.

252   Furthermore, at early stages of diversification we would expect small clustered, rather

253   than widespread, regions of differentiation across the genome. Indeed, signatures of selection

254   were restricted to few regions. Notably, one region on linkage group 38 was particularly

255   strongly diverged (Fig. 4g) and two outlier regions (linkage group 8 (65.239 cM) and 21

256   (9.24 cM)) overlapped with previously identified quantitative trait loci for variation in gill

257 raker number and differential depth selection in lake whitefish respectively[34,35], suggesting a

258 direct link between genetic divergence, genetic architecture, and phenotypic trait variation.

259   Thus, we conclude that these signals of selection and genetic divergence in genomic

260 regions associated with functional traits demonstrate that the expansion of gill raker numbers

261 was driven by natural selection.

262

263 **Gene regulatory and functional basis of rapid divergence.** To investigate the functional

264 molecular bases of trophic niche expansion in ecological context, we conducted a

265 transcriptome-wide analysis of gene expression in gangfisch. While we observed a degree of

266 divergence by depth across 31,872 genes (PCA), the separation was strongest by gill raker

267 number (GRN) (Fig. 5a; Supplementary Fig. 7a). To quantitatively assess the relative

268 contributions of site, depth, sex, gill raker number, and body shape on gene expression

269 divergence we used a linear mixed model. Gill raker number explained the greatest

270 proportion of overall gene expression variation with an average of 5.7%, followed by body

271 shape (explained 2.9%), with only negligible amounts of variation explained by site (1.3%),

272 depth (1.8%), and sex (0.7%) (Supplementary Fig. 7b). The majority of gene expression

273 variation was attributed to residual factors (86.9%), as expected given levels of individual

274 variation and background differences in natural environments[36,37]. Overall, this analysis

275 identified a critical set of genes associated with two key trophic traits, which are important

276 candidates for identifying ecologically relevant biological functions of adaptation and

277 diversification in gangfisch.

278   A single co-expression network was constructed based on 5,438 candidate genes

279 (those within the 90th percentile of expression variation explained by GRN and/or body

280 shape; Supplementary Fig. 7c), which clustered into 14 modules. Five modules were

281 significantly correlated with GRN, one of which (brown) was also correlated with genetic

282　ancestry (Fig. 5c, Supplementary Fig. 7e). A further four modules were significantly

283　correlated with body shape, with one module (blue) also correlated with genetic ancestry (Fig

284　5d, Supplementary Fig. 7f). Co-expression modules that were correlated with GRN and body

285　shape were significantly enriched for 39 pathways (Supplementary Table 7). Modules

286　associated with higher GRN (i.e. pelagic adaptation) showed significant over-representation

287　of pathways (PANTHER) involved in immune response, metabolism, vasodilation and

288　muscle contraction; including integrin signalling (P00042, P=4.12E-03), metabotropic

289　glutamate receptor (P00041, P=4.56E-02), and heme biosynthesis (P02746, P=4.43E-02)

290　(Supplementary Table 7). In contrast, lower GRN and deeper head and body shape (i.e.

291　benthic adaptation) showed significant over-representation of pathways involved in growth

292　and morphogenesis, including TGF-beta signalling (P00052, P=1.10E-02), integrin signalling

293　(P00034, P=5.96E-06), and BMP/activin signalling (P06211, P=4.41E-02) (Supplementary

294　Table 7). In general, many pathways and processes associated with co-expressed gene

295　modules have been linked to trophic and habitat divergence in other fish species[38–40],

296　indicating a crucial and conserved role in niche expansion and ecological adaptation.

297　　　　Of these 39 expression-associated pathways, nine were also found to be associated

298　with candidate loci putatively under selection and/or associated with gill raker number and

299　body shape (N=179 genes), including the TGF-beta receptor signaling (P00052) and integrin

300　signalling (P00034) pathways (Supplementary Table 8). Particularly, the TGF-beta pathway

301　has previously been associated with benthic-pelagic divergence in other salmonids[41] and

302　trophic craniofacial development and diversification, e.g in cichlids and zebrafish[42]. Although

303　none of the candidate loci-associated pathways were significantly enriched, the overlap

304　across analyses suggests that genetic and expression changes affect similar biological

305　functions. The extent to which the observed divergence in gene expression in gangfisch is

306　driven by environmental (i.e. plastic response) vs genetic basis is not determined. However,

307   the significant correlation of two co-expression modules (brown and blue; Fig. 5c-d) with the

308   genetic ancestry coefficient and the overlap with candidate loci pathways strongly suggests

309   that expression divergence is, at least in part, associated with genomic divergence.

310   Overall, we conclude that the rapid evolution of functional diversity in gangfisch is

311   based on functional molecular changes that play a recognised role in development and

312   ecological specialisation in fishes.

313

314   **Conclusions.**

315   By studying the functional phenotypic and genomic diversity of whitefish in Lake Constance,

316   we demonstrate the ability of biodiversity to re-emerge following ecosystem restoration.

317   First, we showed that gangfisch diversified in a functional trait (gill raker number) linked to

318   ecological and morphological diversity and this occurred rapidly, within only a few

319   generations following re-oligotrophication. This new diversity is intraspecific and does not

320   redress, ecologically or genetically, the recognised loss of species richness caused by

321   eutrophication[3,12]. Second, we showed that this phenotypic change has a sparse underlying

322   genetic architecture and that genetic and regulatory changes in recognised adaptive pathways

323   are significantly associated with gill raker number. The phenotypic and genetic correlation of

324   functional diversity (gill raker number) and eco-morphology (body shape) might explain the

325   rapid diversification in both traits. Phenotypic plasticity could facilitate this rapid change,

326   although the presence of genetic divergence and selection at phenotype-associated loci

327   strongly suggests that genetic changes underlie, at least partially, this diversification. Third,

328   we suggest that this rapid niche expansion was possible due to the genetic diversity created

329   through introgression and 'speciation reversal' during eutrophication, in agreement with

330   previous research using microsatellite loci[4]. The introgression of alleles associated with low

331   gill raker numbers from extinct benthic species (kilch or sandfelchen) could have contributed

332 to the emergence of lower gill raker numbers in gangfisch. Overall, our results indicate that

333 ecological diversity can start to re-emerge rapidly following species collapse with

334 hybridisation if few adaptive alleles with large effect underlie functional traits, eco-

335 morphology, and potentially reproductive isolation[11]. Thus, our findings are consistent with

336 the hypothesis that, while ecosystem remediation can have almost immediate benefits for

337 biodiversity, the potential for recovery is likely contingent on genetic architecture, ecological

338 context, and evolutionary history.

339

Methods.

Sampling of European whitefish. Gangfisch (*Coregonus lavaretus macrophthalmus)* were sampled in Lake Constance during spawning season in December 2014. Fish were collected using gill nets at three different spawning sites within the upper lake basin and three depths per site (Supplementary Fig. 1, Supplementary Table 2). Three gill nets with different mesh sizes (32, 42 and 44 mm knot-to-knot mesh size) were set at each depth to sample as much diversity in body size as possible. All fish, with the exception of those to be used for transcriptomic analyses, were immediately euthanized and stored on ice. Specimens to be used for gene expression analyses were held in aerated containers during transportation to ensure fish did not die prior to collection of tissues for RNA extractions. Even though this procedure may introduce stress-related gene expression changes, this did not affect comparisons across depths and sites as all individuals were treated equally. Tissue samples for RNA extractions were collected within five minutes of euthanasia by cutting a one centimeter-thick area of white muscle from above the lateral line, posterior to the operculum and stored in RNALater. We randomly selected an additional 27 individuals per site, depth and mesh size (total of 30 per site and depth including fish sampled for transcriptomics) for phenotypic analysis and tissue collection for genetic analysis. We sampled 20 individuals from each of the other two extant subspecies in Lake Constance, the pelagic blaufelchen (*Coregonus lavaretus wartmanni*) from the upper lake, and the benthic lower lake weissfelchen (*Coregonus lavaretus fera*) in October/November 2015 (Supplementary Fig. 1, Supplemenatry Table 2). For all individuals, muscle tissue and fin clips were taken and stored in absolute ethanol for subsequent genetic analysis. Scientific fish collections in Lake Constance were carried out under a permit from the Regierungspräsidium Tübingen: 'Aktenzeichen 33-4/9220.51-3

365

Collection of eco-morphological data. Several morphological and ecological measurements were taken to determine the breadth of phenotypic and ecological diversity across species, sites and depths. Individuals were weighed and standard length was measured, to the nearest mg and mm, respectively, and then photographed on their left side for morphometric analyses (see Supplementary methods for details). Scales were taken from underneath the ventral fins for aging and back calculation of growth rates based on three scales per fish (see Supplementary methods for details). The first left branchial gill arch was removed and gill rakers (including rudiments) were counted under a dissecting microscope. Furthermore, the length of the central gill raker was measured under a dissecting microscope to the nearest 0.01mm.

Furthermore, we collected 1 mg of white muscle tissue from the left side of the fish, posterior of the operculum, below the lateral line, for stable isotope analyses for gangfisch (N=111, approximately 10 individuals per site and depth), to infer trophic differences in the weeks of feeding before spawning. Lipids were removed prior to the stable isotope analysis. Methods followed[43]. Internal laboratory standards indicated measurement errors (SD) of ± 0.03‰, 0.05‰ for $\delta^{13}C$ and $\delta^{15}N$, respectively.

In addition to stable isotope signatures, we analysed the presence of eye fluke parasites, *Diplostomum spp.*, as more long-term indicators of a littoral-benthic habitat use of gangfisch (see Supplementary methods and[22] for details). To determine *Diplostomum* infection, both left and right eyes were removed and dissected under a dissecting microscope. The lens and *vitreous humour* were examined and infection was recorded as present/absent for all individuals (N=270).

388

Head and body shape analysis. A geometric morphometric approach was used to assess diversity in head and body shape among subspecies (N=269 gangfisch; N=20 weissfelchen, N=20 blaufelchen). Landmarks were chosen based on previous studies in European whitefish[15,44] (Supplementary Fig. 2a) and digitised using tpsDig 2.30[45]. Prior to shape analyses, we used the 'unbend' function in tpsUtil *v*.1.74 to control for the effect of body arching on standing and semi-landmark positions. Procrustes superimposition was conducted in MorphoJ *v*1.06b[46] to standardise landmark configurations. Procrustes coordinates were subjected to a multivariate, pooled within-group regression against the log centroid size to remove shape change associated with body size[47]. Residuals derived from the regression analysis were used as shape variables for all subsequent analyses. To visualise morphological variation, we conducted a canonical variance analysis (CVA), discriminating by subspecies and depth. To examine differences in morphology and assess the degree of phenotypic variation in the context of sampling depths without predefining groups, we performed a single principal component analysis (PCA) on gangfisch from all depths and sites. We first performed a multivariate analysis using the complete set of principal components (N=66) using a MANOVA. Furthermore, we performed univariate three-way ANOVAs on the first three principal components (explaining >68% of total variation) to determine the effect of depth on shape, including site, sex and their interactions as fixed effects.

Analyses of ecological and morphological traits. We estimated correlations of stable isotope values for carbon and nitrogen ($\partial^{13}C$ and $\partial^{15}N$) and trophic morphology (gill raker number (GRN)), using linear models in R. We tested the leverage of potential outliers based on cook's distance. As none of the potential outliers had significant effects on the regression curves, we retained the full dataset. To further test the effect of diet on GRN, we compared the number of gill rakers between individuals that were infected or non-infected by

414 *Diplostomum spp.* parasites (an indicator of long-term littoral-benthic habitat use[22]), using a

415 Wilcoxon rank sum test in R.

416   To estimate the correlation of a range of morphological traits (major axes of body

417 shape (PC1, PC2, PC3), standard length (SL) and the back-calculated total length at age 1, 2

418 and 3, size-corrected gill raker length), and because we were interested in the individual

419 correlations with GRN, we performed linear regression for each trait, with GRN as the

420 response variable, using the *lm* function in R: GRN ~ Trait (e.g. body shape PC2). We also

421 tested the effect of sex on GRN by running a linear model of all traits against GRN, with sex

422 as a co-variable. As sex did not have an effect on GRN, we did not include it in the final

423 models. P-values of all linear regressions were corrected together for multiple testing using

424 Benjamini-Hochberg[48] correction with the *p.adjust* R-package.

425   To estimate the evolutionary rate of shifts in GRN, we calculated the absolute change

426 in the mean GRN as standard deviations per generations, referred to as haldane[25,49], using

427 historical gill raker data for gangfisch collected along a similar depth gradient in 1990[24] and

428 gill raker data from this study in 2014 (see Supplementary methods for sampling details).

429 Haldanes were calculated as follows: $[(\ln(x_2)/sd_{\ln(x)})) - (\ln(x_1)/sd_{\ln(x)})]/t$, where $\ln(x_1)$ and

430 $\ln(x_2)$ are means of natural log-transformed GRNs at each time point, t is the interval between

431 sampling times in number of generations between, and $sd_{\ln(x)}$ is the pooled standard deviation

432 of $\ln(x_2)$ and $\ln(x_1)$[6,25,49]. We used a generation time of three years, based on the age

433 distribution in our dataset at spawning time.

434

435 **Genotyping, bioinformatic processing, and summary statistics.** DNA was extracted from

436 fin clips and muscle tissue using the NucleoSpin Tissue kit (Macherey-Nagel) and double-

437 digest restriction site associated DNA (ddRAD) libraries were prepared for 170 individuals

438 using *MspI* and *PstI-HF* restriction enzymes, following the methods in Recknagel et al.

439 (2015) with modified adapters for Illumina sequencing platforms. Each library was

440 sequenced on a single NextSeq500 lane (75bp paired-end sequencing) including ~10% Phi-X

441 at Glasgow Polyomics. We checked the quality of raw reads for each library using *FastQC*

442 v.0.11.3[51] (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

443 *Stacks* v.1.46[52] was used for demultiplexing, SNP calling, and filtering. First, we

444 processed the raw reads and trimmed reads to 60bp using *process_radtags*. The

445 *denovo_map.pl* pipeline was used for the *de novo* assembly of RAD loci. We used the

446 *rxstacks* module for genotype correction and haplotype pruning using the default settings.

447 The *populations* module was used for calling, extracting, and filtering genotypes using the

448 filtering criteria as described in the supplementary methods. The filtering and conversion of

449 datasets was performed using *Stacks, vcftools v. 0.1.15*[53]*, PLINK v.*1.90[54] (www.cog-

450 genomics.org/plink/1.9/) and *PGDspider v.*2.11.2[55].

451

452 **Population structure and admixture.** We used several approaches to resolve population

453 structure, including all three subspecies, within Lake Constance. First, we used *Admixture*

454 *v.*1.3.0[56] to detect the most likely number of genetic clusters, testing K=1 to K=12.

455 Furthermore, we performed a PCA using the *adegenet* package in R[57,58]. To improve

456 visualisation of population structuring in the multidimensional space, we performed

457 discriminant analyses of principal components using *adegenet*, clustering gangfisch

458 individuals by site and depth. *Genodive v.*2.0b27[59] was used to calculate pairwise Fst[60] among

459 all subspecies and populations, with gangfisch clustered by sampling depth and site (10,000

460 permutations).

461

462 **Evolutionary history of genetic divergence and introgression.** We analysed the

463 evolutionary history of divergence and introgression in Lake Constance whitefish using two

464 different approaches. First, we used *f3*-statistics implemented in the *treemix v*.1.13[61] *threepop*

465 function to test for introgressive hybridization in a three-population comparison. The dataset

466 contained all gangfisch as one focal population, and weissfelchen and blaufelchen as two

467 separate reference populations using the global SNP dataset.

468 Furthermore, we used a coalescence modelling approach implemented in *fastimcoal2*

469 *v*.2.5.2.3[62] based on the information contained in multidimensional site frequency spectra

470 (SFS) to determine the history of gene flow, divergence times, and timing and strength of

471 introgression in whitefish from Lake Constance. Three-population SFSs were created using

472 *δaδi v*.1.6.3[63] based on a specifically filtered SNP dataset (N=22,196 SNPs). The minor

473 folded SFS was used for the analysis, as no trinucleotide substitution matrix is available for

474 salmonids and no outgroup species was sequenced for ancestral state reconstruction and

475 correction. To determine absolute values for divergence times and other parameters, we

476 corrected the number of monomorphic sites as outlined in[64]. A mutation rate of 1×10^{-8} was

477 used[65]. Based on the genetic population structure, we used a model in which blaufelchen split

478 first from the ancestor of weissfelchen and gangfisch and a subsequent split of weissfelchen

479 and gangfisch (Supplementary Fig. 4e). We compared a total of seven different demographic

480 models with different histories and combinations of gene flow and introgression, ranging

481 from strict-isolation to a complete isolation-with-migration model with or without

482 introgression (see Supplementary methods and Supplementary Fig. 4e). We ran 30 iterations

483 for each model, each consisting of 40 rounds of parameter estimation with 100,000

484 coalescence simulations. We inferred the best fitting model based on the AIC [62] and ran it an

485 additional 45 iterations and used the top 10 iterations (based on highest estimated maximum

486 likelihood) for non-parametric bootstrap resampling (*bootstrap* R package) to estimate means

487 and 95%-CI for each parameter[66]. In addition to test the potential effect of selection on the

488 demographic inference, we performed the same analysis using a SNP dataset without sites

489   potentially under selection and differentiated among genetic clusters (see below) (N = 22,100

490   SNPs).

491

492   **Correlation of genetic ancestry and eco-morphology in gangfisch.** To assess how

493   morphology correlates with genetic ancestry in gangfisch, we calculated spearman's rank

494   correlation coefficients for body shape PC2 and GRN with proportion of genetic ancestry

495   inferred by *Admixture*. PC2 of body shape was used for this (and all subsequent analyses) as

496   it depicts the main body shape variation along the depth gradient (Supplementary Fig. 2c). P-

497   values were adjusted for multiple testing using a Benjamini-Hochberg correction using the

498   *p.adjust* R-package.

499

500   **Genome-wide scans for selection.** To detect loci that are differentiated between genetic

501   clusters, we combined two approaches to identify and narrow down candidate loci. First, we

502   calculated the per site pairwise Fst[60] and nucleotide diversity between and within gangfisch

503   from without significant admixture (pure genetic clusters: genetic ancestry coefficient > 0.75

504   or <0.25 in the *Admixture* analysis; N=135) using *vcftools*. To determine whether the basis of

505   eco-morphological variation is similar across subspecies, we compared genetic

506   differentiation between benthic and pelagic gangfisch with that of pelagic blaufelchen and

507   benthic-littoral weissfelchen. We estimated per site pairwise Fst between weissfelchen and

508   blaufelchen, and calculated the spearman's rank coefficient correlation of these Fst values

509   with Fst values calculated between genetic clusters in gangfisch using the *cor.test* function in

510   R. To analyse the distribution of differentiated loci across the genome, we anchored the

511   anonymous RAD-loci to the available RAD-Seq based lake whitefish (*Coregonus*

512   *clupeaformis*) linkage map[34] using the *MapComp* software package[67] and the previously

513   described approach[67,68] using the Atlantic salmon reference genome (ICSASG_v2)[69]. The

514   maximum distance for pairing was set to 2 Mb and conducted in 10 iterations.

515   As the Fst-based analysis ignores the more gradual nature of genetic differentiation

516   along the depth gradient, we also used a (genetic) PCA based approach for detecting loci

517   under selection that does not require the grouping of individuals. We used the *pcadapt* R-

518   package[70] to identify loci that differentiate along the first principal component, as this axis

519   differentiates gangfisch from different spawning depths and therefore different phenotypes.

520   We corrected for multiple testing using a false discovery rate cut-off of 1% using the *qvalue*

521   R-package.

522

523   **Association analysis with eco-morphology.** To identify variants associated with GRN and

524   body shape (PC2), we used three approaches: a linear mixed model (LMM) and a linear

525   Bayesian sparse linear mixed model (BSLMM) implemented in *GEMMA*[71], and a latent-

526   factor linear mixed model approached implemented in *LFMM* (N_{inds}=153)[72]. The analysis for

527   GRN and body shape were performed separately but with the same settings.

528   We imputed missing genotypes using a random forest approach implemented in the

529   *radiator* R-package, using 150 trees. We first estimated phenotype-genotype association for

530   each SNP using a univariate linear mixed model and the Wald test, using α=0.0001 (–

531    $\log10(\alpha) = 4$) as a significance threshold. Second, we used a linear BSLMM to estimate SNP

532   effect sizes (posterior inclusion probability: PIP) on phenotype and further estimate

533   hyperparameters quantitatively describing the genetic architecture underlying trait variation

534   (see Supplementary methods for details). For both analyses in *GEMMA*, we corrected for

535   relatedness and population structure using a centred relatedness matrix, which we calculated

536   using *GEMMA*. Lastly, we used *LFMM*, specifying two latent factors to correct for

537   population structure, and averaged the results across 10 independent runs using default

538   settings. Candidate loci associated with body shape or GRN were detected as those SNPs

539   with FDR < 0.01. *Post hoc* we checked for overlap of phenotype-associated loci across

540   analyses.

541

542   **Annotation of candidate loci.** We identified genes associated with candidate loci

543   (significantly diverged, associated with GRN or body shape, genetically diverged and under

544   selection; N=177) by blasting RAD loci sequences against the Atlantic salmon reference

545   genome (ICSASG_v2)[69] using blastn in Blast+[73], and identifying genes containing candidate

546   loci or within 50kb of a candidate locus. We only kept loci with an E-value below 1e-13 and

547   the best (or equally good results if a locus blasted to different locations) blast results. We

548   used the PANTHER (Protein Analysis Through Evolutionary Relationships) classification

549   tool[74] to identify functional pathways associated with candidate loci and overrepresented

550   biological process gene ontology terms, using the human gene ontology set as background.

551

552   **Transcriptomics.** Total RNA was extracted from white muscle tissue from 27 individuals

553   (three per site and depth) and RNA-seq libraries were prepared and sequenced by Glasgow

554   Polyomics research facility (University of Glasgow) (see Supplementary methods for details).

555   Individual cDNA libraries were synthesised for all 27 samples using the TruSeq Stranded

556   mRNA Sample Preparation kit (Illumina, San Diego, CA), with a Poly-A selection step.

557   Libraries were sequenced on the NextSeq 500 system (Illumina, San Diego, CA) using

558   paired-end sequencing (75 bp from each end), at a sequencing depth of 25-35 million reads

559   per library, yielding approximately 822 million reads in total.

560   Raw reads were processed prior to alignment, by removing adapter sequences with

561   *Scythe v*0.9944 BETA and trimming low quality reads with *Trimmomatic v*0.36[75]. FastQC

562   *v*0.11.2[51] was used to assess read quality before and after pre-processing. The resulting reads

563 were aligned to a transcriptome assembly for European whitefish[76], consisting of 33,697

564 annotated transcripts. Reads were aligned using *Bowtie2* (--all, --local, --no-mixed, --no-

565 discordant)[77], and transcript abundance was quantified with *eXpress v*1.5.1[78]. Count data

566 were processed using *DESeq2 v*3.5[79], transcripts with <20 reads across all samples removed

567 (yielding a final dataset of 31,872 transcripts), and count data $\log_2$ transformed.

568 Transcriptional profiles were visualised using a PCA on the full dataset using the *plotPCA*

569 function in *DESeq2* and plots were generated using the *ggpplot2 v*2.2.1.9000.

570   We used a linear mixed model implemented in the R-package *variancePartition*

571 *v*3.6[80] to identify genes associated with GRN and body shape (BS-PC2) as fixed variables,

572 including site, depth and sex as random effects (see Supplementary methods for details). We

573 retained genes for a subsequent network analysis based on the amount of expression variation

574 explained by either GRN or BS-PC2, namely if the expression variation explained by GRN or

575 BS-PC2 was above the 90th-quantile of the variation distribution for all genes associated with

576 the respective trait.

577   Subsequently, a Weighted Gene Co-Expression Network Analysis (WGCNA)

578 approach[81] using the *WGCNA* R package[82] was used to identify co-expression modules

579 associated with trophic divergence based on the pre-filtered gene set (see Supplementary

580 methods for details). Network modules were defined using the dynamic *treecut* algorithm,

581 with a minimum module size of 30 genes and a cut height of 0.981. The module *eigengene*

582 distance threshold was set to 0.25 to merge highly similar modules.

583   To identify genetic pathways associated with trophic niche divergence in gangfisch,

584 we performed functional enrichment analyses on co-expressed modules. Separate enrichment

585 analyses were performed for each module, using the Over-Representation Analysis (ORA)

586 function in PANTHER[74]. Module genes were compared against the background set of 31,872

587 transcripts to determine over-representation of pathways. All functional analyses were

588   conducted using gene symbols from the annotated transcriptome assembly, and significance

589   was determined using Fisher's Exact test with Bonferroni correction (FDR < 0.05).

590

591   **Data availability.** The sequence datasets have been deposited in NCBI Sequence Read

592   Archive with the BioProject accession code PRJNA497182 (corresponding to BioSample

593   accessions SAMN10250325 to SAMN10250521). Phenotype and ecological data are

594   available at the 'Enlighten: Research Data' repository of the University of Glasgow:

595   http://dx.doi.org/10.5525/gla.researchdata.680

596

References.

1. Hautier, Y. *et al.* Anthropogenic environmental changes affect ecosystem stability via biodiversity. *Science (80-.).* **348,** 336–340 (2015).

2. Hendry, A. P., Gotanda, K. M. & Svensson, E. I. Human influences on evolution, and the ecological and societal consequences. *Phil. Trans. R. Soc. B* **372,** 20160028 (2017).

3. Des Roches, S. *et al.* The ecological importance of intraspecific variation. *Nat. Ecol. Evol.* **2,** (2017).

4. Vonlanthen, P. *et al.* Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482,** 357–62 (2012).

5. Taylor, E. B. *et al.* Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Mol. Ecol.* **15,** 343–355 (2006).

6. Hendry, A. P., Farrugia, T. H. O. J. & Kinnison, M. T. Human influences on rates of phenotypic change in wild animal populations. *Mol. Ecol.* **17,** 20–29 (2008).

7. Rudman, S. M. & Schluter, D. Ecological impacts of reverse speciation in threespine stickleback. *Curr. Biol.* **26,** 490–495 (2016).

8. Bullock, J. M., Aronson, J., Newton, A. C., Pywell, R. F. & Rey-Benayas, J. M. Restoration of ecosystem services and biodiversity: conflicts and opportunities. *Trends Ecol. Evol.* **26,** 541–549 (2011).

9. Alexander, T. J., Vonlanthen, P. & Seehausen, O. Does eutrophication-driven evolution change aquatic ecosystems? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372,** (2017).

10. Gilman, R. T. & Behm, J. E. Hybridization, species collapse, and species reemergence after disturbance to premating mechanisms of reproductive isolation. *Evolution (N. Y).* **65,** 2592–2605 (2011).

622 11. Yeaman, S. & Whitlock, M. C. The genetic architecture of adaptation under migration-
623 selection balance. *Evolution (N. Y)*. **65,** 1897–1911 (2011).

624 12. Schluter, D. Evidence for Ecological Speciation and Its Alternative. *Science (80-.)*.
625 **323,** 737–741 (2009).

626 13. Maan, M. E. & Seehausen, O. Ecology, sexual selection and speciation. *Ecol. Lett.* **14,**
627 591–602 (2011).

628 14. Nümann, W. Attempt of a classification of the coregonines in the Lake of Constance
629 with regard to the so-called Blaufelchen by combination of several characteristics.
630 *Achiv Fur Hydriobiologie* **82,** 500–521 (1978).

631 15. Hirsch, P. E., Eckmann, R., Oppelt, C. & Behrmann-Godel, J. Phenotypic and genetic
632 divergence within a single whitefish form - detecting the potential for future
633 divergence. *Evol. Appl.* **6,** 1119–1132 (2013).

634 16. Jochimsen, M. C., Kümmerlin, R. & Straile, D. Compensatory dynamics and the
635 stability of phytoplankton biomass during four decades of eutrophication and
636 oligotrophication. *Ecol. Lett.* **16,** 81–89 (2013).

637 17. Thomas, G., Quoss, H., Hartmann, J. & Eckmann, R. Human-induced changes in the
638 reproductive traits of Lake Constance common whitefish (Coregonus lavaretus). *J.*
639 *Evol. Biol.* **22,** 88–96 (2009).

640 18. Lundsgaard-Hansen, B., Matthews, B., Vonlanthen, P., Taverna, a. & Seehausen, O.
641 Adaptive plasticity and genetic divergence in feeding efficiency during parallel
642 adaptive radiation of whitefish (Coregonus spp.). *J. Evol. Biol.* **26,** 483–498 (2013).

643 19. Harrod, C., Mallela, J. & Kahilainen, K. K. Phenotype-environment correlations in a
644 putative whitefish adaptive radiation. *J. Anim. Ecol.* **79,** 1057–68 (2010).

645 20. Østbye, K., Bernatchez, L., Naesje, T. F., Himberg, K.-J. M. & Hindar, K.
646 Evolutionary history of the European whitefish Coregonus lavaretus (L.) species

647   complex as inferred from mtDNA phylogeography and gill-raker numbers. *Mol. Ecol.*

648   **14,** 4371–87 (2005).

649   21.   Quevedo, M., Svanbäck, R. & Eklöv, P. Intrapopulation niche partitioning in a

650   generalist predator limits food web connectivity. *Ecology* **90,** 2263–2274 (2009).

651   22.   Behrmann-Godel, J. Parasite identification, succession and infection pathways in perch

652   fry (Perca fluviatilis): new insights through a combined morphological and genetic

653   approach. *Parasitology* **140,** 509–520 (2013).

654   23.   Vonlanthen, P. *et al.* Divergence along a steep ecological gradient in lake whitefish

655   (Coregonus sp.). *J. Evol. Biol.* **22,** 498–514 (2009).

656   24.   Luczynski, M., Rösch, R., Vuorinen, J. A. & Brzuzan, P. Biochemical genetic study of

657   sympatric Lake Constance whitefish (Coregonus lavaretus) populations: Blaufelchen

658   and Gangfisch. *Aquat. Sci.* **57,** 136–143 (1995).

659   25.   Gingerich, P. D. Quantification and comparison of evolutionary rates. *Am. J. Sci.* **293,**

660   453–478 (1993).

661   26.   Hamilton, J. A. & Miller, J. M. Adaptive introgression as a resource for management

662   and genetic conservation in a changing climate. *Conserv. Biol.* **30,** 33–41 (2016).

663   27.   Shafer, A. B. A. & Wolf, J. B. W. Widespread evidence for incipient ecological

664   speciation: A meta-analysis of isolation-by-ecology. *Ecol. Lett.* **16,** 940–950 (2013).

665   28.   Laporte, M. *et al.* RAD-QTL Mapping Reveals Both Genome-Level Parallelism and

666   Different Genetic Architecture Underlying the Evolution of Body Shape in Lake

667   Whitefish (Coregonus clupeaformis) Species Pairs. *G3 Genes|Genomes|Genetics* **5,**

668   1481–1491 (2015).

669   29.   Chaves, J. A. *et al.* Genomic variation at the tips of the adaptive radiation of Darwin's

670   finches. *Mol. Ecol.* **25,** 5282–5295 (2016).

671   30.   Pfeifer, S. P. *et al.* The evolutionary history of Nebraska deer mice: local adaptation in

672 the face of strong gene flow. *Mol. Biol. Evol.* **35,** 792–806 (2018).

673 31. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of

674 speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23,** 3133–57

675 (2014).

676 32. Burri, R. Interpreting differentiation landscapes in the light of long-term linked

677 selection. *Evol. Lett.* **1,** 118–131 (2017).

678 33. Delmore, K. E. *et al.* Comparative analysis examining patterns of genomic

679 differentiation across multiple episodes of population divergence in birds. *Evol. Lett.* **2,**

680 76–87 (2018).

681 34. Gagnaire, P.-A., Normandeau, E., Pavey, S. a & Bernatchez, L. Mapping phenotypic,

682 expression and transmission ratio distortion QTL using RAD markers in the Lake

683 Whitefish (Coregonus clupeaformis). *Mol. Ecol.* **22,** 3036–48 (2013).

684 35. Jacobs, A., Womack, R., Chen, M., Gharbi, K. & Elmer, K. Significant Synteny and

685 Co-localization of Ecologically Relevant Quantitative Trait Loci Within and Across

686 Species of Salmonid Fishes. *Genetics* **207,** 741–754 (2017).

687 36. Ranz, J. M. & Machado, C. A. Uncovering evolutionary patterns of gene expression

688 using microarrays. *Trends Ecol. Evol.* **21,** 29–37 (2006).

689 37. Gibson, G. The environmental contribution to gene expression profiles. *Nat. Rev.*

690 *Genet.* **9,** 575–581 (2008).

691 38. Jeukens, J., Renaut, S., St-Cyr, J., Nolte, A. W. & Bernatchez, L. The transcriptomics

692 of sympatric dwarf and normal lake whitefish (Coregonus clupeaformis spp.,

693 Salmonidae) divergence as revealed by next-generation sequencing. *Mol. Ecol.* **19,**

694 5389–5403 (2010).

695 39. Park, P. J. & Bell, M. A. Variation of telencephalon morphology of the threespine

696 stickleback (Gasterosteus aculeatus) in relation to inferred ecology. *J. Evol. Biol.* **23,**

697   1261–1277 (2010).

698   40.   Conejeros, P. *et al.* Differentiation of Sympatric Arctic Char Morphotypes Using

699   Major Histocompatibility Class II Genes Differentiation of Sympatric Arctic Char

700   Morphotypes Using Major Histocompatibility Class II Genes. *Trans. Am. Fish. Soc.*

701   **143,** 586–594 (2014).

702   41.   Jacobs, A. *et al.* Convergence in form and function overcomes non-parallel

703   evolutionary histories in Arctic charr . *bioRxiv* (2018). doi:10.1101/265272

704   42.   Ahi, E. Signalling pathways in trophic skeletal development and morphogenesis:

705   Insights from studies on teleost fish. *Dev. Biol.* **420,** 11–31 (2016).

706   43.   Yohannes, E., Grimm, C., Rothhaupt, K.-O. & Behrmann-Godel, J. The Effect of

707   Parasite Infection on Stable Isotope Turnover Rates of $\delta15N$, $\delta13C$ and $\delta34S$ in

708   Multiple Tissues of Eurasian Perch Perca fluviatilis. *PLoS One* **12,** (2017).

709   44.   Siwertsson, A., Knudsen, R., Adams, C. E., Præbel, K. & Amundsen, P. A. Parallel

710   and non-parallel morphological divergence among foraging specialists in European

711   whitefish (Coregonus lavaretus). *Ecol. Evol.* **3,** 1590–1602 (2013).

712   45.   Rohlf, F. J. TpsDig. *Dep. Ecol. Evol. State Univ. New York, Stony Brook, NY* (2004).

713   at

714   <http://scholar.google.com/scholar?q=TpsDig&btnG=&hl=en&num=20&as_sdt=0%2

715   C22 VN - readcube.com>

716   46.   Klingenberg, C. P. MorphoJ: An integrated software package for geometric

717   morphometrics. *Mol. Ecol. Resour.* **11,** 353–357 (2011).

718   47.   Klingenberg, C. P. & McIntyre, G. S. Geometric Morphometrics of Developmental

719   Instability: Analyzing Patterns of Fluctuating Asymmetry with Procrustes Methods.

720   *Evolution (N. Y).* **52,** 1363 (1998).

721   48.   Benjamini, Y. & Hochberg, Y. Benjamini Y, Hochberg Y. Controlling the false

722   discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*

723   **57,** 289–300 (1995).

724   49.   Haldane, J. B. S. Suggestions as to Quantitative Measurement of Rates of Evolution.

725   *Evolution (N. Y).* **3,** 51–56 (1949).

726   50.   Recknagel, H., Jacobs, A., Herzyk, P. & Elmer, K. R. Double-digest RAD sequencing

727   using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms.

728   *Mol. Ecol. Resour.* **15,** 1316–1329 (2015).

729   51.   Andrews, S. FastQC: A quality control tool for high throughput sequence data.

730   *Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/*

731   http://www.bioinformatics.babraham.ac.uk/projects/ (2010). doi:citeulike-article-

732   id:11583827

733   52.   Catchen, J. M. *et al.* Stacks: Building and Genotyping Loci De Novo From Short-Read

734   Sequences. *G3 Genes, Genomes, Genet.* **1,** 171–182 (2011).

735   53.   Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27,** 2156–

736   2158 (2011).

737   54.   Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and

738   richer datasets. *Gigascience* **4,** 1–16 (2015).

739   55.   Lischer, H. E. L. & Excoffier, L. PGDSpider: An automated data conversion tool for

740   connecting population genetics and genomics programs. *Bioinformatics* **28,** 298–299

741   (2012).

742   56.   Alexander, D. H. & Novembre, J. Fast Model-Based Estimation of Ancestry in

743   Unrelated Individuals. *Genome Res.* **19,** 1655–1664 (2009).

744   57.   Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers.

745   *Bioinformatics* **24,** 1403–1405 (2008).

746   58.   Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide

747   SNP data. *Bioinformatics* **27,** 3070–3071 (2011).

748   59.   Meirmans, P. G. & Tienderen, P. H. genotype and genodive: two programs for the

749   analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* **4,** 792–794

750   (2004).

751   60.   Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population

752   structure. *Evolution (N. Y)*. **38,** 1358–1370 (1984).

753   61.   Pickrell, J. K. & Pritchard, J. K. Inference of Population Splits and Mixtures from

754   Genome-Wide Allele Frequency Data. *PLoS Genet.* **8,** (2012).

755   62.   Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust

756   Demographic Inference from Genomic and SNP Data. *PLoS Genet.* **9,** (2013).

757   63.   Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring

758   the Joint Demographic History of Multiple Populations from Multidimensional SNP

759   Frequency Data. *PLoS Genet.* **5,** (2009).

760   64.   Kautt, A. F., Machado-Schiaffino, G. & Meyer, A. Multispecies Outcomes of

761   Sympatric Speciation after Admixture with the Source Population in Two Radiations

762   of Nicaraguan Crater Lake Cichlids. *PLoS Genet.* **12,** (2016).

763   65.   Rougeux, C., Bernatchez, L. & Gagnaire, P.-A. A. Modeling the Multiple Facets of

764   Speciation-with-Gene-Flow toward Inferring the Divergence History of Lake

765   Whitefish Species Pairs (Coregonus clupeaformis). *Genome Biol. Evol.* **9,** 2057–2074

766   (2017).

767   66.   Egger, B., Rösti, M., Böhne, A., Roth, O. & Salzburger, W. Demography and genome

768   divergence of lake and stream populations of an East African cichlid fish. *Mol. Ecol.*

769   **26,** 5016–5030 (2017).

770   67.   Sutherland, B. J. G. *et al.* Salmonid Chromosome Evolution as Revealed by a Novel

771   Method for Comparing RADseq Linkage Maps. *Genome Biol. Evol.* **8,** 3600–3617

772   (2016).

773   68.   Moore, J. *et al.* Genomics and telemetry suggest a role for migration harshness in

774   determining overwintering habitat choice, but not gene flow, in anadromous Arctic

775   Char. *Mol. Ecol.* **26,** 6784–6800 (2017).

776   69.   Lien, S. *et al.* The Atlantic salmon genome provides insights into rediploidization.

777   *Nature* **533,** 200–205 (2016).

778   70.   Luu, K., Bazin, E. & Blum, M. G. pcadapt: an R package to perform genome scans for

779   selection based on principal component analysis. *Mol. Ecol. Resour.* **17,** 67–77 (2017).

780   71.   Zhou, X., Carbonetto, P. & Stephens, M. Polygenic Modeling with Bayesian Sparse

781   Linear Mixed Models. *PLoS Genet.* **9,** (2013).

782   72.   Frichot, E., Schoville, S. D., Bouchard, G. & François, O. Testing for associations

783   between loci and environmental gradients using latent factor mixed models. *Mol. Biol.*

784   *Evol.* **30,** 1687–1699 (2013).

785   73.   Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,**

786   421 (2009).

787   74.   Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene

788   function analysis with the panther classification system. *Nat. Protoc.* **8,** 1551–1566

789   (2013).

790   75.   Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina

791   sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

792   76.   Carruthers, M. *et al.* De novo transcriptome assembly, annotation and comparison of

793   four ecological and evolutionary model salmonid fish species. *BMC Genomics* **19,** 32

794   (2018).

795   77.   Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.*

796   *Methods* **9,** (2012).

797   78.   Roberts, A. & Pachter, L. Streaming fragment assignment for real-time analysis of

798   sequencing experiments. *Nat. Methods* **10,** 71–73 (2013).

799   79.   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and

800   dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

801   80.   Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in

802   complex gene expression studies. *BMC Bioinformatics* **17,** 483 (2016).

803   81.   Zhang, B. & Horvath, S. A General Framework for Weighted Gene Co-Expression

804   Network Analysis. *Stat. Appl. Genet. Mol. Biol.* **4,** (2005).

805   82.   Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation

806   network analysis. *BMC Bioinformatics* **9,** 1–13 (2008).

807

815 **Author contributions:** JB-G and KRE designed the experiment. AJ, MC, KRE and JB-G

816 collected data. MC generated and analysed eco-morphological and transcriptomic data. AJ

817 generated and analysed genomic data, analysed eco-morphological and stable isotope data.

818 RE analysed life history data. EY generated stable isotope data. CEA, JB-G and KRE

819 supervised the project. AJ with MC, JB-G and KRE wrote the paper. All authors commented

820 on and approved the final manuscript.

821

822 **Author Information:** Reprints and permissions information is available at
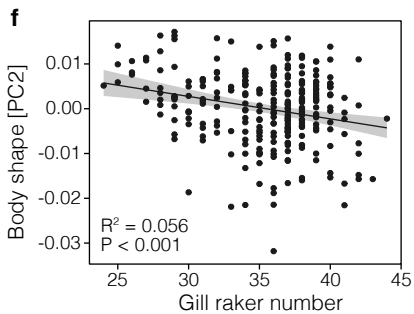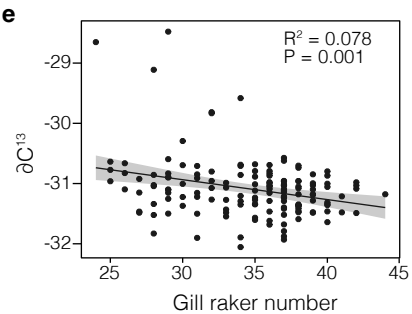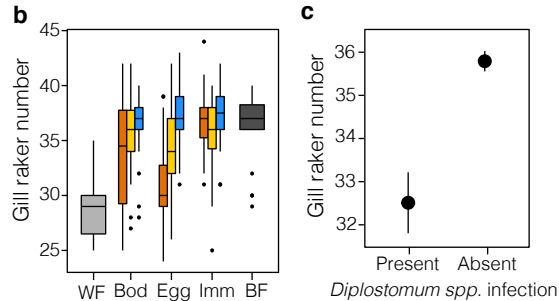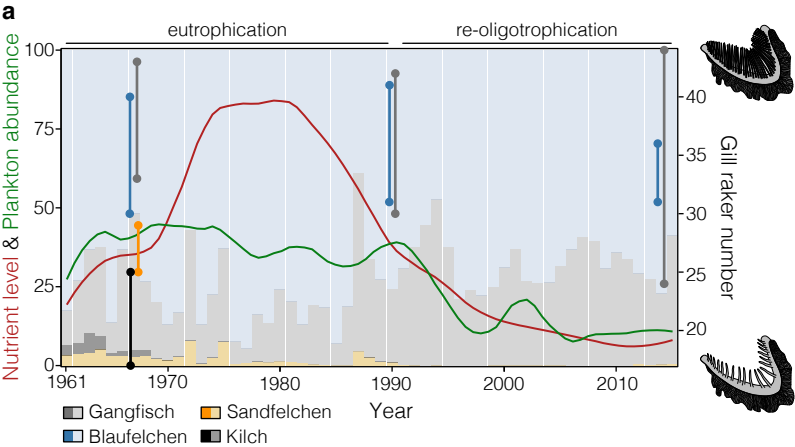
823 www.nature.com/reprints. The authors declare no competing interests. Correspondence and

824 requests for materials should be addressed to jasminca.behrmann@uni-konstanz.de and/or

825 kathryn.elmer@glasgow.ac.uk

826

Figure legends:

829 **Figure 1. Phenotypic and ecological diversity. a,** Historical environmental and gill raker

830 data during eutrophication and re-oligotrophication of Lake Constance from 1961 until 2015.

831 The red and green curve show historical fluctuations in total phosphorus (annual mean; μL^{-1})

832 and the total zooplankton abundance ((individuals per m^{-2} x 10^4)/10^5). The bar plots show the

833 annual catch composition of whitefish in Lake Constance by subspecies (in percent of total

834 catch) over the same time period. The range in gill raker number for gangfisch (grey),

835 blaufelchen (blue), sandfelchen (orange) and kilch (black) are shown for the years 1967, 1990

836 and 2014. Note that data for kilch and sandfelchen were only available for the year 1967, as

837 they went extinct or were very rare following the eutrophication. Illustrations of gill raker

838 arches next to the plot illustrate the two extremes (N_{low}=24; N_{high}=44) of the gill raker

839 spectrum in contemporary gangfisch. **b,** Boxplots [bar = median, box range = range between

840 third and first quartile (interquartile range; IQR), whiskers = extend to furthest point (highest

841 or lowest) no further than 1.5 times the IQR, points = outlier] showing the distribution gill

842 raker number by subspecies, and by sampling site and depth for gangfisch (N = 309). Species

843 and sampling sites: WF = weissfelchen, BF = blaufelchen; gangfisch from: Bod = Bodman,

844 Egg = Egg and Imm = Immenstaad. **c,** Difference in the number of gill rakers (mean ± SEM)

845 between gangfisch infected and non-infected with *Diplostomum spp*. (Wilcoxon rank sum

846 test: P < 0.001; N = 269). **d,** Canonical variant plot showing the variation in body shape

847 between whitefish, with individuals grouped by subspecies and gangfisch individuals further

848 split by spawning depth. Wireframes depict body shape at outer most point of CV1 (40.6%).

849 **e-f,** Correlation of gill raker number with carbon stable isotope signatures (**e**; linear model:

850 $F_{1,146}$=13.35, P < 0.001) and PC2 of body shape (**f**; linear model: $F_{1,266}$=16.76, P < 0.001) in

851 gangfisch. Grey areas depict the standard error around the regression line.

a

eutrophication | re-oligotrophication

Nutrient level & Plankton abundance / Red / Green

Gangfisch | Sandfelchen
Blaufelchen | Kilch

Gill raker number

Year

b

Gill raker number

WF | Bod | Egg | Imm | BF

c

Gill raker number

Present | Absent

Diplostomum spp. infection

d

shallow | intermediate | deep

Egg | Imm | Bod | WF | BF

CV2 (38.4%)

CV1 (40.6%)

e

$R^2 = 0.078$
$P = 0.001$

∂C^{13}

Gill raker number

f

Body shape [PC2]

$R^2 = 0.056$
$P < 0.001$

Gill raker number

852 **Figure 2. Rate of change in gill raker number and range. a**, Distribution of gill rakers in

853 gangfisch in 1990 and 2014 illustrated by a density plot and mean ± SD. The distribution of

854 individual data points is shown by jittered points. **b**, Evolutionary rates of phenotypic change

855 plotted in (absolute) haldanes, standard deviations of change per generation, for the shift in

856 the mean gill raker number in gangfisch in Lake Constance from this study (orange, N=1)

857 and previously published rates of phenotypic change for other salmonid species (red, N=132)

858 as well as for different taxa (white, N=2224)[6]. Absolute values of haldanes and generations

859 were log-transformed.

860

a

2014

1990

Gill raker number

b

Haldanes

Generations
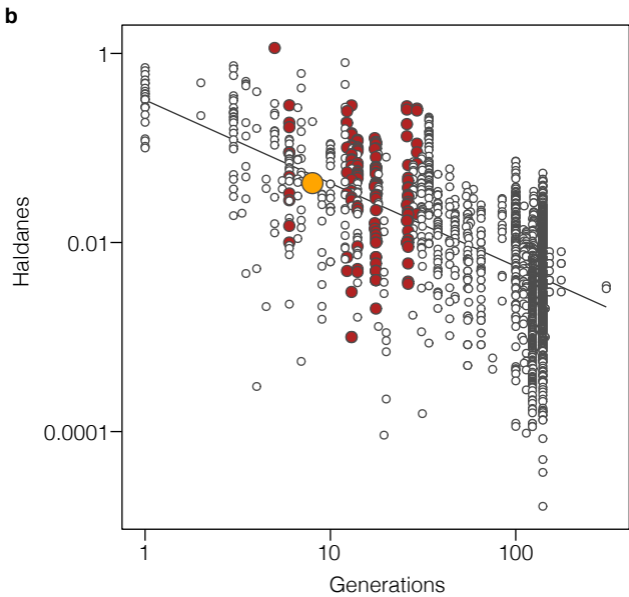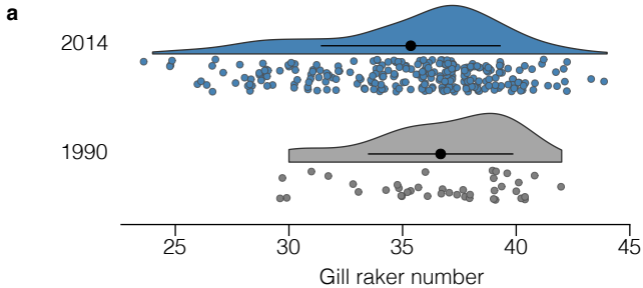
861 **Figure 3. Evolutionary history of introgression. a**, The *f3*-statistic revealed significant

862 signatures of introgressive hybridization in gangfisch (GF) from weissfelchen (WF) and

863 blaufelchen (BF) (GF:BF,WF; *f3* = -0.0011 ± 0.00014 SEM, z-score = -7.97). **b**, The most

864 likely demographic model for the evolutionary history of whitefish in Lake Constance based

865 on the full SNP dataset. Divergence times are in years (95%-CI) and the proportion of

866 admixture (95%-CI) between subspecies is given above the solid black arrows. The

867 proportion of admixture describes the proportion of the gene pool in gangfisch that was

868 replaced by either WF or BF. Detailed parameter values for migration rates (grey arrows) and

869 effective population sizes are given in Supplementary Table 3. The proportions of genetic

870 ancestry for each individual by subspecies are given below.

871

a

b
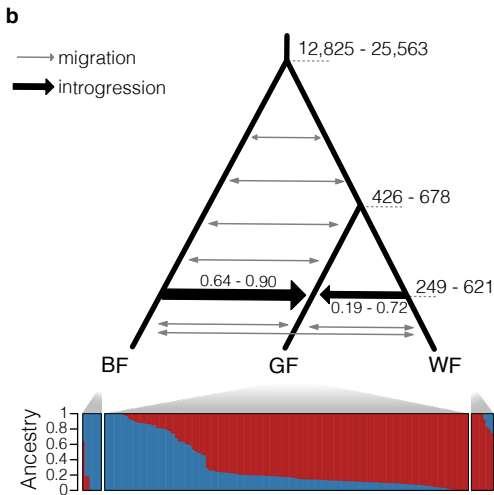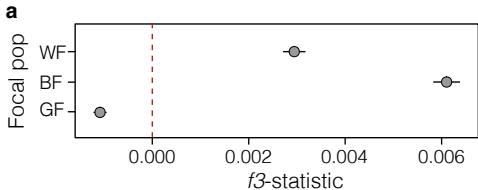
872 **Figure 4. Genotype-phenotype associations and signatures of selection. a**, Correlation

873 between the genetic ancestry coefficient and number of gill rakers in gangfisch (Spearman

874 test: rho=-0.6007; P<0.001; N=153). Histograms and density lines show the parameter

875 distributions. Individuals with an ancestry coefficient above 0.75 were defined as 'benthic',

876 whereas individuals with an ancestry coefficient below 0.25 as 'pelagic', as these are more

877 similar to weissfelchen and blaufelchen, respectively. **b**, Frequency of the pelagic alleles

878 (mean ± SD) at loci strongly associated with gill raker number (N_{loci}=6), by subspecies and

879 spawning depth in gangfisch. We defined the allele with a higher frequency in blaufelchen

880 compared to weissfelchen as the 'pelagic' allele. **c**, Correlation between genotype-association

881 with gill raker number (LFMM-GRN) and with body shape PC2 (LFMM-BS) across all

882 SNPs (N=12,976). The association results from *LFMM* are given by the z-score. Loci

883 associated with GRN and BS in LFMM are highlighted red (N=6). Loci only associated with

884 GRN are highlighted in yellow (N=93), whereas loci only associated with BS are highlighted

885 in cyan (N=48). Locus 56834, which is strongly associated with GRN and BS is annotated. **d**

886 **and e**, Gill raker number and body shape (PC2) for gangfisch that are homozygous or

887 heterozygous for the benthic (B) or pelagic (P) alleles on locus 56834 (ANOVA-GRN:

888 $F_{2,126}$=29.82, P < 0.001; ANOVA-BS: $F_{2,127}$=7.97; P < 0.001). The dotplot shows the

889 distribution of values and the mean ± SD. **f**, Correlation of association with gill raker number

890 (LMM) and signal of selection (PCAdapt) for 12,976 SNPs in gangfisch. Dashed lines show

891 the applied significance thresholds for each analysis. The size of dots indicates the effect size

892 of each SNP on the variation in gill raker number and the colour gradient shows the degree of

893 genetic differentiation (standardised Fst: ZFst) between benthic and pelagic gangfisch. **g**,

894 Manhattan plot showing the distribution of genetic differentiation between benthic and

895 pelagic gangfisch along the lake whitefish linkage map for a subset of 3,350 loci.

896 Significantly differentiated SNPs (ZFst > 4, dashed line) are highlighted in red.
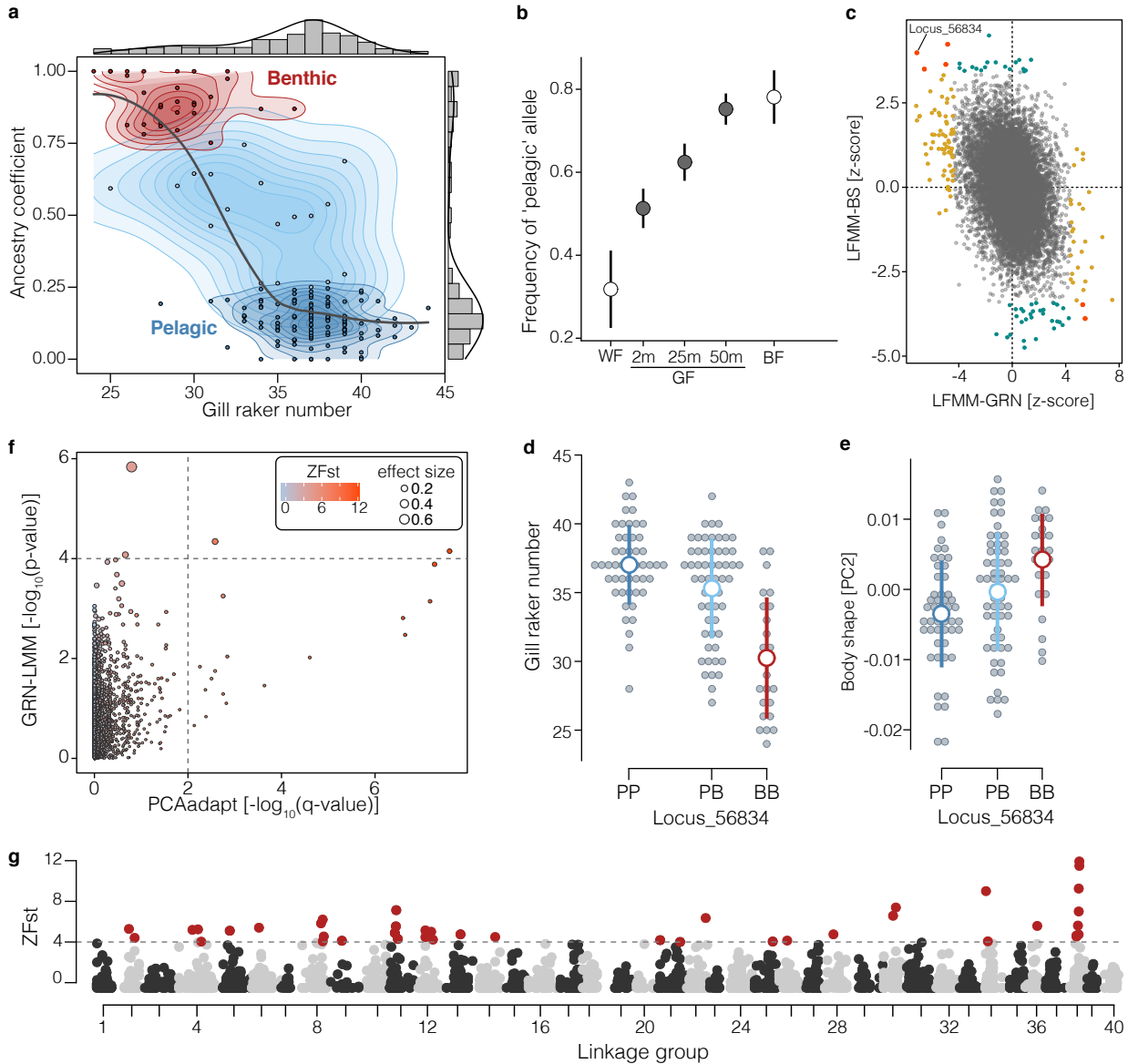
897 **Figure 5. Functional gene expression variation in gangfisch. a,** Transcriptional profiles of

898 gangfisch collected from all three sampling sites (Bodman represented as circles, Egg as

899 triangles and Immenstaad as squares), and spawning depths in Lake Constance analysed

900 using a principal component analysis based on the complete set of filtered genes (>20 read

901 counts across all samples; N=31,872 genes). Gangfisch are coloured based on their number of

902 gill rakers, using a gradient scale (low GRN 'benthic' = red, high GRN 'pelagic' = blue).

903 Each point represents an individual (N=27, three per depth, per site), and the centroids for

904 each depth group are plotted and labelled. **b,** Correlation between WGCNA module

905 eigengenes and measured traits (site, depth, sex, genetic ancestry coefficient (GAC), gill

906 raker number (GRN) and body shape (BS-PC2); N=27). Each row corresponds to a module

907 (identified on the left side by its colour; colours are assigned arbitrarily by WGCNA), and

908 each column corresponds to a trait. The number of genes present within each module is given

909 in parentheses below the module colour label. Pearson's correlation coefficients for module-

910 trait relationships are given in each cell, with the corresponding p-values in parenthesis. All

911 p-values are corrected for multiple testing using Benjamini-Hochberg and significant

912 correlations (FDR < 0.05) are highlighted in bold. Cells are coloured based on their

913 correlation (see legend). No significant module associations were identified for site, depth or

914 sex. **c,** Associations of module eigengene expression for the brown module with gill raker

915 number (GRN) and genetic ancestry coefficient (GAC). **d,** Associations of module eigengene

916 expression for the blue module with body shape (BS-PC2) and genetic ancestry coefficient

917 (GAC). In panels **c** and **d**, the grey area shows the standard error around the regression line.

918 Individuals are coloured by genetic ancestry (ancestry coefficient above > 0.75 for respective

919 clusters based on SNP-based *Admixture* analysis). R^2-values for significant module-trait

920 relationships are given in the figures.

921

a PC2 (9%) vs PC1 (15%)

Gill raker number: 25–40 (color scale, red to blue)

Gangfisch sampling sites:
- Egg (triangle)
- Immenstaad (square)
- Bodman (circle)

Labels: Shallow, Deep, Intermediate

b

	Site	Depth	Sex	GAC	GRN	BS-PC2
Red (N=320)	0.043 (0.9)	0.42 (0.1)	0.041 (1)	0.3 (0.2)	**0.54 (0.01)**	0.083 (0.7)
Green (N=402)	-0.24 (0.6)	0.27 (0.3)	-0.0048 (1)	0.024 (0.9)	0.33 (0.2)	0.03 (0.9)
Yellow (N=463)	-0.2 (0.6)	0.38 (0.1)	-0.033 (1)	0.19 (0.4)	**0.52 (0.02)**	0.21 (0.4)
Brown (N=612)	0.28 (0.6)	0.32 (0.2)	0.23 (1)	**0.65 (0.003)**	**0.6 (0.006)**	0.35 (0.1)
Purple (N=129)	-0.094 (0.8)	0.5 (0.1)	-0.26 (1)	0.35 (0.1)	0.31 (0.2)	0.44 (0.06)
MidnightBlue (N=58)	0.34 (0.4)	0.2 (0.4)	-0.011 (1)	0.48 (0.05)	**0.62 (0.006)**	0.16 (0.5)
Black (N=258)	-0.4 (0.3)	0.063 (0.8)	0.001 (1)	0.13 (0.6)	-0.16 (0.4)	**0.69 (0.001)**
Cyan (N=174)	-0.17 (0.7)	0.24 (0.3)	-0.0074 (1)	0.22 (0.4)	0.35 (0.1)	0.31 (0.2)
Turquoise (N=1,746)	0.13 (0.7)	-0.45 (0.1)	0.13 (1)	-0.2 (0.4)	**-0.53 (0.01)**	0.15 (0.5)
Tan (N=119)	-0.067 (0.9)	-0.13 (0.5)	-0.15 (1)	-0.43 (0.08)	-0.16 (0.4)	**-0.55 (0.01)**
LightCyan (N=54)	0.14 (0.7)	-0.17 (0.5)	-0.063 (1)	-0.21 (0.4)	0.19 (0.4)	**-0.61 (0.004)**
GreenYellow (N=120)	-0.038 (0.9)	0.23 (0.4)	-0.13 (1)	-0.43 (0.08)	-0.2 (0.4)	-0.43 (0.06)
Blue (N=766)	0.22 (0.6)	-0.28 (0.3)	-0.0022 (1)	**-0.59 (0.008)**	-0.28 (0.2)	**-0.65 (0.002)**
Pink (N=217)	0.38 (0.3)	-0.39 (0.1)	0.066 (1)	-0.39 (0.1)	-0.44 (0.05)	-0.29 (0.2)

c Brown module

$R^2_{GAC} = 0.65$
$R^2_{GRN} = 0.6$

Module expression (PC1) vs Gill raker number

Genetic ancestry: Benthic (red), Pelagic (blue)

d Blue Module

$R^2_{GAC} = -0.59$
$R^2_{BS-PC2} = -0.65$

Module expression (PC1) vs Body shape [PC2]