
There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

[http://eprints.gla.ac.uk/171335/](http://eprints.gla.ac.uk/171335/)

Deposited on: 07 November 2018
Rapid niche expansion by selection on functional genomic variation after ecosystem recovery

Arne Jacobs\textsuperscript{1†}, Madeleine Carruthers\textsuperscript{1†}, Reiner Eckmann\textsuperscript{2}, Elizabeth Yohannes\textsuperscript{2}, Colin E. Adams\textsuperscript{1,3}, Jasminca Behrmann-Godel\textsuperscript{2*}, & Kathryn R. Elmer\textsuperscript{1*}

Affiliation:

\textsuperscript{1}Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G12 8QQ, Scotland, UK

\textsuperscript{2}Limnological Institute, University of Konstanz, Konstanz, 78457 Germany

\textsuperscript{3}Scottish Centre for Ecology and the Natural Environment, University of Glasgow, Rowardennan, G63 0AW, Scotland, UK

*Corresponding

† These authors contributed equally to this work
Introduction.

It is well recognised that environmental degradation caused by human activities can result in dramatic losses of species and diversity. However, comparatively little is known about the ability of biodiversity to re-emerge following ecosystem recovery. Here we show that a European whitefish subspecies, the gangfisch (*Coregonus lavaretus macrophthalmus*), rapidly increased its ecologically functional diversity following the restoration of Lake Constance after anthropogenic eutrophication. In fewer than 10 generations, gangfisch evolved a greater range of gill raker number to utilise a broader ecological niche. A sparse genetic architecture underlies this variation in gill raker number. Several co-expressed gene modules and genes showing signals of positive selection were associated with gill raker number and body shape. These were enriched for biological pathways related to trophic niche expansion in fishes. Our findings demonstrate the potential of functional diversity to expand following habitat restoration, given a fortuitous combination of genetic architecture, genetic diversity, and selection.
Anthropogenic activities are one of the major drivers of environmental change, often with critical impacts on ecosystem health\textsuperscript{1,2}. Environmental degradation caused by human disturbance can be extremely challenging for organisms, resulting in rapid evolutionary and ecosystem changes\textsuperscript{3}, interbreeding and hybridisation, extirpation, or even extinction\textsuperscript{2,4–6}. For example, hybridisation across closely related species can result in a loss of species richness (‘speciation reversal’) and associated functional diversity in adaptive traits\textsuperscript{4,5,7}. In some cases, polluted environments can be improved by remediation efforts, but the potential for diversity to re-emerge following ecosystem restoration remains a major unanswered issue in evolutionary and conservation biology, with important implications for policy and practical efforts\textsuperscript{8,9}. Theoretical models have predicted that diversity can recover but only under particular genetic, ecological and evolutionary circumstances\textsuperscript{10}. These models suggested that the probability of functional diversity to re-emerge from a collapsed species group (e.g. ‘hybrid swarm’) is higher when environmental disturbance is temporary (as longer lasting disturbances, even weak ones, risk the loss of adaptive genetic diversity), and when functional trait diversity has a simple genetic basis and is linked to assortative mating or mating-related traits\textsuperscript{10,11}. For example, it has been shown that fish often mate with individuals that are similar\textsuperscript{12,13}; thus, if functional traits are linked to e.g. body shape, divergence in a functional trait can lead to morphological divergence and further on to genetic divergence via trait-based assortative mating\textsuperscript{10}. However, re-emergence of functional diversity following a species collapse has not been observed nor empirically tested to date.

In many aquatic ecosystems, high levels of nutrient pollution from agriculture and sewage during the mid-20\textsuperscript{th} century caused cultural eutrophication. The algae blooms and consequent negative impacts on water quality caused the collapse of natural trophic and
habitat gradients in many freshwater lakes. In the Alpine lakes, this resulted in the dramatic loss of whitefish ecological and functional diversity, highlighted by the average reduction in the range of gill raker numbers by 14% across lakes\(^4\). For example, in one of the largest lakes, Lake Constance, the benthic specialists (kilch and sandfelchen) went extinct or collapsed, with extensive hybridisation among those and the remaining pelagic (blaufelchen and gangfisch) and littoral-benthic (weissfelchen) subspecies\(^4\). This depleted functional diversity, evidenced as a 28% reduction in the range of gill raker number across whitefish species\(^4,14\) (Fig. 1a, Supplementary Table 1). As a result, individuals with low gill raker numbers disappeared from Lake Constance during and after eutrophication\(^4,15\). However, concerted effort from the 1980s to reduce polluting nutrient inputs enabled Lake Constance to revert rapidly to its pristine oligotrophic state\(^16\) (Fig. 1a). This provides an excellent system to study biodiversity recovery after ecosystem remediation, as long-term data are available for the fish and lake conditions, natural recolonisation from other whitefish populations is unlikely, and, in contrast to other alpine lakes, Constance has not been subjected to external stocking\(^15,17\). Here, we explored the variation in ecological, morphological, genomic, and gene expression traits to assess the material underpinning a rapid re-emergence of functional diversity in whitefish.

**Results and Discussion.**

**Functional and eco-morphological divergence.** To examine contemporary phenotypic diversity and its relationship to habitat use and trophic niche in whitefish, we focused on the gangfisch because it is the only subspecies that has modified its spawning depth since ecosystem recovery\(^12\). Research proposed that gangfisch may be ecologically expanding because spawning depth is strongly associated with trophic ecology in European
whitefish\textsuperscript{15,18}. Thus, we assessed gill rakers and body shape across spawning depths (shallow: 2-5 m, intermediate: 20-25 m, and deep: below 50 m) at multiple sites in the lake (Supplementary Figs. 1 and 2a). Gill rakers are important trophic traits in many postglacial fish species; many, dense and long rakers are used to filter-feed zooplankton while few, sparse and short rakers are associated with benthos feeding, and gill raker number is well documented to be genetically based\textsuperscript{4,19,20}. Indeed, in line with expectations we found gill raker number (GRN) to be significantly structured by depth (Fig. 1b; $F_{2,260}=17.27$, $P<0.001$), with shallow-spawning individuals having fewer gill rakers (GRN$_{\text{shallow}}=33.9 \pm 4.7$ standard error of the mean (SEM)) than deep-spawning individuals (GRN$_{\text{deep}}=37.0 \pm 2.8$ SEM). In addition, we detected a significant effect of depth on gill raker length (Supplementary Fig. 2; GRL: $F_{2,265}=4.824$, $P=0.0088$), and number and length were positively correlated, with individuals with fewer gill rakers also having shorter gill rakers (Supplementary Fig. 3a; GRL: $R^2=0.028$; $F_{1,267}=8.76$, $P=0.0034$).

Individuals spawning at different depths also differed in body shape (MANCOVA: Pillai’s Trace$_{\text{Depth}}=0.722$, df=(1.66), $P<0.001$), an important eco-morphological trait (Supplementary Fig. 2); shallow-spawners were deeper bodied and had a sub-terminal mouth compared to deep-spawning individuals, which had more slender bodies and terminal mouths (Fig. 1d; Supplementary Fig. 2b,c). Body shape divergence with depth was mostly explained by PC2 (Supplementary Fig. 2c) and variation in body shape was correlated to GRN (Fig. 1f; BS-PC2: $R^2=0.056$; $F_{1,266}=16.76$, $P<0.001$; Supplementary Fig. 3b,c), suggesting its association with trophic ecology in gangfisch. Additionally, shallow-spawning gangfisch were longer (standard length SL$_{\text{shallow}}=28.6 \pm 0.7$ cm vs SL$_{\text{deep}}=25.7 \pm 0.6$ cm) and heavier (body weight$_{\text{shallow}}=327.5 \pm 12.9$ g vs body weight$_{\text{deep}}=225.2 \pm 5.6$ g) than deep-spawning individuals (SL: $F_{2,261}=34.69$, $P<0.001$; body weight: $F_{2,261}=38.49$, $P<0.001$) (Supplementary Fig. 2d), evidencing that multiple phenotypic traits contribute to the gradient along the depth
Although divergence in body shape and standard length were correlated with GRN (Fig. 1, Supplementary Fig. 3), body length was only (negatively) correlated with GRN from age three onwards (back-calculated from scales; Supplementary Fig. 3e), indicating increased growth of putatively benthic whitefish later in ontogeny. However, the observed patterns were heterogeneous across sampling sites, as indicated by the effect of site and depth by site interaction on variation in body shape (MANCOVA: Pillai’s Trace<sub>site</sub>=1.243, df=(2,132), P<0.001; Pillai’s Trace<sub>site*depth</sub>=1.054, df=(2,132), P<0.001). Individuals from Immenstaad showed in general less divergence along the depth gradient (Supplementary results). In comparison to the other extant whitefish subspecies, gangfisch were intermediate in body shape to weissfelchen and blaufelchen, with shallow-spawning gangfisch being more similar to benthic-littoral weissfelchen and deep-spawning gangfisch being similar to pelagic blaufelchen (Fig. 1d).

To test if phenotypic diversity is correlated with trophic and habitat diversity in gangfisch, and therefore reflects functional utility, we assessed the correlation of gill raker number with ecological indicators: stable isotope signature and parasite infection. We found that the diversity associated with spawning depth was related to trophic and ecological differences (Fig.1; Supplementary Figs. 2,3). Specifically, gill raker number was correlated to carbon stable isotope signatures (Fig. 1c; R²=0.07751, P<0.001, df=147), indicating a more littoral-benthic diet in individuals with fewer gill rakers<sup>21</sup>. Fewer gill rakers were also associated with higher infection by Diplostomum eye flukes (Fig. 1c; Wilcoxon rank sum test, P<0.001), trematode parasites that infect snails in the shallow zone of the lake. Fish-infective stages hatch from the snails and therefore indicate littoral-benthic habitat use<sup>22</sup>. Overall, these observed differences in morphology and functional traits with spawning depth are typical of benthic-pelagic divergence in European whitefish trophic specialists along the
Thus, we focused on variation in the key functional trait of gill raker number as a proxy of overall trophic and eco-morphological diversity.

**Rapid trait expansion.** Identifying how functional diversity changes in response to rapid environmental change is important for understanding its drivers. We compared gill raker number (GRN) between the samples collected in 2014 for this study and samples collected in 1990\(^2\). This demonstrated that the functional trait diversity arose rapidly, as it was not present in gangfisch at the end of the eutrophic era in 1990 (Fig. 1a, 2a; range in GRN: 1990=30-42\(^2\), 2014=24-44). The range in GRN increased drastically over this 24-year period (corresponding to about eight whitefish generations), with a notable decrease in the minimum number from 30 to 24. The mean number of gill rakers decreased by 0.3423 standard deviations (SD; haldane numerator) over this period, equivalent to an evolutionary rate of -0.0428 haldanes (Fig. 2)\(^2\). This is comparable to rates of phenotypic evolution in several examples of rapid evolution, such as change in gill raker number in marine-freshwater sticklebacks or hindlimb length in anole lizards, and lies within the top 12% of reported rates of rapid phenotypic evolution\(^6\) (Fig. 2b). The shift in functional phenotypic diversity coincides temporally with the re-oligotrophication of Lake Constance. Thus, we propose that restored habitat gradients drove the emergence of functional diversity (GRN) in gangfisch by selection. This rapid expansion in gangfisch partially, but not completely, recovers the historical functional diversity that existed across the whitefish species complex in Lake Constance (pre-eutrophication GRN range across species: 17-43)\(^14,15\) (Fig. 1a, Supplementary Table S1).

**Evolutionary history of introgression.** To examine the evolutionary genetic changes underlying the observed rapid diversification, we analysed 13,589 genome-wide single
nucleotide polymorphisms (SNPs) within and across whitefish subspecies in Lake Constance. We found two genetic clusters, with blaufelchen and weissfelchen at extremes, and gangfisch showing an intermediate gradient of ancestry (Figs. 3b; Supplementary Fig. 4a,b), suggesting strong historic admixture and recent segregation in gangfisch. Indeed, we detected multiple signals consistent with historical genetic introgression from blaufelchen and weissfelchen into gangfisch, including the intermediacy of gangfisch in genetic ordination plots (Supplementary Fig. 4a-d), the divergence of gangfisch along the eigenvector distinguishing subspecies (Supplementary Fig. 4c,d), and support from introgression tests (Fig. 3a, f3 statistics=-0.00108362, z-score=-7.967).

To reconstruct the history of divergence and introgression between all extant subspecies, we used a demographic modelling approach based on the joint site frequency spectrum, built from the full SNP dataset and a dataset excluding differentiated SNPs (i.e. putatively neutral, Supplementary Fig. 4b). Based on the full dataset, we found the most likely demographic history to be one of strong and recent introgression from blaufelchen (mean admixture proportion=0.78) and weissfelchen (mean admixture proportion=0.45) into gangfisch (Fig. 3b, Supplementary Table 3), leading to the formation of a gangfisch hybrid swarm. Demographic analysis based on the neutral SNP dataset suggested a simple isolation-with-migration model to be more likely, although models including introgression were also highly supported (Supplementary Table 4). The weaker signal of introgression in the neutral SNP dataset suggests that loci which introgressed into gangfisch from other sympatric whitefish species are associated with its diversification. Thus, introgressive hybridisation across species in Lake Constance, including those that are now extinct, potentially provided the genetic material needed for the rapid eco-morphological diversification in gangfisch after ecosystem recovery.
Genetic basis of functional and eco-morphological traits. To assess the genetic underpinnings of eco-morphological changes, we first tested the correlation of functional diversity and genetic ancestry. Gill raker number (GRN) in gangfisch was highly correlated with genetic ancestry (Fig. 4a; Spearman test, r_{GRN}=-0.6007, P<0.001), suggesting genetic isolation by ecological adaptation. Sparsely-rakered gangfisch (‘benthic gangfisch’) were in general genetically more similar to the benthic-littoral weissfelchen, whereas densely-rakered gangfisch (‘pelagic gangfisch’) were genetically closer to pelagic blaufelchen (Fig. 4a, Supplementary Fig. 4a-d).

To further assess the genetic basis of functional diversity (variation in GRN) in gangfisch, we investigated its genetic architecture. Polygenic models suggested that GRN is controlled by a sparse genetic architecture, with most of the phenotypic variance (phenotypic variance explained by all SNPs: PVE_{GRN}=0.9991) being explained by a small expected number (predicted to be between 68 and 117 (95% CI)) of large effect loci (phenotypic variance explained by sparse effect loci: PGE_{GRN}=0.9718) (Supplementary Table 5). In comparison, we found that body shape, a highly polygenic trait in fishes that was also correlated with genetic ancestry (\rho_{BS-PC2}=0.334, P<0.001), had a more polygenic architecture, with only a small proportion of phenotypic variance explained by large effect loci (Supplementary Table 5; PVE_{BS-PC2}=0.8636; PGE_{BS-PC2}=0.2296). These results suggest differences in the underlying architecture of these traits, with the large proportion of variance explained (PVE) likely reflecting the high heritability of these traits and the sparse genetic architecture in the case of GRN (PGE_{GRN}>0.9), with similarly high values reported for beak size in Darwin’s finches (PVE=0.947, PGE=0.775) or dorsal brightness in deer mice (PVE=0.69, PGE=0.83). Statistical difficulties involved in estimating trait architectures or correlations with population structure could inflate these values. However, for this study the large difference in PGE for GRN compared to body shape (97% vs 23%) is particularly
informative and highlights the different architectures underlying these traits. Such a simple genetic architecture may facilitate the rapid expansion in gill raker number.

Next, we used genome-wide approaches to identify the loci underlying phenotypic variation. Association analyses using linear-mixed models (GEMMA) identified six SNPs associated with variation in GRN, whereas a latent-factor linear mixed model (LFMM) identified a set of 99 associated SNPs (Fig. 4; Supplementary Table 6). All SNPs identified using GEMMA were also identified in LFMM. Individuals homozygous for these shared loci were on opposite ends of the gill raker spectrum and heterozygous individuals intermediate (Fig. 4d; Supplementary Fig. 5a). Blaufelchen and weissfelchen were at either extreme of ‘pelagic’ allele frequencies (allele associated with higher GRN), which in gangfisch varied with spawning depth (Fig. 4b). Thus, gill raker number is likely controlled by a common genetic basis across species. Furthermore, using the same approaches, we identified 56 SNPs (two using GEMMA, 54 in LFMM) associated with body shape in gangfisch (Fig. 4c; Supplementary Table 6). Although these SNPs did not overlap across analyses, one SNP located in the MTHFD1L gene associated with body shape was also strongly associated with GRN (Locus_56834; Fig. 4c; Supplementary Table 6).

Thus, we suggest that a sparse genetic architecture shared across ecologically distinct whitefish subspecies underlies the rapid expansion in gill raker number. Genetic linkage or pleiotropy are potential mechanisms for explaining this rapid and correlated diversification of body shape and gill raker number.

**Genetic differentiation.** To understand if gill raker number (GRN) expansion was driven by natural selection, we tested patterns of differentiation and diversity across the genome. Using pcdadapt, we found signatures of selection at 19 SNPs, including two of the six shared SNPs that had been associated with gill raker number (Fig. 4f, Supplementary Fig. 5b;
Forty of the 99 SNPs associated with gill raker number in both genome-wide association analyses, including those identified with *pcadapt*, were genetically differentiated (ZFst>4; Supplementary Fig. 6a) between benthic and pelagic genetic clusters in gangfisch (Fig. 4a) and showed reduced diversity in one of the two genetic clusters (Supplementary Fig. 6b,c; Supplementary Table 6). This is reflective of a response to positive selection, as genetic differentiation is less affected by genomic features, such as low recombination regions, at early stages of divergence, and supports a role of selection driving diversification in gill raker number. Only three of the 56 body shape-associated SNPs were significantly differentiated between the genetic clusters in gangfisch, suggesting that selection is acting less dramatically on body shape (Supplementary Fig. 5c). However, the shared SNP associated with GRN and body shape (Locus_56834) showed strong differentiation. Patterns of genetic differentiation were correlated across subspecies comparisons (Spearman’s rho=0.11; P<0.001; Methods), with 20 out of 165 SNPs differentiated in gangfisch also differentiated (ZFst>4) between blaufelchen and weissfelchen (Supplementary Fig. 6f). This suggests that genetically differentiated regions, including those associated with phenotypic divergence (e.g. shared Locus_56834) in gangfisch, are also involved in the divergence between weissfelchen and blaufelchen, either due to adaptive introgression or shared responses to selection. However, the majority of outlier SNPs were private to gangfisch, indicating a unique response to selection.

Furthermore, at early stages of diversification we would expect small clustered, rather than widespread, regions of differentiation across the genome. Indeed, signatures of selection were restricted to few regions. Notably, one region on linkage group 38 was particularly strongly diverged (Fig. 4g) and two outlier regions (linkage group 8 (65.239 cM) and 21 (9.24 cM)) overlapped with previously identified quantitative trait loci for variation in gill
raker number and differential depth selection in lake whitefish respectively\textsuperscript{34,35}, suggesting a direct link between genetic divergence, genetic architecture, and phenotypic trait variation.

Thus, we conclude that these signals of selection and genetic divergence in genomic regions associated with functional traits demonstrate that the expansion of gill raker numbers was driven by natural selection.

**Gene regulatory and functional basis of rapid divergence.** To investigate the functional molecular bases of trophic niche expansion in ecological context, we conducted a transcriptome-wide analysis of gene expression in gangfisch. While we observed a degree of divergence by depth across 31,872 genes (PCA), the separation was strongest by gill raker number (GRN) (Fig. 5a; Supplementary Fig. 7a). To quantitatively assess the relative contributions of site, depth, sex, gill raker number, and body shape on gene expression divergence we used a linear mixed model. Gill raker number explained the greatest proportion of overall gene expression variation with an average of 5.7%, followed by body shape (explained 2.9%), with only negligible amounts of variation explained by site (1.3%), depth (1.8%), and sex (0.7%) (Supplementary Fig. 7b). The majority of gene expression variation was attributed to residual factors (86.9%), as expected given levels of individual variation and background differences in natural environments\textsuperscript{36,37}. Overall, this analysis identified a critical set of genes associated with two key trophic traits, which are important candidates for identifying ecologically relevant biological functions of adaptation and diversification in gangfisch.

A single co-expression network was constructed based on 5,438 candidate genes (those within the 90\textsuperscript{th} percentile of expression variation explained by GRN and/or body shape; Supplementary Fig. 7c), which clustered into 14 modules. Five modules were significantly correlated with GRN, one of which (brown) was also correlated with genetic
ancestry (Fig. 5c, Supplementary Fig. 7e). A further four modules were significantly correlated with body shape, with one module (blue) also correlated with genetic ancestry (Fig 5d, Supplementary Fig. 7f). Co-expression modules that were correlated with GRN and body shape were significantly enriched for 39 pathways (Supplementary Table 7). Modules associated with higher GRN (i.e. pelagic adaptation) showed significant over-representation of pathways (PANTHER) involved in immune response, metabolism, vasodilation and muscle contraction; including integrin signalling (P00042, P=4.12E-03), metabotropic glutamate receptor (P00041, P=4.56E-02), and heme biosynthesis (P02746, P=4.43E-02) (Supplementary Table 7). In contrast, lower GRN and deeper head and body shape (i.e. benthic adaptation) showed significant over-representation of pathways involved in growth and morphogenesis, including TGF-beta signalling (P00052, P=1.10E-02), integrin signalling (P00034, P=5.96E-06), and BMP/activin signalling (P06211, P=4.41E-02) (Supplementary Table 7). In general, many pathways and processes associated with co-expressed gene modules have been linked to trophic and habitat divergence in other fish species38–40, indicating a crucial and conserved role in niche expansion and ecological adaptation.

Of these 39 expression-associated pathways, nine were also found to be associated with candidate loci putatively under selection and/or associated with gill raker number and body shape (N=179 genes), including the TGF-beta receptor signaling (P00052) and integrin signalling (P00034) pathways (Supplementary Table 8). Particularly, the TGF-beta pathway has previously been associated with benthic-pelagic divergence in other salmonids41 and trophic craniofacial development and diversification, e.g in cichlids and zebrafish42. Although none of the candidate loci-associated pathways were significantly enriched, the overlap across analyses suggests that genetic and expression changes affect similar biological functions. The extent to which the observed divergence in gene expression in gangfisch is driven by environmental (i.e. plastic response) vs genetic basis is not determined. However,
the significant correlation of two co-expression modules (brown and blue; Fig. 5c-d) with the genetic ancestry coefficient and the overlap with candidate loci pathways strongly suggests that expression divergence is, at least in part, associated with genomic divergence.

Overall, we conclude that the rapid evolution of functional diversity in gangfisch is based on functional molecular changes that play a recognised role in development and ecological specialisation in fishes.

Conclusions.

By studying the functional phenotypic and genomic diversity of whitefish in Lake Constance, we demonstrate the ability of biodiversity to re-emerge following ecosystem restoration. First, we showed that gangfisch diversified in a functional trait (gill raker number) linked to ecological and morphological diversity and this occurred rapidly, within only a few generations following re-oligotrophication. This new diversity is intraspecific and does not redress, ecologically or genetically, the recognised loss of species richness caused by eutrophication. Second, we showed that this phenotypic change has a sparse underlying genetic architecture and that genetic and regulatory changes in recognised adaptive pathways are significantly associated with gill raker number. The phenotypic and genetic correlation of functional diversity (gill raker number) and eco-morphology (body shape) might explain the rapid diversification in both traits. Phenotypic plasticity could facilitate this rapid change, although the presence of genetic divergence and selection at phenotype-associated loci strongly suggests that genetic changes underlie, at least partially, this diversification. Third, we suggest that this rapid niche expansion was possible due to the genetic diversity created through introgression and ‘speciation reversal’ during eutrophication, in agreement with previous research using microsatellite loci. The introgression of alleles associated with low gill raker numbers from extinct benthic species (kilch or sandfelchen) could have contributed
to the emergence of lower gill raker numbers in gangfisch. Overall, our results indicate that ecological diversity can start to re-emerge rapidly following species collapse with hybridisation if few adaptive alleles with large effect underlie functional traits, ecophenotype, and potentially reproductive isolation\textsuperscript{11}. Thus, our findings are consistent with the hypothesis that, while ecosystem remediation can have almost immediate benefits for biodiversity, the potential for recovery is likely contingent on genetic architecture, ecological context, and evolutionary history.
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, Supplementary 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
**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\(^4\). 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\(^2\) 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).
**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 v1.06b\(^{46}\) 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 (\(\delta^{13}C\) and \(\delta^{15}N\)) 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
*Diplostomum spp.* parasites (an indicator of long-term littoral-benthic habitat use), using a Wilcoxon rank sum test in R.

To estimate the correlation of a range of morphological traits (major axes of body shape (PC1, PC2, PC3), standard length (SL) and the back-calculated total length at age 1, 2 and 3, size-corrected gill raker length), and because we were interested in the individual correlations with GRN, we performed linear regression for each trait, with GRN as the response variable, using the *lm* function in R: GRN ~ Trait (e.g. body shape PC2). We also tested the effect of sex on GRN by running a linear model of all traits against GRN, with sex as a co-variable. As sex did not have an effect on GRN, we did not include it in the final models. P-values of all linear regressions were corrected together for multiple testing using Benjamini-Hochberg correction with the *p.adjust* R-package.

To estimate the evolutionary rate of shifts in GRN, we calculated the absolute change in the mean GRN as standard deviations per generations, referred to as haldane, using historical gill raker data for gangfisch collected along a similar depth gradient in 1990 and gill raker data from this study in 2014 (see Supplementary methods for sampling details). Haldanes were calculated as follows: \[ \frac{(\ln(x_2)/sd_{\ln(x_2)}) - (\ln(x_1)/sd_{\ln(x_1)})}{t}, \] where \(\ln(x_1)\) and \(\ln(x_2)\) are means of natural log-transformed GRNs at each time point, \(t\) is the interval between sampling times in number of generations between, and \(sd_{\ln(x)}\) is the pooled standard deviation of \(\ln(x_2)\) and \(\ln(x_1)\). We used a generation time of three years, based on the age distribution in our dataset at spawning time.

**Genotyping, bioinformatic processing, and summary statistics.** DNA was extracted from fin clips and muscle tissue using the NucleoSpin Tissue kit (Macherey-Nagel) and double-digest restriction site associated DNA (ddRAD) libraries were prepared for 170 individuals using *MspI* and *PstI-HF* restriction enzymes, following the methods in Recknagel et al.
(2015) with modified adapters for Illumina sequencing platforms. Each library was sequenced on a single NextSeq500 lane (75bp paired-end sequencing) including ~10% Phi-X at Glasgow Polyomics. We checked the quality of raw reads for each library using FastQC v.0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

Stacks v.1.46 was used for demultiplexing, SNP calling, and filtering. First, we processed the raw reads and trimmed reads to 60bp using process_radtags. The denovo_map.pl pipeline was used for the de novo assembly of RAD loci. We used the rxstacks module for genotype correction and haplotype pruning using the default settings. The populations module was used for calling, extracting, and filtering genotypes using the filtering criteria as described in the supplementary methods. The filtering and conversion of datasets was performed using Stacks, vcftools v. 0.1.15, PLINK v.1.90 (www.cog-genomics.org/plink/1.9/) and PGDspider v.2.11.2.

Population structure and admixture. We used several approaches to resolve population structure, including all three subspecies, within Lake Constance. First, we used Admixture v.1.3.0 to detect the most likely number of genetic clusters, testing K=1 to K=12. Furthermore, we performed a PCA using the adegenet package in R. To improve visualisation of population structuring in the multidimensional space, we performed discriminant analyses of principal components using adegenet, clustering gangfisch individuals by site and depth. Genodive v.2.0b27 was used to calculate pairwise Fst among all subspecies and populations, with gangfisch clustered by sampling depth and site (10,000 permutations).

Evolutionary history of genetic divergence and introgression. We analysed the evolutionary history of divergence and introgression in Lake Constance whitefish using two
different approaches. First, we used $f_3$-statistics implemented in the `treemix v.1.13` function to test for introgressive hybridization in a three-population comparison. The dataset contained all gangfisch as one focal population, and weissfelchen and blaufelchen as two separate reference populations using the global SNP dataset.

Furthermore, we used a coalescence modelling approach implemented in `fastimcoal2 v.2.5.2.3` based on the information contained in multidimensional site frequency spectra (SFS) to determine the history of gene flow, divergence times, and timing and strength of introgression in whitefish from Lake Constance. Three-population SFSs were created using $\delta a \delta i$ v.1.6.3 based on a specifically filtered SNP dataset (N=22,196 SNPs). The minor folded SFS was used for the analysis, as no trinucleotide substitution matrix is available for salmonids and no outgroup species was sequenced for ancestral state reconstruction and correction. To determine absolute values for divergence times and other parameters, we corrected the number of monomorphic sites as outlined in. A mutation rate of $1 \times 10^{-8}$ was used. Based on the genetic population structure, we used a model in which blaufelchen split first from the ancestor of weissfelchen and gangfisch and a subsequent split of weissfelchen and gangfisch (Supplementary Fig. 4e). We compared a total of seven different demographic models with different histories and combinations of gene flow and introgression, ranging from strict-isolation to a complete isolation-with-migration model with or without introgression (see Supplementary methods and Supplementary Fig. 4e). We ran 30 iterations for each model, each consisting of 40 rounds of parameter estimation with 100,000 coalescence simulations. We inferred the best fitting model based on the AIC and ran it an additional 45 iterations and used the top 10 iterations (based on highest estimated maximum likelihood) for non-parametric bootstrap resampling (bootstrap R package) to estimate means and 95%-CI for each parameter. In addition to test the potential effect of selection on the demographic inference, we performed the same analysis using a SNP dataset without sites.
potentially under selection and differentiated among genetic clusters (see below) (N = 22,100 SNPs).

**Correlation of genetic ancestry and eco-morphology in gangfisch.** To assess how morphology correlates with genetic ancestry in gangfisch, we calculated spearman’s rank correlation coefficients for body shape PC2 and GRN with proportion of genetic ancestry inferred by Admixture. PC2 of body shape was used for this (and all subsequent analyses) as it depicts the main body shape variation along the depth gradient (Supplementary Fig. 2c). P-values were adjusted for multiple testing using a Benjamini-Hochberg correction using the `p.adjust` R-package.

**Genome-wide scans for selection.** To detect loci that are differentiated between genetic clusters, we combined two approaches to identify and narrow down candidate loci. First, we calculated the per site pairwise Fst and nucleotide diversity between and within gangfisch from without significant admixture (pure genetic clusters: genetic ancestry coefficient > 0.75 or <0.25 in the Admixture analysis; N=135) using vcftools. To determine whether the basis of eco-morphological variation is similar across subspecies, we compared genetic differentiation between benthic and pelagic gangfisch with that of pelagic blaufelchen and benthic-littoral weissfelchen. We estimated per site pairwise Fst between weissfelchen and blaufelchen, and calculated the spearman’s rank coefficient correlation of these Fst values with Fst values calculated between genetic clusters in gangfisch using the `cor.test` function in R. To analyse the distribution of differentiated loci across the genome, we anchored the anonymous RAD-loci to the available RAD-Seq based lake whitefish (*Coregonus clupeaformis*) linkage map using the MapComp software package and the previously
described approach\textsuperscript{67,68} using the Atlantic salmon reference genome (ICSASG\_v2)\textsuperscript{69}. The maximum distance for pairing was set to 2 Mb and conducted in 10 iterations.

As the Fst-based analysis ignores the more gradual nature of genetic differentiation along the depth gradient, we also used a (genetic) PCA based approach for detecting loci under selection that does not require the grouping of individuals. We used the \texttt{pcadapt} \textit{R} package\textsuperscript{70} to identify loci that differentiate along the first principal component, as this axis differentiates gangfisch from different spawning depths and therefore different phenotypes. We corrected for multiple testing using a false discovery rate cut-off of 1\% using the \texttt{qvalue} \textit{R}-package.

\textbf{Association analysis with eco-morphology.} To identify variants associated with GRN and body shape (PC2), we used three approaches: a linear mixed model (LMM) and a linear Bayesian sparse linear mixed model (BSLMM) implemented in \textit{GEMMA}\textsuperscript{71}, and a latent-factor linear mixed model approached implemented in \textit{LFMM} (N\textsubscript{inds}=153)\textsuperscript{72}. The analysis for GRN and body shape were performed separately but with the same settings.

We imputed missing genotypes using a random forest approach implemented in the \textit{radiator} \textit{R}-package, using 150 trees. We first estimated phenotype-genotype association for each SNP using a univariate linear mixed model and the Wald test, using $\alpha=0.0001$ ($-\log_{10}(\alpha) = 4$) as a significance threshold. Second, we used a linear BSLMM to estimate SNP effect sizes (posterior inclusion probability: PIP) on phenotype and further estimate hyperparameters quantitatively describing the genetic architecture underlying trait variation (see Supplementary methods for details). For both analyses in \textit{GEMMA}, we corrected for relatedness and population structure using a centred relatedness matrix, which we calculated using \textit{GEMMA}. Lastly, we used \textit{LFMM}, specifying two latent factors to correct for population structure, and averaged the results across 10 independent runs using default
settings. Candidate loci associated with body shape or GRN were detected as those SNPs with FDR < 0.01. *Post hoc* we checked for overlap of phenotype-associated loci across analyses.

**Annotation of candidate loci.** We identified genes associated with candidate loci (significantly diverged, associated with GRN or body shape, genetically diverged and under selection; N=177) by blasting RAD loci sequences against the Atlantic salmon reference genome (ICSASG_v2)\(^69\) using blastn in Blast\(^+\),\(^73\) and identifying genes containing candidate loci or within 50kb of a candidate locus. We only kept loci with an E-value below 1e-13 and the best (or equally good results if a locus blasted to different locations) blast results. We used the PANTHER (Protein Analysis Through Evolutionary Relationships) classification tool\(^74\) to identify functional pathways associated with candidate loci and overrepresented biological process gene ontology terms, using the human gene ontology set as background.

**Transcriptomics.** Total RNA was extracted from white muscle tissue from 27 individuals (three per site and depth) and RNA-seq libraries were prepared and sequenced by Glasgow Polyomics research facility (University of Glasgow) (see Supplementary methods for details). Individual cDNA libraries were synthesised for all 27 samples using the TruSeq Stranded mRNA Sample Preparation kit (Illumina, San Diego, CA), with a Poly-A selection step. Libraries were sequenced on the NextSeq 500 system (Illumina, San Diego, CA) using paired-end sequencing (75 bp from each end), at a sequencing depth of 25-35 million reads per library, yielding approximately 822 million reads in total.

Raw reads were processed prior to alignment, by removing adapter sequences with *Scythe* v0.9944 BETA and trimming low quality reads with *Trimmomatic* v0.36\(^75\). FastQC v0.11.2\(^51\) was used to assess read quality before and after pre-processing. The resulting reads
were aligned to a transcriptome assembly for European whitefish\textsuperscript{76}, consisting of 33,697 annotated transcripts. Reads were aligned using Bowtie2 (--all, --local, --no-mixed, --no-discordant)\textsuperscript{77}, and transcript abundance was quantified with eXpress v1.5.1\textsuperscript{78}. Count data were processed using DESeq2 v3.5\textsuperscript{79}, transcripts with <20 reads across all samples removed (yielding a final dataset of 31,872 transcripts), and count data log\textsubscript{2} transformed. Transcriptional profiles were visualised using a PCA on the full dataset using the plotPCA function in DESeq2 and plots were generated using the ggpplot2 v2.2.1.9000.

We used a linear mixed model implemented in the R-package variancePartition v3.6\textsuperscript{80} to identify genes associated with GRN and body shape (BS-PC2) as fixed variables, including site, depth and sex as random effects (see Supplementary methods for details). We retained genes for a subsequent network analysis based on the amount of expression variation explained by either GRN or BS-PC2, namely if the expression variation explained by GRN or BS-PC2 was above the 90\textsuperscript{th}-quantile of the variation distribution for all genes associated with the respective trait.

Subsequently, a Weighted Gene Co-Expression Network Analysis (WGCNA) approach\textsuperscript{81} using the WGCNA R package\textsuperscript{82} was used to identify co-expression modules associated with trophic divergence based on the pre-filtered gene set (see Supplementary methods for details). Network modules were defined using the dynamic treecut algorithm, with a minimum module size of 30 genes and a cut height of 0.981. The module eigengene distance threshold was set to 0.25 to merge highly similar modules.

To identify genetic pathways associated with trophic niche divergence in gangfisch, we performed functional enrichment analyses on co-expressed modules. Separate enrichment analyses were performed for each module, using the Over-Representation Analysis (ORA) function in PANTHER\textsuperscript{74}. Module genes were compared against the background set of 31,872 transcripts to determine over-representation of pathways. All functional analyses were
conducted using gene symbols from the annotated transcriptome assembly, and significance was determined using Fisher’s Exact test with Bonferroni correction (FDR < 0.05).

**Data availability.** The sequence datasets have been deposited in NCBI Sequence Read Archive with the BioProject accession code PRJNA497182 (corresponding to BioSample accessions SAMN10250325 to SAMN10250521). Phenotype and ecological data are available at the ‘Enlighten: Research Data’ repository of the University of Glasgow: http://dx.doi.org/10.5525/gla.researchdata.680
References.


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


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


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


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


Acknowledgements: We thank H. Thiele, M. Schmid, W. Kornberger and A. Sulger for assistance with specimen or data collection, D. Straile for providing background data, and P. Hirsch, H. Recknagel and A. Yurchenko for comments and advice. This work was funded by Marie Curie CIG (321999) to KRE, BBSRC Westbio DTP studentship awarded to MC with CEA and KRE (BB/J013854/1), ERASMUS+ (JB-G, KRE), Fisheries Society of the British Isles Research Grant (KRE, AJ), and AFF funding Univ. Konstanz to EY and JB-G.

Author contributions: JB-G and KRE designed the experiment. AJ, MC, KRE and JB-G collected data. MC generated and analysed eco-morphological and transcriptomic data. AJ generated and analysed genomic data, analysed eco-morphological and stable isotope data. RE analysed life history data. EY generated stable isotope data. CEA, JB-G and KRE supervised the project. AJ with MC, JB-G and KRE wrote the paper. All authors commented on and approved the final manuscript.

Author Information: Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Correspondence and requests for materials should be addressed to jasminca.behrmann@uni-konstanz.de and/or kathryn.elmer@glasgow.ac.uk
Figure legends:

Figure 1. Phenotypic and ecological diversity. a, Historical environmental and gill raker data during eutrophication and re-oligotrophication of Lake Constance from 1961 until 2015. The red and green curve show historical fluctuations in total phosphorus (annual mean; \( \mu \text{L}^{-1} \)) and the total zooplankton abundance ((individuals per m\(^2\) x 10\(^3\)/10\(^5\)). The bar plots show the annual catch composition of whitefish in Lake Constance by subspecies (in percent of total catch) over the same time period. The range in gill raker number for gangfisch (grey), blaufelchen (blue), sandfelchen (orange) and kilch (black) are shown for the years 1967, 1990 and 2014. Note that data for kilch and sandfelchen were only available for the year 1967, as they went extinct or were very rare following the eutrophication. Illustrations of gill raker arches next to the plot illustrate the two extremes (N\(_{\text{low}}=24; \ N_{\text{high}}=44\) of the gill raker spectrum in contemporary gangfisch. b, Boxplots [bar = median, box range = range between third and first quartile (interquartile range; IQR), whiskers = extend to furthest point (highest or lowest) no further than 1.5 times the IQR, points = outlier] showing the distribution gill raker number by subspecies, and by sampling site and depth for gangfisch (N = 309). Species and sampling sites: WF = weissfelchen, BF = blaufelchen; gangfisch from: Bod = Bodman, Egg = Egg and Imm = Immenstaad. e, Difference in the number of gill rakers (mean ± SEM) between gangfisch infected and non-infected with Diplostomum spp. (Wilcoxon rank sum test: \( P < 0.001; \ N = 269\)). d, Canonical variant plot showing the variation in body shape between whitefish, with individuals grouped by subspecies and gangfisch individuals further split by spawning depth. Wireframes depict body shape at outer most point of CV1 (40.6%). e-f, Correlation of gill raker number with carbon stable isotope signatures (e; linear model: \( F_{1,146}=13.35, \ P < 0.001 \)) and PC2 of body shape (f; linear model: \( F_{1,266}=16.76, \ P < 0.001 \)) in gangfisch. Grey areas depict the standard error around the regression line.
Figure 2. Rate of change in gill raker number and range. a, Distribution of gill rakers in gangfisch in 1990 and 2014 illustrated by a density plot and mean ± SD. The distribution of individual data points is shown by jittered points. b, Evolutionary rates of phenotypic change plotted in (absolute) haldanes, standard deviations of change per generation, for the shift in the mean gill raker number in gangfisch in Lake Constance from this study (orange, N=1) and previously published rates of phenotypic change for other salmonid species (red, N=132) as well as for different taxa (white, N=2224)⁶. Absolute values of haldanes and generations were log-transformed.
Figure 3. Evolutionary history of introgression. **a**, The $f_3$-statistic revealed significant signatures of introgressive hybridization in gangfisch (GF) from weissfelchen (WF) and blaufelchen (BF) (GF:BF,WF; $f_3 = -0.0011 \pm 0.00014$ SEM, $z$-score = -7.97). **b**, The most likely demographic model for the evolutionary history of whitefish in Lake Constance based on the full SNP dataset. Divergence times are in years (95%-CI) and the proportion of admixture (95%-CI) between subspecies is given above the solid black arrows. The proportion of admixture describes the proportion of the gene pool in gangfisch that was replaced by either WF or BF. Detailed parameter values for migration rates (grey arrows) and effective population sizes are given in Supplementary Table 3. The proportions of genetic ancestry for each individual by subspecies are given below.
**Figure 4. Genotype-phenotype associations and signatures of selection.**

a, Correlation between the genetic ancestry coefficient and number of gill rakers in gangfisch (Spearman test: $\rho = -0.6007; P < 0.001; N = 153$). Histograms and density lines show the parameter distributions. Individuals with an ancestry coefficient above 0.75 were defined as ‘benthic’, whereas individuals with an ancestry coefficient below 0.25 as ‘pelagic’, as these are more similar to weissfelchen and blaufelchen, respectively. b, Frequency of the pelagic alleles (mean ± SD) at loci strongly associated with gill raker number ($N_{\text{loci}} = 6$), by subspecies and spawning depth in gangfisch. We defined the allele with a higher frequency in blaufelchen compared to weissfelchen as the ‘pelagic’ allele. c, Correlation between genotype-association with gill raker number (LFMM-GRN) and with body shape PC2 (LFMM-BS) across all SNPs ($N = 12,976$). The association results from LFMM are given by the $z$-score. Loci associated with GRN and BS in LFMM are highlighted red ($N = 6$). Loci only associated with GRN are highlighted in yellow ($N = 93$), whereas loci only associated with BS are highlighted in cyan ($N = 48$). Locus 56834, which is strongly associated with GRN and BS is annotated. d and e, Gill raker number and body shape (PC2) for gangfisch that are homozygous or heterozygous for the benthic (B) or pelagic (P) alleles on locus 56834 (ANOVA-GRN: $F_{2,126} = 29.82$, $P < 0.001$; ANOVA-BS: $F_{2,127} = 7.97$; $P < 0.001$). The dotplot shows the distribution of values and the mean ± SD. f, Correlation of association with gill raker number (LMM) and signal of selection (PCAdapt) for 12,976 SNPs in gangfisch. Dashed lines show the applied significance thresholds for each analysis. The size of dots indicates the effect size of each SNP on the variation in gill raker number and the colour gradient shows the degree of genetic differentiation (standardised $Fst$: $ZFst$) between benthic and pelagic gangfisch. g, Manhattan plot showing the distribution of genetic differentiation between benthic and pelagic gangfisch along the lake whitefish linkage map for a subset of 3,350 loci. Significantly differentiated SNPs ($ZFst > 4$, dashed line) are highlighted in red.
**Figure 5. Functional gene expression variation in gangfisch.** a, Transcriptional profiles of gangfisch collected from all three sampling sites (Bodman represented as circles, Egg as triangles and Immenstaad as squares), and spawning depths in Lake Constance analysed using a principal component analysis based on the complete set of filtered genes (>20 read counts across all samples; N=31,872 genes). Gangfisch are coloured based on their number of gill rakers, using a gradient scale (low GRN ‘benthic’ = red, high GRN ‘pelagic’ = blue). Each point represents an individual (N=27, three per depth, per site), and the centroids for each depth group are plotted and labelled. b, Correlation between WGCNA module eigengenes and measured traits (site, depth, sex, genetic ancestry coefficient (GAC), gill raker number (GRN) and body shape (BS-PC2); N=27). Each row corresponds to a module (identified on the left side by its colour; colours are assigned arbitrarily by WGCNA), and each column corresponds to a trait. The number of genes present within each module is given in parentheses below the module colour label. Pearson’s correlation coefficients for module-trait relationships are given in each cell, with the corresponding p-values in parenthesis. All p-values are corrected for multiple testing using Benjamini-Hochberg and significant correlations (FDR < 0.05) are highlighted in bold. Cells are coloured based on their correlation (see legend). No significant module associations were identified for site, depth or sex. c, Associations of module eigengene expression for the brown module with gill raker number (GRN) and genetic ancestry coefficient (GAC). d, Associations of module eigengene expression for the blue module with body shape (BS-PC2) and genetic ancestry coefficient (GAC). In panels c and d, the grey area shows the standard error around the regression line. Individuals are coloured by genetic ancestry (ancestry coefficient above > 0.75 for respective clusters based on SNP-based *Admixture* analysis). R²-values for significant module-trait relationships are given in the figures.