

Rapid niche expansion by selection on functional genomic variation after ecosystem recovery

Arne Jacobs^{1,4}, Madeleine Carruthers^{1,4}, Reiner Eckmann², Elizabeth Yohannes², Colin E. Adams^{1,3}, Jasminca Behrmann-Godel^{2*} and Kathryn R. Elmer^{1*}

It is well recognized 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 ten generations, gangfisch evolved a greater range of gill raker numbers (GRNs) to utilize a broader ecological niche. A sparse genetic architecture underlies this variation in GRN. Several co-expressed gene modules and genes showing signals of positive selection were associated with GRN 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^{1,2}. Environmental degradation caused by human disturbance can be extremely challenging for organisms, resulting in rapid evolutionary and ecosystem changes³, interbreeding and hybridization, extirpation, or even extinction^{2,4–6}. For example, hybridization across closely related species can result in a loss of species richness ('speciation reversal') and associated functional diversity in adaptive traits^{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^{8,9}. Theoretical models have predicted that diversity can recover but only under particular genetic, ecological and evolutionary circumstances¹⁰. These models suggest that the probability of functional diversity re-emerging from a collapsed species group (for example, a '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^{10,11}. For example, it has been shown that fish often mate with individuals that are similar^{12,13}; thus, if functional traits are linked to, for example, body shape, divergence in a functional trait can lead to morphological divergence and further on to genetic divergence via trait-based assortative mating¹⁰. 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-twentieth 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 (GRNs) by 14% across lakes⁴. For example, in one of the largest lakes, Lake Constance, the benthic specialists (kilch and sandfelchen) went extinct or collapsed, with extensive hybridization among these and the remaining pelagic (blaufelchen and gangfisch) and littoral–benthic (weissfelchen) subspecies⁴. This depleted functional diversity is evidenced as a 28% reduction in the range of GRNs across whitefish species^{4,14} (Fig. 1a and Supplementary Table 1). As a result, individuals with low GRNs 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 state¹⁶ (Fig. 1a). This provides an excellent system for the study of biodiversity recovery after ecosystem remediation, as long-term data are available for the fish and lake conditions, natural recolonization from other whitefish populations is unlikely, and, in contrast with other alpine lakes, Constance has not been subjected to external stocking^{15,17}. Here, we explore 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 with 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 recovery¹². Research proposed that gangfisch may be ecologically expanding because spawning depth is strongly associated with trophic ecology in European whitefish^{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,

¹Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. ²Limnological Institute, University of Konstanz, Konstanz, Germany. ³Scottish Centre for Ecology and the Natural Environment, University of Glasgow, Glasgow, UK. ⁴These authors contributed equally: Arne Jacobs, Madeleine Carruthers. *e-mail: jasminca.behrmann@uni-konstanz.de; kathryn.elmer@glasgow.ac.uk

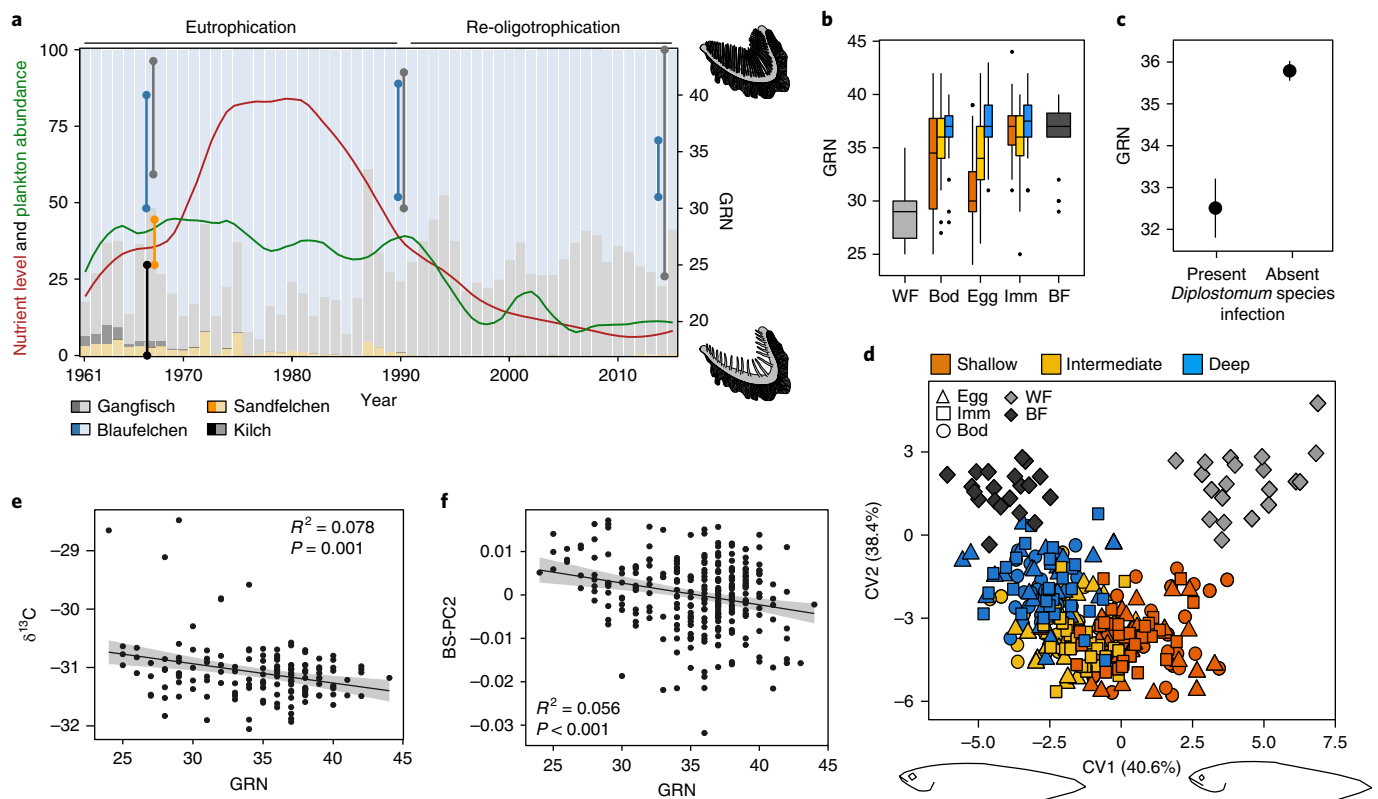


Fig. 1 | Phenotypic and ecological diversity. **a**, Historical environmental and gill raker data during eutrophication and re-oligotrophication of Lake Constance from 1961 until 2015. The red and green curves show historical fluctuations in total phosphorus (annual mean; μL^{-1}) and total zooplankton abundance ((individuals $\text{m}^{-2} \times 10^4$)/ 10^5), respectively. The bar plots show the annual catch composition of whitefish in Lake Constance by subspecies (as a percentage of the total catch) over the same time period. The ranges in GRNs for gangfish (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 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 gangfish. **b**, Boxplots (bar = median; box range = range between third and first quartile (interquartile range); whiskers = extend to furthest point (highest or lowest) no further than 1.5 times the interquartile range; points = outliers) showing the distribution of GRNs by subspecies, and by sampling site and depth for gangfish ($n = 309$). BF, blaufelchen; Bod, gangfish from Bodman; Egg, gangfish from Egg; Imm, gangfish from Immenstaad; WF, weissfelchen. **c**, Difference between the GRNs (mean \pm s.e.m.) of gangfish infected and not infected with *Diplostomum* species (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 gangfish individuals further split by spawning depth. Wireframes depict body shapes at the outermost points of canonical variant 1 (CV1; 40.6%). **e, f**, Correlation of GRN with carbon stable isotope signatures (**e**; linear model: $F_{1,146} = 13.35$, $P < 0.001$) and BS-PC2 (**f**; linear model: $F_{1,266} = 16.76$, $P < 0.001$) in gangfish. Grey areas depict the standard error around the regression line.

dense and long rakers are used to filter-feed zooplankton, while few, sparse and short rakers are associated with benthos feeding, and GRN is well documented to be genetically based^{4,19,20}. Indeed, in line with expectations, we found 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 ($\text{GRN}_{\text{shallow}} = 33.9 \pm 4.7$ s.e.m.) than deep-spawning individuals ($\text{GRN}_{\text{deep}} = 37.0 \pm 2.8$ s.e.m.). 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: effect size (R^2) = 0.028, $F_{1,267} = 8.76$, $P = 0.0034$).

Individuals spawning at different depths also differed in body shape (multivariate analysis of covariance, Pillai's trace_{depth} = 0.722, d.f. = (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 with deep-spawning individuals, which had more slender bodies and terminal mouths (Fig. 1d and Supplementary Fig. 2b,c). Body shape divergence with depth was mostly explained by principal component 2 (PC2; hereafter,

referred to as BS-PC2) (Supplementary Fig. 2c) and variation in body shape was correlated with 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 gangfish. Additionally, shallow-spawning gangfish were longer (standard length_{shallow} = 28.6 ± 0.7 cm versus standard length_{deep} = 25.7 ± 0.6 cm) and heavier (body weight_{shallow} = 327.5 ± 12.9 g versus body weight_{deep} = 225.2 ± 5.6 g) than deep-spawning individuals (standard length: $F_{2,261} = 34.69$, $P < 0.001$; body weight: $F_{2,261} = 38.49$, $P < 0.001$) (Supplementary Fig. 2d), providing evidence that multiple phenotypic traits contribute to the gradient along the depth axis. Although divergence in body shape and standard length were correlated with GRN (Fig. 1 and 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 (multivariate analysis of covariance, Pillai's trace_{site} = 1.243, d.f. = (2,132), $P < 0.001$; Pillai's trace_{site×depth} = 1.054,

d.f. = (2,132), $P < 0.001$). Individuals from Immenstaad showed, in general, less divergence along the depth gradient (Supplementary Results). Compared with the other extant whitefish subspecies, gangfisch were intermediate in body shape to weissfelchen and blauefelchen, with shallow-spawning gangfisch being more similar to benthic–littoral weissfelchen and deep-spawning gangfisch being similar to pelagic blauefelchen (Fig. 1d).

To test whether phenotypic diversity is correlated with trophic and habitat diversity in gangfisch, and therefore reflects functional utility, we assessed the correlation of GRN 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 and Supplementary Figs. 2 and 3). Specifically, GRN was correlated with carbon stable isotope signatures (Fig. 1e; $R^2 = 0.07751$, $P < 0.001$, d.f. = 147), indicating a more littoral–benthic diet in individuals with fewer gill rakers²¹. 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 use²². 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 depth gradient^{19,23}. Thus, we focused on variation in the key functional trait of GRN 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 GRNs between the samples collected in 2014 for this study and samples collected in 1990²⁴. 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 (Figs. 1a and 2a; ranges in GRNs: 1990 = 30–42 (ref. ²⁴); 2014 = 24–44). The range in GRNs increased drastically over this 24-year period (corresponding to about 8 whitefish generations), with a notable decrease in the minimum number from 30 to 24. The mean GRN decreased by 0.3423 s.d. (haldane numerator) over this period, equivalent to an evolutionary rate of -0.0428 haldanes (Fig. 2)²⁵. This is comparable to rates of phenotypic evolution in several examples of rapid evolution, such as changes in GRN in marine–freshwater sticklebacks or hindlimb length in anole lizards, and lies within the top 12% of reported rates of rapid phenotypic evolution⁶ (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 and Supplementary Table 1).

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 blauefelchen and weissfelchen at extremes, and gangfisch showing an intermediate gradient of ancestry (Fig. 3b and Supplementary Fig. 4a,b), suggesting strong historic admixture and recent segregation in gangfisch⁴. Indeed, we detected multiple signals consistent with historical genetic introgression from blauefelchen and weissfelchen into gangfisch, including the intermediacy of gangfisch in genetic ordination plots (Supplementary Fig. 4a–d), divergence of gangfisch along the eigenvector distinguishing subspecies (Supplementary Fig. 4c,d) and

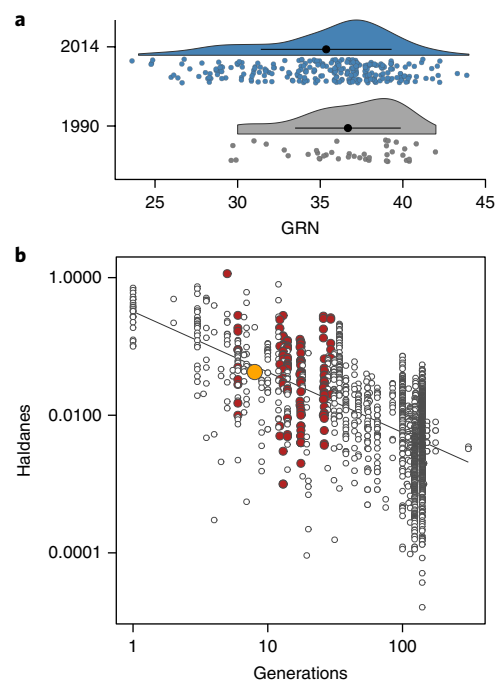


Fig. 2 | Rate of change in GRN and GRN ranges. **a**, Distribution of gill rakers in gangfisch in 1990 and 2014 illustrated by density plots showing means \pm s.d. The distribution of individual data points is shown by jittered points. **b**, Evolutionary rates of phenotypic change plotted in (absolute) haldanes (standard deviations of change per generation) for the shift in the mean GRN 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, $n = 2,224$)⁶. Absolute values of haldanes and generations were log-transformed.

support from introgression tests (Fig. 3a; f_3 statistic = -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 (that is, 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 blauefelchen (mean admixture proportion = 0.78) and weissfelchen (mean admixture proportion = 0.45) into gangfisch (Fig. 3b and Supplementary Table 3), leading to the formation of a gangfisch hybrid swarm⁴. 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 that introgressed into gangfisch from other sympatric whitefish species are associated with its diversification. Thus, introgressive hybridization across species in Lake Constance, including those that are now extinct⁴, potentially provided the genetic material needed for the rapid eco-morphological diversification in gangfisch after ecosystem recovery²⁶.

Genetic basis of functional and eco-morphological traits. To assess the genetic underpinnings of eco-morphological changes, we first tested the correlation between functional diversity and genetic ancestry. GRN in gangfisch was highly correlated with genetic ancestry (Fig. 4a; Spearman's rank correlation coefficient, $r_{\text{GRN}} = -0.6007$, $P < 0.001$), suggesting genetic isolation by ecological adaptation²⁷. Sparsely rakered gangfisch (benthic gangfisch) were,

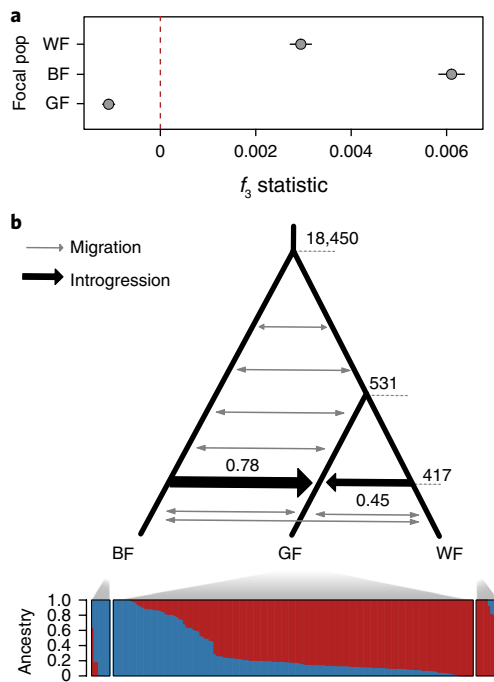


Fig. 3 | Evolutionary history of introgression. **a**, The f_3 statistic revealed significant signatures of introgressive hybridization from weissfelchen (WF) and blaufelchen (BF) into gangfisch (GF) ($f_3 = -0.0011 \pm 0.00014$ s.e.m, z score = -7.97). **b**, Most likely demographic model for the evolutionary history of whitefish in Lake Constance based on the full SNP dataset. Mean divergence times are shown to the right in years, and the mean proportions of admixture between subspecies are given above or below the solid black arrows (see Supplementary Table 3 for 95% CI). The proportion of admixture describes the proportion of the gene pool in GF 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.

in general, genetically more similar to the benthic–littoral weissfelchen, whereas densely rakered gangfisch (pelagic gangfisch) were genetically closer to pelagic blaufelchen (Fig. 4a and 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); here, $PVE_{GRN} = 0.9991$) being explained by a small expected number (predicted to be 91 (68,117) mean (upper and lower bound of 95% credible interval)) of large-effect loci (phenotypic variance explained by large-effect loci (PGE); here, $PGE_{GRN} = 0.9718$) (Supplementary Table 5). In comparison, we found that body shape—a highly polygenic trait in fishes²⁸ that was also correlated with genetic ancestry ($r_{BS-PC2} = 0.334$, $P < 0.001$)—had a more polygenic architecture, with only a small proportion of PVE 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 PVE explained probably 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 ($PVE = 0.947$; $PGE = 0.775$)²⁹ and dorsal brightness in deer mice ($PVE = 0.69$; $PGE = 0.83$)³⁰. 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 with body shape (97 versus 23%) is particularly informative and highlights the different architectures underlying these traits. Such a simple genetic architecture may facilitate the rapid expansion in GRN¹¹.

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 and Supplementary Table 6). All SNPs identified using GEMMA were also identified by the LFMM. Individuals homozygous for these shared loci were on opposite ends of the gill raker spectrum, and heterozygous individuals were intermediate (Fig. 4d and Supplementary Fig. 5a). Blaufelchen and weissfelchen were at either extreme of ‘pelagic’ allele frequencies (the allele associated with higher GRN), which in gangfisch varied with spawning depth (Fig. 4b). Thus, GRN is probably controlled by a common genetic basis across species. Furthermore, using the same approaches, we identified 56 SNPs (two using GEMMA and 54 by the LFMM) associated with body shape in gangfisch (Fig. 4c and 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,e and Supplementary Table 6).

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

Genetic differentiation. To understand whether GRN expansion was driven by natural selection, we tested patterns of differentiation and diversity across the genome. Using pcadapt, we found signatures of selection at 19 SNPs, including 2 of the 6 shared SNPs that had been associated with GRN (Fig. 4f, Supplementary Fig. 5b and Supplementary Table 6). Of the 99 SNPs associated with GRN in both genome-wide association analyses, including those identified with pcadapt, 40 were genetically differentiated (Z -transformed standardized degree of genetic variation (ZF_{st}) > 4 ; Supplementary Fig. 6a) between benthic and pelagic genetic clusters in gangfisch (Fig. 4a) and showed reduced diversity in 1 of the 2 genetic clusters (Supplementary Fig. 6b,c and Supplementary Table 6). This is reflective of a response to positive selection³¹, as genetic differentiation is less affected by genomic features, such as low recombination regions, at early stages of divergence^{32,33}, and supports a role of selection driving diversification in GRN. Only 3 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 ($r = 0.11$, $P < 0.001$; Methods), with 20 out of 165 SNPs differentiated in gangfisch also differentiated ($ZF_{st} > 4$) between blaufelchen and weissfelchen (Supplementary Fig. 6f). This suggests that genetically differentiated regions, including those associated with phenotypic divergence (for example, 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 groups 8 (65.239 cM) and 21 (9.24 cM)) overlapped with previously

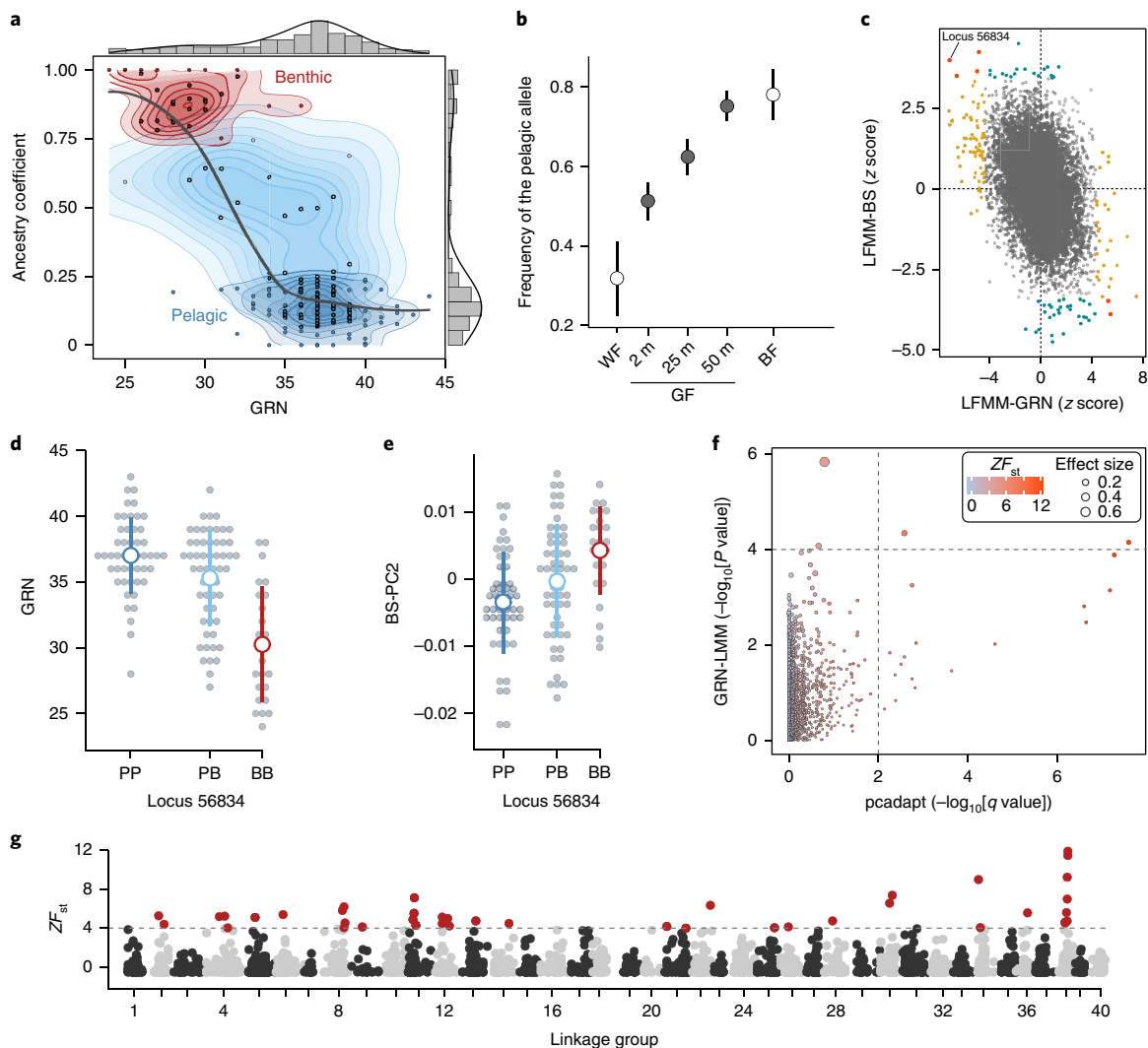


Fig. 4 | Genotype-phenotype associations and signatures of selection. **a**, Correlation between the GAC and GRN in gangfish ($r = -0.6007$, $P < 0.001$, $n = 153$). Histograms and density lines show the parameter distributions. Individuals with an ancestry coefficient above 0.75 were defined as ‘benthic’ and individuals with an ancestry coefficient below 0.25 were defined as ‘pelagic’, as these are more similar to weissfelchen and blauefelchen, respectively. **b**, Frequencies of pelagic alleles (means \pm s.d.) at loci strongly associated with GRN ($n_{\text{loci}} = 6$), by subspecies (weissfelchen (WF) or blauefelchen (BF)) and spawning depth in gangfish (GF). We defined the allele with a higher frequency in BF compared with WF as the ‘pelagic’ allele. **c**, Correlation between genotype association with GRN as determined by the LFMM (LFMM-GRN) and genotype association with BS-PC2 as determined by the LFMM (LFMM-BS) across all SNPs ($n = 12,976$). The association results from the LFMM are given by z scores. Loci associated with GRN and body shape in the LFMM are highlighted in red ($n = 6$). Loci only associated with GRN are highlighted in yellow ($n = 93$), whereas loci only associated with body shape are highlighted in cyan ($n = 48$). Locus 56834, which is strongly associated with GRN and body shape, is labelled. **d, e**, GRN (**d**) and BS-PC2 (**e**) for GF that are homozygous or heterozygous for the benthic (B) or pelagic (P) alleles on locus 56834 (analysis of variance for GRN: $F_{2,127} = 29.82$, $P < 0.001$; analysis of variance for body shape: $F_{2,127} = 7.97$, $P < 0.001$). Dot plots show the distribution of values and means \pm s.d. **f**, Correlation between the linear mixed model for GRN (GRN-LMM) and signal of selection (pcadapt) for 12,976 SNPs in GF. Dashed lines show the applied significance thresholds for each analysis. The sizes of the dots indicate the effect size of each SNP on the variation in GRN, and the colour gradient shows ZF_{st} between benthic and pelagic GF. **g**, Manhattan plot showing the distribution of genetic differentiation between benthic and pelagic GF along the lake whitefish linkage map for a subset of 3,350 loci. Significantly differentiated SNPs ($ZF_{\text{st}} > 4$; dashed line) are highlighted in red.

identified quantitative trait loci for variation in GRN and differential depth selection in lake whitefish, respectively^{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 GRNs was driven by natural selection.

Gene regulatory and functional basis of rapid divergence. To investigate the functional molecular bases of trophic niche

expansion in the ecological context, we conducted a transcriptome-wide analysis of gene expression in gangfish. While we observed a degree of divergence by depth across 31,872 genes (principal component analysis (PCA)), the separation was strongest by GRN (Fig. 5a and Supplementary Fig. 7a). To quantitatively assess the relative contributions of site, depth, sex, GRN and body shape on gene expression divergence, we used a linear mixed model. GRN explained the greatest proportion of overall gene expression variation with an average of 5.7%, followed by body shape (explained 2.9%). Only negligible amounts of variation were explained by site

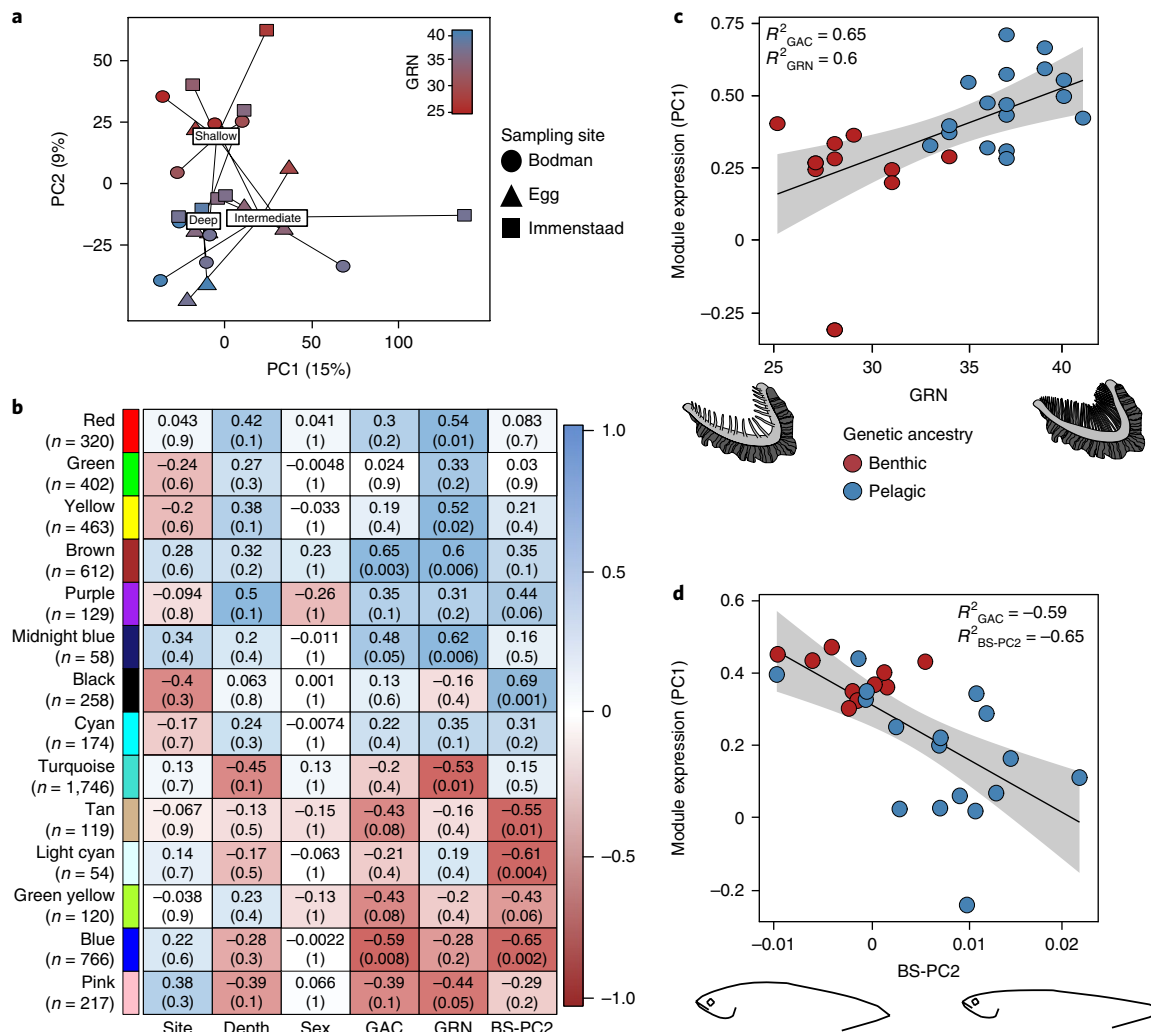


Fig. 5 | Functional gene expression variation in gangfish. **a**, Transcriptional profiles of gangfish collected from all three sampling sites (Bodman, circles; Egg, triangles; and Immenstaad, squares) and spawning depths in Lake Constance analysed using a PCA based on the complete set of filtered genes (>20 read counts across all samples; $n = 31,872$ genes). Gangfish are coloured based on their GRN, using a gradient scale (low GRN/'benthic' = red; high GRN/'pelagic' = blue). Each point represents an individual ($n = 27$; that is, 3 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, GAC, GRN and 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 correction. 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 GRN and GAC. **d**, Associations of module eigengene expression for the blue module with BS-PC2 and GAC. In **c** and **d**, grey areas show s.e.m. around the regression line. Individuals are coloured by genetic ancestry (ancestry coefficient >0.75 for respective clusters based on SNP-based ADMIXTURE analysis). R^2 values for significant module-trait relationships are provided.

(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^{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 gangfish.

A single co-expression network was constructed based on 5,438 candidate genes (those within the 90th percentile of expression variation explained by GRN and/or body shape; Supplementary Fig. 7c), which clustered into 14 modules (Fig. 5b). Five modules were significantly correlated with GRN, one of which (brown) was also correlated with genetic ancestry (Fig. 5c and Supplementary

Fig. 7e). A further four modules were significantly correlated with body shape, with one module (blue) also correlated with genetic ancestry (Fig. 5d and 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 (that is, pelagic adaptation) showed significant over-representation of pathways (as assessed by protein analysis through evolutionary relationships (PANTHER)) involved in immune response, metabolism, vasodilation and muscle contraction, including integrin signalling (P00042 (PANTHER accession); $P = 4.12 \times 10^{-3}$), metabotropic glutamate receptor (P00041; $P = 4.56 \times 10^{-2}$) and haem biosynthesis (P02746; $P = 4.43 \times 10^{-2}$) (Supplementary Table 7). In contrast, lower GRN and deeper head and body shape (that is, benthic adaptation) showed

significant over-representation of pathways involved in growth and morphogenesis, including transforming growth factor- β (TGF- β) signalling (P00052; $P=1.10 \times 10^{-2}$), integrin signalling (P00034; $P=5.96 \times 10^{-6}$), and bone morphogenetic protein and activin signalling (P06211; $P=4.41 \times 10^{-2}$) (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 species^{38–40}, indicating a crucial and conserved role in niche expansion and ecological adaptation.

Of these 39 expression-associated pathways, 9 were also found to be associated with candidate loci putatively under selection and/or associated with GRN and body shape ($n=179$ genes), including the TGF- β receptor signalling (P00052) and integrin signalling (P00034) pathways (Supplementary Table 8). In particular, the TGF- β pathway has previously been associated with benthic–pelagic divergence in other salmonids⁴¹, as well as trophic craniofacial development and diversification (for example, in cichlids and zebrafish)⁴². 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 gangfish is driven by an environmental (that is, plastic response) versus 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 (GAC), and the overlap with candidate loci pathways, strongly suggest that expression divergence is, at least in part, associated with genomic divergence.

Overall, we conclude that the rapid evolution of functional diversity in gangfish is based on functional molecular changes that play a recognized role in development and ecological specialization 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 gangfish diversified in a functional trait (GRN) 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 recognized loss of species richness caused by eutrophication^{3,12}. Second, we showed that this phenotypic change has a sparse underlying genetic architecture and that genetic and regulatory changes in recognized adaptive pathways are significantly associated with GRN. The phenotypic and genetic correlation of functional diversity (GRN) 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 loci⁴. The introgression of alleles associated with low GRNs from extinct benthic species (kilch or sandfelchen) could have contributed to the emergence of lower GRNs in gangfish. Overall, our results indicate that ecological diversity can start to re-emerge rapidly following species collapse with hybridization if few adaptive alleles with large effects underlie functional traits, eco-morphology and potentially reproductive isolation¹¹. Thus, our findings are consistent with the hypothesis that, while ecosystem remediation can have almost immediate benefits for biodiversity, the potential for recovery is probably contingent on genetic architecture, ecological context and evolutionary history.

Methods

Sampling of European whitefish. Gangfish (*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 and Supplementary Table 2). Three gill nets with different mesh sizes (32, 42 and 44 mm knot-to-knot mesh sizes) 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 that fish did not die before the 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 5 min of euthanasia by cutting a 1 cm-thick area of white muscle from above the lateral line, posterior to the operculum, and stored in RNAlater. A total of 30 individuals per site and depth, including fish sampled for transcriptomics, were selected for phenotypic analysis and tissue collection for genetic analysis ($n=270$). We sampled 20 individuals from each of the other 2 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 and 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 milligram and millimetre, respectively). They were then photographed on their left side for morphometric analyses (see Supplementary Methods for details). Scales were taken from underneath the ventral fins for ageing 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 0.01 mm.

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

In addition to stable isotope signatures, we analysed the presence of *Diplostomum* species (eye fluke parasites), as more long-term indicators of littoral–benthic habitat use of gangfish (see Supplementary Methods and ref. 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 or 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 ($n=269$ gangfish; $n=20$ weissfelchen; and $n=20$ blaufelchen). Landmarks were chosen based on previous studies in European whitefish^{15,44} (Supplementary Fig. 2a) and digitized using tpsDig 2.30 (ref. ⁴⁵). Before shape analyses, we used the ‘unbend’ function in tpsUtil version 1.74 to control for the effect of body arching on standing and semi-landmark positions. Procrustes superimposition was conducted in MorphoJ version 1.06b (ref. ⁴⁶) to standardize landmark configurations. Procrustes coordinates were subjected to a multivariate, pooled within-group regression against the log-transformed centroid size to remove shape change associated with body size⁴⁷. Residuals derived from the regression analysis were used as shape variables for all subsequent analyses. To visualize morphological variation, we conducted a canonical variance analysis, 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 PCA on gangfish from all depths and sites. We first performed a multivariate analysis using the complete set of principal components ($n=66$) and multivariate analysis of variance. Furthermore, we performed univariate three-way analyses of variance on the first three principal components (explaining >68% of the 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 ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) and trophic morphology (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 GRN between individuals that were infected or not infected by *Diplostomum* species parasites (an indicator of long-term littoral–benthic habitat use²²), using a Wilcoxon rank sum test in R.

To estimate the correlation of a range of morphological traits (major axes of body shape (principal components 1, 2 and 3), standard length and back-calculated total length at ages 1, 2 and 3, and size-corrected gill raker length), and because we were interested in the individual correlations with GRN, we performed individual linear regressions for each morphological trait, with GRN as the response variable, using the `lm` function in R (for example, $\text{GRN} \sim \text{BS-PC2}$). 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–Hochberg³⁸ 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 generation, referred to as haldanes^{25,49}, using historical gill raker data for gangfish collected along a similar depth gradient in 1990²⁴ and gill raker data from this study in 2014 (see Supplementary Methods for sampling details). Haldanes were calculated as follows: $((\ln[x_2]/s.d._{\ln[x_2]} - \ln[x_1]/s.d._{\ln[x_1]})/t)$, 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 (number of generations) and *s.d.* _{$\ln[x]$} is the pooled standard deviation of $\ln[x_1]$ and $\ln[x_2]$ ^{6,25,49}. 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 libraries were prepared for 170 individuals using *MspI* and *PstI*-*HF* restriction enzymes, following the methods in ref. ⁵⁰ with modified adapters for Illumina sequencing platforms. Each library was sequenced on a single NextSeq 500 lane (75-base pair (bp) paired-end sequencing) including ~10% PhiX at Glasgow Polyomics. We checked the quality of raw reads for each library using FastQC version 0.11.3 (ref. ⁵¹).

Stacks version 1.46 (ref. ⁵²) was used for demultiplexing, SNP calling and filtering. First, we processed the raw reads and trimmed reads to 60 bp using `process_radtags`. The `denovo_map.pl` pipeline was used for the de novo assembly of restriction site associated DNA (RAD) loci. We used the `rxstacks` module for genotype correction and haplotype pruning, applying 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 version 0.1.15 (ref. ⁵³), PLINK version 1.90 (ref. ⁵⁴ and www.cog-genomics.org/plink/1.9/) and PGDSpider version 2.11.2 (ref. ⁵⁵).

Population structure and admixture. We used several approaches to resolve population structure, including all three subspecies, within Lake Constance. First, we used ADMIXTURE version 1.3.0 (ref. ⁵⁶) 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^{57,58}. To improve visualization of population structuring in the multidimensional space, we performed discriminant analyses of principal components using adegenet, clustering gangfish individuals by site and depth. GenoDive version 2.0b27 (ref. ⁵⁹) was used to calculate the pairwise standardized degree of genetic variation (F_{ST})⁶⁰ among all subspecies and populations, with gangfish 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 version 1.13 (ref. ⁶¹) threepop function to test for introgressive hybridization in a three-population comparison. The dataset contained all gangfish 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 fastsimcoal2 version 2.5.2.3 (ref. ⁶²) based on the information contained in multidimensional site frequency spectra to determine the history of gene flow, divergence times, and timing and strength of introgression in whitefish from Lake Constance. Three-population site frequency spectra were created using `δaδi` version 1.6.3 (ref. ⁶³) based on a specifically filtered SNP dataset ($n = 22,196$ SNPs). The minor folded site frequency spectrum 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 ref. ⁶⁴. A mutation rate of 1×10^{-8} was used⁶⁵. Based on the genetic population structure, we used a model in which blaufelchen split first from the ancestor of weissfelchen and gangfish, followed by a subsequent split of weissfelchen and gangfish (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 Akaike information criterion⁶², ran it for 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% CIs for each parameter⁶⁶. In addition to testing 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 gangfish. To assess how morphology correlates with genetic ancestry in gangfish, we calculated values of *r* for BS-PC2 and GRN with the proportion of genetic ancestry inferred by ADMIXTURE. BS-PC2 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 Benjamini–Hochberg correction with the `p.adjust` R package.

Genome-wide scans for selection. To detect loci that were differentiated between genetic clusters, we combined two approaches to identify and narrow down candidate loci. First, we calculated the per-site pairwise F_{ST} ⁶⁰ and nucleotide diversity between and within gangfish without significant admixture (pure genetic clusters: $GAC > 0.75$ 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 gangfish with that of pelagic blaufelchen and benthic-littoral weissfelchen. We estimated per-site pairwise F_{ST} between weissfelchen and blaufelchen, and calculated the *r* values of these F_{ST} values, with F_{ST} values calculated between genetic clusters in gangfish 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-sequencing-based lake whitefish (*Coregonus clupeaformis*) linkage map³⁴ using the MapComp software package⁶⁷ and the previously described approach^{67,68} using the Atlantic salmon reference genome (ICSASG_v2)⁶⁹. The maximum distance for pairing was set to two megabases and conducted in ten iterations.

As the F_{ST} -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 `pcadapt` R package⁷⁰ to identify loci that differentiate along the first principal component, as this axis differentiates gangfish from different spawning depths, and therefore different phenotypes. We corrected for multiple testing using a false discovery rate cut-off of 1% using the `qvalue` R package.

Association analysis with eco-morphology. To identify variants associated with GRN and BS-PC2, we used three approaches: a linear mixed model and a Bayesian sparse linear mixed model implemented in GEMMA⁷¹, and a LFMM approach implemented in LFMM ($n = 153$ individuals)⁷². The analyses for GRN and body shape were performed separately but using the same settings.

We imputed missing genotypes using a random forest approach implemented in the radiator R package, using 150 trees. We first estimated phenotype–genotype association for each SNP using a univariate linear mixed model and the Wald test, with $\alpha = 0.0001$ ($-\log_{10}(\alpha) = 4$) as a significance threshold. Second, we used a Bayesian sparse linear mixed model to estimate SNP effect sizes (posterior inclusion probability) on phenotype and further estimate hyperparameters quantitatively describing the genetic architecture underlying trait variation (see Supplementary Methods for details). For both analyses in GEMMA, we corrected for relatedness and population structure using a centred relatedness matrix, which we calculated using GEMMA. Lastly, we used LFMM, specifying two latent factors to correct for population structure, and averaged the results across ten independent runs using default settings. Candidate loci associated with body shape or GRN were detected as SNPs with a false discovery rate (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 50 kilobases of a candidate locus. We only kept loci with an *E* value below 1×10^{-13} and the best (or equally good if a locus blasted to different locations) blast results. We used the PANTHER classification tool⁷⁴ to identify functional pathways associated with candidate loci and over-represented 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 sequencing libraries were prepared and sequenced at the Glasgow Polyomics research facility (University of Glasgow)

(see Supplementary Methods for details). Individual complementary DNA libraries were synthesized for all 27 samples using the TruSeq Stranded mRNA Sample Prep Kit (Illumina), with a poly-A selection step. Libraries were sequenced on the NextSeq 500 system (Illumina) 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 before alignment by removing adapter sequences with Scythe version 0.9944 BETA and trimming low-quality reads with Trimmomatic version 0.36 (ref. ⁷³). FastQC version 0.11.2 (ref. ³¹) was used to assess read quality before and after pre-processing. The resulting reads were aligned to a transcriptome assembly for European whitefish⁷⁶, consisting of 33,697 annotated transcripts. Reads were aligned using Bowtie2 (--all, --local, --no-mixed, --no-discordant)⁷⁷, and transcript abundance was quantified with eXpress version 1.5.1 (ref. ⁷⁸). Count data were processed using DESeq2 version 3.5 (ref. ⁷⁹), transcripts with <20 reads across all samples were removed (yielding a final dataset of 31,872 transcripts) and count data were log₂ transformed. Transcriptional profiles were visualized by PCA on the full dataset using the plotPCA function in DESeq2, and plots were generated using ggplot2 version 2.2.1.9000.

We used a linear mixed model implemented in the R package variancePartition version 3.6 (ref. ⁸⁰) to identify genes associated with GRN and 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 90th quantile of the variation distribution for all genes associated with the respective trait.

Subsequently, a weighted gene co-expression network analysis (WGCNA) approach⁸¹ using the WGCNA R package⁸² 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 gangfish, 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 PANTHER⁸⁴. 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).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The sequence datasets have been deposited in the National Center for Biotechnology Information 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: <https://doi.org/10.5525/gla.researchdata.680>.

Received: 22 March 2018; Accepted: 2 November 2018;

Published online: 3 December 2018

References

- Hautier, Y. et al. Anthropogenic environmental changes affect ecosystem stability via biodiversity. *Science* **348**, 336–340 (2015).
- Hendry, A. P., Gotanda, K. M. & Svensson, E. I. Human influences on evolution, and the ecological and societal consequences. *Philos. Trans. R. Soc. B* **372**, 20160028 (2017).
- Des Roches, S. et al. The ecological importance of intraspecific variation. *Nat. Ecol. Evol.* **2**, 57–64 (2018).
- Vonlanthen, P. et al. Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482**, 357–362 (2012).
- 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).
- 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).
- Rudman, S. M. & Schluter, D. Ecological impacts of reverse speciation in threespine stickleback. *Curr. Biol.* **26**, 490–495 (2016).
- 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).
- Alexander, T. J., Vonlanthen, P. & Seehausen, O. Does eutrophication-driven evolution change aquatic ecosystems? *Philos. Trans. R. Soc. B* **372**, 20160041 (2017).
- Gilman, R. T. & Behm, J. E. Hybridization, species collapse, and species reemergence after disturbance to premating mechanisms of reproductive isolation. *Evolution* **65**, 2592–2605 (2011).
- Yeaman, S. & Whitlock, M. C. The genetic architecture of adaptation under migration–selection balance. *Evolution* **65**, 1897–1911 (2011).
- Schluter, D. Evidence for ecological speciation and its alternative. *Science* **323**, 737–741 (2009).
- Maan, M. E. & Seehausen, O. Ecology, sexual selection and speciation. *Ecol. Lett.* **14**, 591–602 (2011).
- Nümann, W. Attempt of a classification of the coregonines in the Lake of Constance with regard to the so-called blauefchen by combination of several characteristics. *Arch. Hydrobiol.* **82**, 500–521 (1978).
- Hirsch, P. E., Eckmann, R., Oppelt, C. & Behrmann-Godel, J. Phenotypic and genetic divergence within a single whitefish form—detecting the potential for future divergence. *Evol. Appl.* **6**, 1119–1132 (2013).
- Jochimsen, M. C., Kümmerlin, R. & Straile, D. Compensatory dynamics and the stability of phytoplankton biomass during four decades of eutrophication and oligotrophication. *Ecol. Lett.* **16**, 81–89 (2013).
- Thomas, G., Quoss, H., Hartmann, J. & Eckmann, R. Human-induced changes in the reproductive traits of Lake Constance common whitefish (*Coregonus lavaretus*). *J. Evol. Biol.* **22**, 88–96 (2009).
- Lundsgaard-Hansen, B., Matthews, B., Vonlanthen, P., Taverna, A. & Seehausen, O. Adaptive plasticity and genetic divergence in feeding efficiency during parallel adaptive radiation of whitefish (*Coregonus* spp.). *J. Evol. Biol.* **26**, 483–498 (2013).
- Harrod, C., Mallela, J. & Kahilainen, K. K. Phenotype–environment correlations in a putative whitefish adaptive radiation. *J. Anim. Ecol.* **79**, 1057–1068 (2010).
- Østbye, K., Bernatchez, L., Naesje, T. F., Himberg, K.-J. M. & Hindar, K. Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Mol. Ecol.* **14**, 4371–4387 (2005).
- Quevedo, M., Svanbäck, R. & Eklöv, P. Intrapopulation niche partitioning in a generalist predator limits food web connectivity. *Ecology* **90**, 2263–2274 (2009).
- Behrmann-Godel, J. Parasite identification, succession and infection pathways in perch fry (*Perca fluviatilis*): new insights through a combined morphological and genetic approach. *Parasitology* **140**, 509–520 (2013).
- Vonlanthen, P. et al. Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *J. Evol. Biol.* **22**, 498–514 (2009).
- Luczynski, M., Rösch, R., Vuorinen, J. A. & Brzuzan, P. Biochemical genetic study of sympatric Lake Constance whitefish (*Coregonus lavaretus*) populations: blauefchen and gangfish. *Aquat. Sci.* **57**, 136–143 (1995).
- Gingerich, P. D. Quantification and comparison of evolutionary rates. *Am. J. Sci.* **293**, 453–478 (1993).
- Hamilton, J. A. & Miller, J. M. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv. Biol.* **30**, 33–41 (2016).
- Shafer, A. B. A. & Wolf, J. B. W. Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecol. Lett.* **16**, 940–950 (2013).
- Laporte, M. et al. RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in lake whitefish (*Coregonus clupeaformis*) species pairs. *G3* **5**, 1481–1491 (2015).
- Chaves, J. A. et al. Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol. Ecol.* **25**, 5282–5295 (2016).
- Pfeifer, S. P. et al. The evolutionary history of Nebraska deer mice: local adaptation in the face of strong gene flow. *Mol. Biol. Evol.* **35**, 792–806 (2018).
- Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
- Burri, R. Interpreting differentiation landscapes in the light of long-term linked selection. *Evol. Lett.* **1**, 118–131 (2017).
- Delmore, K. E. et al. Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evol. Lett.* **2**, 76–87 (2018).
- Gagnaire, P.-A., Normandeau, E., Pavey, S. A. & Bernatchez, L. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the lake whitefish (*Coregonus clupeaformis*). *Mol. Ecol.* **22**, 3036–3048 (2013).
- Jacobs, A., Womack, R., Chen, M., Gharbi, K. & Elmer, K. Significant synteny and co-localization of ecologically relevant quantitative trait loci within and across species of salmonid fishes. *Genetics* **207**, 741–754 (2017).
- Ranz, J. M. & Machado, C. A. Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol.* **21**, 29–37 (2006).
- Gibson, G. The environmental contribution to gene expression profiles. *Nat. Rev. Genet.* **9**, 575–581 (2008).

38. Jeukens, J., Renaut, S., St-Cyr, J., Nolte, A. W. & Bernatchez, L. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol. Ecol.* **19**, 5389–5403 (2010).
39. Park, P. J. & Bell, M. A. Variation of telencephalon morphology of the threespine stickleback (*Gasterosteus aculeatus*) in relation to inferred ecology. *J. Evol. Biol.* **23**, 1261–1277 (2010).
40. Conejeros, P. et al. Differentiation of sympatric Arctic char morphotypes using major histocompatibility class II genes. *Trans. Am. Fish. Soc.* **143**, 586–594 (2014).
41. Jacobs, A. et al. Convergence in form and function overcomes non-parallel evolutionary histories in Arctic charr. Preprint at <https://doi.org/10.1101/265272> (2018).
42. Ahi, E. Signalling pathways in trophic skeletal development and morphogenesis: insights from studies on teleost fish. *Dev. Biol.* **420**, 11–31 (2016).
43. Yohannes, E., Grimm, C., Rothhaupt, K.-O. & Behrmann-Godel, J. The effect of parasite infection on stable isotope turnover rates of $\delta^{15}\text{N}$, $\delta^{13}\text{C}$ and $\delta^{34}\text{S}$ in multiple tissues of Eurasian perch *Perca fluviatilis*. *PLoS ONE* **12**, e0169058 (2017).
44. Siwertsson, A., Knudsen, R., Adams, C. E., Præbel, K. & Amundsen, P. A. Parallel and non-parallel morphological divergence among foraging specialists in European whitefish (*Coregonus lavaretus*). *Ecol. Evol.* **3**, 1590–1602 (2013).
45. Rohlf, F. J. *TpsDig* (Department of Ecology and Evolution, State University of New York, Stony Brook, 2004); <http://life.bio.sunysb.edu/morph/>
46. Klingenberg, C. P. MorphoJ: an integrated software package for geometric morphometrics. *Mol. Ecol. Resour.* **11**, 353–357 (2011).
47. Klingenberg, C. P. & McIntyre, G. S. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution* **52**, 1363–1375 (1998).
48. Benjamini, Y., Hochberg, Y., Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
49. Haldane, J. B. S. Suggestions as to quantitative measurement of rates of evolution. *Evolution* **3**, 51–56 (1949).
50. Recknagel, H., Jacobs, A., Herzyk, P. & Elmer, K. R. Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. *Mol. Ecol. Resour.* **15**, 1316–1329 (2015).
51. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data* (2010); <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
52. Catchen, J. M. et al. Stacks: building and genotyping loci de novo from short-read sequences. *G3* **1**, 171–182 (2011).
53. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
54. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 1–16 (2015).
55. Lischer, H. E. L. & Excoffier, L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298–299 (2012).
56. Alexander, D. H. & Novembre, J. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
57. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
58. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
59. Meirmans, P. G. & Tienderen, P. H. genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* **4**, 792–794 (2004).
60. Weir, B. S. & Cockerham, C. C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
61. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
62. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
63. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
64. Kautt, A. F., Machado-Schiaffino, G. & Meyer, A. Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. *PLoS Genet.* **12**, e1006157 (2016).
65. Rougeux, C., Bernatchez, L. & Gagnaire, P.-A. A. Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biol. Evol.* **9**, 2057–2074 (2017).
66. Egger, B., Rösti, M., Böhne, A., Roth, O. & Salzburger, W. Demography and genome divergence of lake and stream populations of an East African cichlid fish. *Mol. Ecol.* **26**, 5016–5030 (2017).
67. Sutherland, B. J. G. et al. Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps. *Genome Biol. Evol.* **8**, 3600–3617 (2016).
68. Moore, J. et al. Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic char. *Mol. Ecol.* **26**, 6784–6800 (2017).
69. Lien, S. et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200–205 (2016).
70. Luu, K., Bazin, E. & Blum, M. G. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* **17**, 67–77 (2017).
71. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).
72. Frichot, E., Schoville, S. D., Bouchard, G. & François, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* **30**, 1687–1699 (2013).
73. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
74. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
76. Carruthers, M. et al. De novo transcriptome assembly, annotation and comparison of four ecological and evolutionary model salmonid fish species. *BMC Genomics* **19**, 32 (2018).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
78. Roberts, A. & Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat. Methods* **10**, 71–73 (2013).
79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
80. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016).
81. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
82. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 1–13 (2008).

Acknowledgements

We thank H. Thiele, M. Schmid, W. Kornberger and A. Sulger for assistance with specimen and 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 a Marie Curie Action Career Integration Grant (321999) to K.R.E., a BBSRC WestBio Doctoral Training Partnership studentship to M.C., C.E.A. and K.R.E. (BB/J013854/1), ERASMUS+ (to J.B.-G. and K.R.E.), a Fisheries Society of the British Isles Research Grant (to K.R.E. and A.J.), and AFF funding from the University of Konstanz to E.Y. and J.B.-G.

Author contributions

J.B.-G. and K.R.E. designed the experiment. A.J., M.C., K.R.E. and J.B.-G. collected the data. M.C. generated and analysed the eco-morphological and transcriptomic data. A.J. generated and analysed the genomic data, and analysed the eco-morphological and stable isotope data. R.E. analysed the life history data. E.Y. generated the stable isotope data. C.E.A., J.B.-G. and K.R.E. supervised the project. A.J. wrote the paper, along with M.C., J.B.-G. and K.R.E. All authors commented on and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41559-018-0742-9>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.B. or K.R.E.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2018

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- ☐ ☒ Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection See full documentation in the Methods section.

Data analysis See full documentation in the Methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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>.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Investigation of phenotypic diversification of European whitefish (<i>Coregonus lavaretus</i>) in Lake Constance following the re-oligotrophication of the lake. Phenotypic (gill raker; geometric morphometrics of body shape) and ecological (stable isotopes, parasites) investigation of whitefish diversity, and genotyping of 170 whitefish using ddRADseq and gene expression analysis of 27 individuals using RNAseq.
Research sample	European whitefish (<i>Coregonus lavaretus</i>) subspecies from Lake Constance: 270 gangfisch (<i>C. l. macrophthalmus</i>); 20 blaufelchen (<i>C. l. wartmanii</i>) and 20 weissfelchen (<i>C. l. fera</i>) were used in total.
Sampling strategy	270 gangfisch were caught at spawning time along the spawning depth gradient at three different sites across Lake Constance, based on known spawning sites and depths. Fish were caught using gill nets with different mesh sizes to cover the full size range. Furthermore, blaufelchen and weissfelchen were also caught during spawning time using gill nets, but at different sites and times.
Data collection	See Methods section for detailed description of data collection. In general, data were collected by A.J, M.C, J.B.G, R.E and K.R.E
Timing and spatial scale	European whitefish samples were collected in December 2014 (gangfisch) and October/November 2015 (blaufelchen and weissfelchen)
Data exclusions	Data were excluded from analysis if technical reasons existed (e.g. bad data quality of stable isotope data or low sequencing coverage).
Reproducibility	Samples from different spawning sites in the lake were analysed and compared to other subspecies to ensure reproducibility. The extensive analysis was not repeated in other lakes for reasons mentioned in the ms.
Randomization	Individuals were randomly sampled from gill nets for each site, depth and mesh size.
Blinding	Blinding was not necessary as all statistics were directly computed from phenotypic, ecological, genomic and transcriptomic data.
Did the study involve field work?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

Field work, collection and transport

Field conditions	Detailed conditions were not recorded. Air temperature during sampling was below 0 degrees Celsius, and the water temperature was around 4 degree Celsius.
Location	Gangfisch were sampled at three different sites (Egg, Immenstaad, Bodman; see Methods) at three different water depths (2m, 25m, 50m). Blaufelchen were sampled over open water of Romanshorn and weissfelchen in the lower lake. (see methods and supplementary data for more detail.
Access and import/export	Scientific fish collections in Lake Constance were carried out under a permit from the Regierungspräsidium Tübingen: 'Aktenzeichen 33-4/9220.51-3'
Disturbance	Disturbances were minimized by localised sampling in the early morning in collaboration with local fishermen.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Unique biological materials

Policy information about [availability of materials](#)

Obtaining unique materials Tissues / materials are stored in the collection of K.R.E at the University of Glasgow and can be examined on site on request.


Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Not applicable.
Wild animals	European whitefish were caught using gill nets and either directly euthanised and stored on ice, or transported back to the laboratory in aerated containers and euthanised directly before taking tissue samples for gene expression analysis (see Methods for details).
Field-collected samples	Not applicable.

In the format provided by the authors and unedited.

Rapid niche expansion by selection on functional genomic variation after ecosystem recovery

Arne Jacobs ^{1,4}, Madeleine Carruthers^{1,4}, Reiner Eckmann², Elizabeth Yohannes², Colin E. Adams ^{1,3},
Jasminca Behrmann-Godel ^{2*} and Kathryn R. Elmer ^{1*}

¹Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. ²Limnological Institute, University of Konstanz, Konstanz, Germany. ³Scottish Centre for Ecology and the Natural Environment, University of Glasgow, Glasgow, UK. ⁴These authors contributed equally: Arne Jacobs, Madeleine Carruthers. *e-mail: jasminca.behrmann@uni-konstanz.de; kathryn.elmer@glasgow.ac.uk

Supplementary information: Rapid niche expansion by selection on functional genomic variation after ecosystem recovery

Arne Jacobs, Madeleine Carruthers, Reiner Eckmann, Elizabeth Yohannes, Colin E. Adams, Jasminca Behrmann-Godel, & Kathryn R. Elmer

A. Supplementary methods

Eco-morphological data.

For geometric morphometric analyses, individuals were photographed on their left side using a Canon EOS 7D digital camera and 40 mm lens, using a standardized distance for all individuals and including a scale.

For aging and back calculation of growth rates, scales were cleaned, dry mounted between two glass plates and inspected under a microscope. Age was determined from annular marks on scales, and their radii (i.e. the distances between the scale centre and each annual mark) were measured from digital images to the nearest 0.01mm. Length-at-age was then back-calculated using a regression of fish length on scale radius that is based on our unpublished database spanning several years of sampling whitefish in Lake Constance. Three scales per fish were analysed, and back-calculated lengths were averaged.

Ecological and trophic niche use of gangfish was inferred using stable isotope data and *Diplostomum spp.* (eye fluke) infection data. Stable isotope data from muscle tissue give a detailed account of trophic niche over the last few weeks before sampling, allowing one to examine differences in trophic niche use during the autumn/winter feeding season. In contrast, eye fluke infection data give a more long-term view on habitat use during the spring and summer, as fish can only get infected by *Diplostomum* cercariae in that period. These cercariae shed from benthic snails, their first intermediate hosts, which live exclusively in shallow water down to a maximum depth of 15m²². Due to the location of their host, the low mobility of cercariae and the fact that they have to find a (fish) host within 48h, means that they are good indicators for a littoral-benthic habitat use by gangfish during the spring and summer.

Sampling of historical gangfish samples.

The historical gangfish samples from 1990 used for the analysis of gill raker number in Luczynski *et al.*²⁴ were sampled in an almost identical way to the contemporary sample used in this study. First, in both sampling campaigns, gangfish were sampled during spawning time, which is a period of two to three weeks during the winter, around November to December. Second, samples were caught at a comparable depth gradient ranging from shallow water (2m) to deeper water (40m or 50m). Although Luczynski *et al.*²⁴ only fished to a depth of 40m, this is highly comparable to the habitat at 50m, due to the bathymetry of Lake Constance. From the shore onwards, Lake Constance displays a shallow littoral zone reaching water depths of around 2-3m before a steep drop-off. Afterwards, this shallow zone is followed by a steep slope, reaching depths of around 20-30m. At this point, the slope becomes less steep (moderate slope) and continues to deeper depths or reaches the flat bottom (30m and below) (Supplementary Fig. 1). Thus, both sampling campaigns have sampled in similar bathymetric areas: 2m – shallow littoral zone, 25m – end of the steep slope, 40/50m – moderate slope or flat lake bottom.

Third, the range in mesh size used in both sampling campaigns were identical (32 – 48mm), covering the size range of adult spawning gangfisch. Thus, we were confident that the methods used in both sampling campaigns are similar enough to allow accurate comparisons of gill raker counts for gangfisch across sampling times.

Genotyping and bioinformatics.

Stacks v.1.46⁵² was used for the assembly of RAD loci and genotype calling. We used the following settings. First, the *denovo_map.pl* pipeline was used for the *de novo* assembly of RAD loci, allowing a maximum of two nucleotide mismatches for primary reads (M=2), a minimum stack depth of six (m=6), a maximum of three nucleotide mismatches when aligning secondary reads (N=3), and a maximum number of one nucleotide difference when building the catalogue (n=1). The *rxstacks* module was used for genotype correction and haplotype pruning using an upper bound of 0.05 for the genotype error rate and a significance threshold of 5%, an overall log-likelihood cut-off of -10 and by removing loci that were confounded in more than 25% of individuals and showed excess of haplotypes within individuals. The *populations* module was used for extracting and filtering genotypes. A genotype was retained in the population genomic dataset if it fulfilled the following criteria: (i) it was present in at least 75% of individuals within a population (subspecies) in at least two out the three subspecies, (ii) global minor allele frequency (MAF) of ≥ 0.05 (iii) an observed heterozygosity of < 0.50 , and (iv) a minimum coverage of 6x. Furthermore, only one SNP per locus was retained for analyses to reduce the impact of linkage. The filtering criteria for the demographic modelling differed slightly from the described filtering in that no MAF filter was applied in order to retain informative low frequency SNPs and a genotype had to be present in 90% of individuals within a dataset to reduce the impact of missing data on the calculation of site frequency spectrum (see below).

Demographic modelling.

To investigate the evolutionary history of divergence and introgression of European whitefish in Lake Constance, we tested seven different demographic models using two different SNP datasets. First, we used a dataset including all SNPs in the dataset (N=22,196 SNPs) and second we used a dataset without loci showing signs of differentiation or selection (N=22,100 SNPs).

We tested seven different models (Supplementary Fig. 4) describing different combinations of gene flow and introgression based on the same species relationship: strict isolation (SI), (SCint) secondary contact and introgression; (SC_IM_int) secondary contact and introgression after isolation-with-migration of weissfelchen (WF) and gangfisch (GF) and introgression; (SCchange_int) ancient secondary contact without introgression and isolation with migration followed by introgression; (IM) isolation-with-migration; (IMint) isolation-with-migration and introgression; (IM_SC_int) IM with secondary contact between WF and GF and introgression. The SI model describes the divergence between species without gene flow, whereas IM describes the divergence under ongoing gene flow.

Furthermore, we tested an IM model that included a strong admixture event (IMint) from weissfelchen (WF) and blaufelchen (BF) into gangfisch (GF). We also tested a variation of this model that did not include gene flow between WF and GF after divergence before admixture (IM_SC_int). Additionally, we tested three different models describing a secondary contact scenario. First, we tested a model in which all species diverged without gene flow and a came into secondary under admixture, at which point gene flow started (SCint). Second, The SC_IM_int model is a variation of this model under which WF and GF diverged under gene flow. Lastly, the SCchange_int model is a variation of SCint model, at which BF and the ancestor of GF and WF diverged without gene flow, but gene flow started before the split of GF and WF. Again, this model also included an admixture event from WF and BF into GF. All models inferred bi-directional rates of gene flow, as the direction of gene flow cannot be inferred using folded site-frequency spectra.

Genome-wide association analyses.

Genotypes were imputed using a random forest approach implemented in the *radiator* R-package (150 trees). To test the genotype imputation accuracy, we calculated Spearman correlations between allele frequencies before and after imputation using the *cor.test* function in R.

The BSLMM analysis was performed using the following settings: 10 independent BSLMM chains for 25 million steps with a burn-in of 5 million steps, recording of parameters every 100 steps and writing of results every 10,000 steps. We calculated means and 95% credible intervals [95% CIs] for all hyperparameters in R. We averaged PIP (posterior inclusion probability) across all runs and used a cut-off of PIP > 0.1 to identify SNPs with significant effect on gill raker number (GRN) or body shape^{29,30}.

Transcriptomic analyses.

Total RNA was extracted from ~30g of white muscle tissue from 27 individuals (three per depth, per site), using PureLink RNA Mini kits (Life Technologies, Carlsbad, CA). Extractions were carried out following the manufacturer's instructions, with the exception of an additional homogenisation step using a FastPrep-24 (MP Biomedicals) prior to isolation⁸³. RNA was quantified using the HS Assay kit for the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA) and quality was assessed with a 2200 TapeStation (Agilent, Santa Clara, CA). High quality RNA was achieved, with A260/280 ratios between 1.9 and 2.1 and RNA Integrity Numbers above 8.4.

Raw reads were processed prior to alignment. Adapter sequences were removed with *Scythe* v0.9944 BETA (<https://github.com/vsbuffalo/scythe/>)⁸⁴, and low quality reads were trimmed with *Trimmomatic* v0.36⁷⁴. Leading and trailing bases with a Phred quality score <20 were removed and 4 bp sliding window approach was used to trim reads where the average Phred score fell below 20. Minimum read length was set to 50 bp. Processing removed ~4% of the total reads, resulting in a final set of ~790 million cleaned reads.

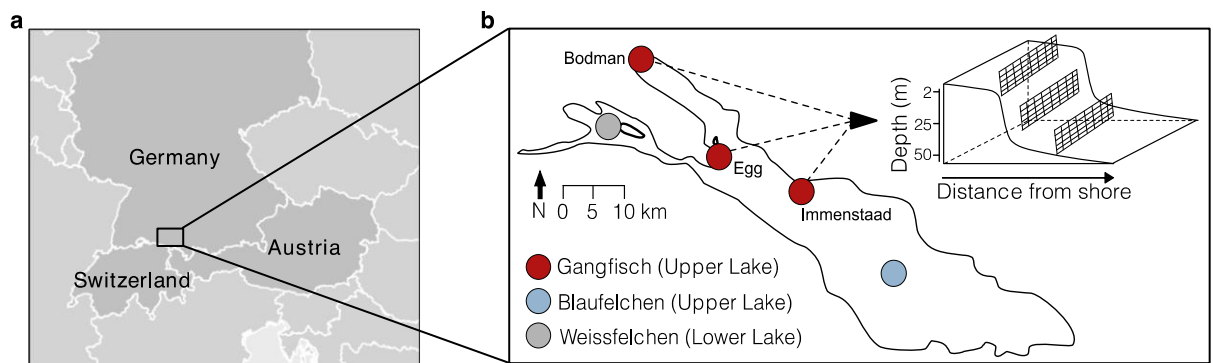
We used the R-package *variancePartition* v3.6⁷⁹ to identify genes associated with trophic modality (i.e. GRN and body shape) and pre-filter the gene expression dataset to remove background signals. *VariancePartition* uses a linear mixed model approach to calculate the percentage of expression variation explained by multiple traits of interest (site, depth, sex, GRN, and body shape), with all remaining expression variation not attributable to the specified traits termed residual variation. Batch effects were removed using an initial LMM, with ‘batch’ as a ‘random’ effect, as per the *variancePartition* pipeline⁷⁹. The residuals from the initial model then formed the response for a second LMM, with site, depth, sex, GRN, and body shape PC2 as explanatory variables. Categorical variables (site, depth and sex) were set as random effects and continuous variables (GRN, body shape PC2) as fixed effects.

A Weighted Gene Co-expression Network Analysis (WGCNA) approach was used to conduct a meta-analysis of gene expression divergence between gangfish inhabiting different trophic niches within Lake Constance. A single gene co-expression network was constructed for all 27 samples, from the log₂ scaled count data, using the *WGCNA* R package⁸¹ following the standard procedure. Briefly, WGCNA generated pair-wise Pearson’s correlations for all genes, and for all samples, which acts as a measure of gene co-expression. Gene co-expression measures were then raised to a chosen power ($\beta=9$); soft-thresholding power was chosen based on the biologically motivated criterion of scale free topology (Supplementary Fig. 7d;⁸⁰), selecting the lowest integers above the model fit of $R^2=0.9$. Next the topological overlap matrix (TOM) was calculated from the adjacency matrix to measure network interconnectedness (the strength of a co-expression relationship between any two genes with respect to all other genes in the network)⁸⁵. Genes with highly similar co-expression relationships were clustered using average linkage hierarchical clustering of the topological overlap dissimilarity measure (1-TOM). Finally, network modules were defined using the dynamic treecut algorithm, with a minimum module size of 30 genes and a cut height of 0.992. The module eigengene distance threshold was set to 0.25 to merge highly similar modules.

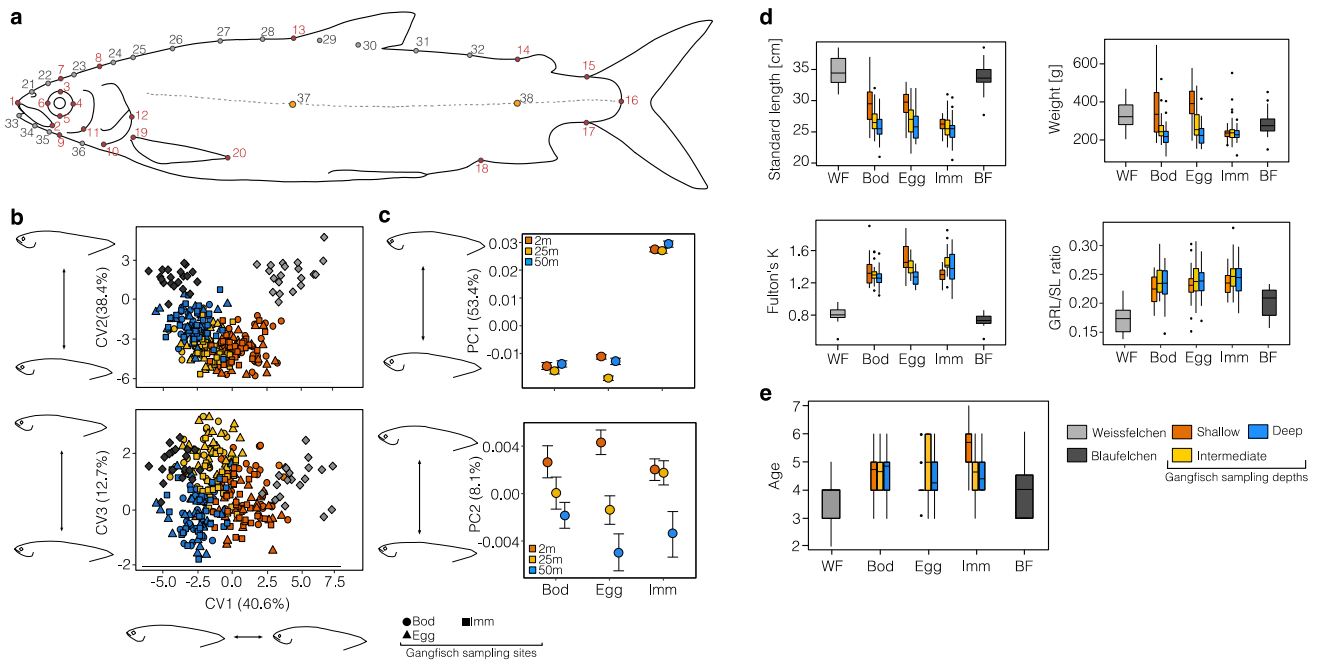
B. Supplementary results.

Heterogeneity across sampling sites.

While patterns of phenotypic divergence along the depth gradient were largely consistent across all three sites, we found that individuals from the sampling site Immenstaad consistently showed less divergence along the depth gradient compared to individuals from either Bodman or Egg and also differed in body shape compared to the other two sampling locations (Supplementary Fig. 2; BS-PC1: 53.4% of total variation). Furthermore, individuals from Immenstaad showed a lower degree of divergence in gill raker number across spawning depths (Fig. 1b). These patterns may partially be explained by the difference in environmental pressures experienced at Immenstaad, such as increased wind and wave exposure⁸⁶, compared to both Bodman and Egg. Additionally, it was found that spatio-temporal differences in zooplankton population dynamics exist between different parts of Lake Constance¹⁶, namely differences between the Überlingersee (where Bodman and Egg are situated) and the main body of the upper lake (where Immenstaad is situated). Differences in the availability and abundance of pelagic food resources at different locations within the lake may influence the extent of phenotypic divergence between gangfisch inhabiting different parts of the lake, by altering the selection pressures associated within resource acquisition. Differences in body shape in individuals from Immenstaad may be a plastic response due to the different environmental pressures in that part of the lake, as no consistent genetic differences were observed between this and the other sites (Supplementary Fig. 6e). The lower degree of phenotypic divergence along the depth gradient in some traits could be potentially explained by the lower degree of genetic admixture in gangfisch from Immenstaad (Supplementary Fig. 4a,b). However, the heterogeneity in the extent of divergence across sampling sites did not affect the overall contemporary phenotypic diversity observed in gangfisch, and overall there was consistent divergence across depths and sites for a range of traits.

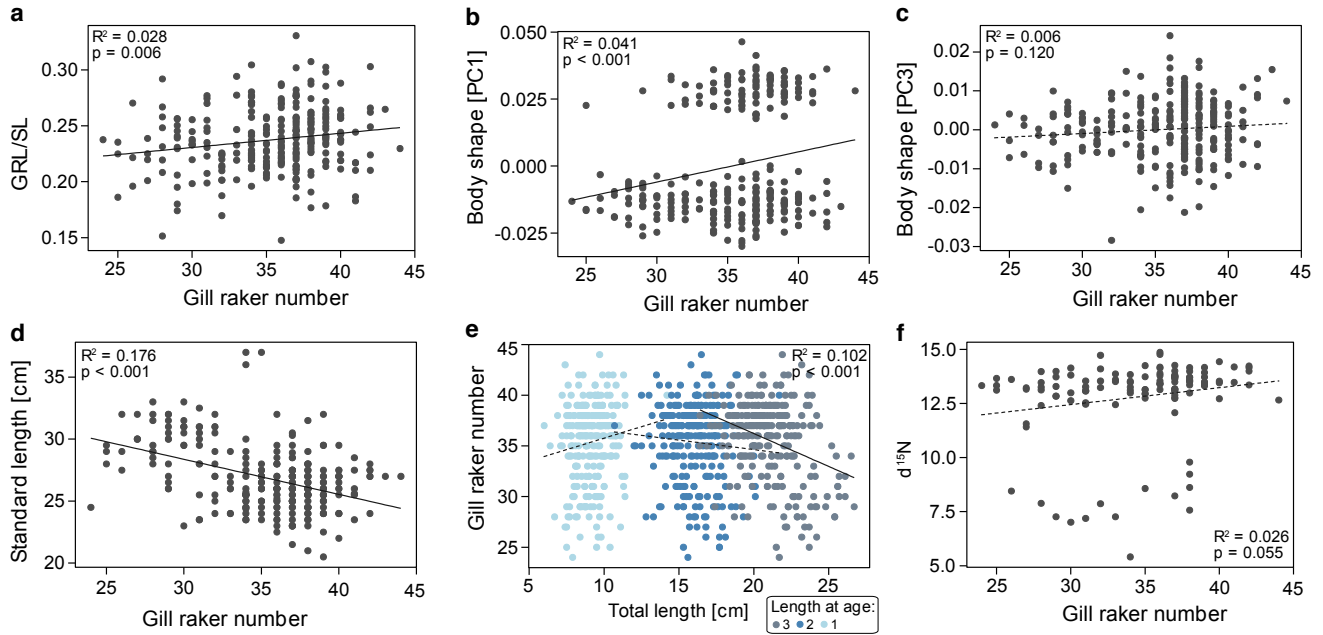


Supplementary Figure 1 | Summary of sampling sites and datasets. **a**, Map of central Europe showing the location of Lake Constance. **b**, A close-up view of Lake Constance showing the location of sampling sites for gangfisch (red circles) and blaufelchen (blue circle) within Upper Lake Constance, and weissfelchen (grey circle) within Lower Lake Constance. Gangfisch from all three sites (Bodman, Egg and Immenstaad) were sampled at three different spawning depths; shallow: 2m, intermediate/steep slope: 25m, and deep/bottom or gentle slope: 50m.

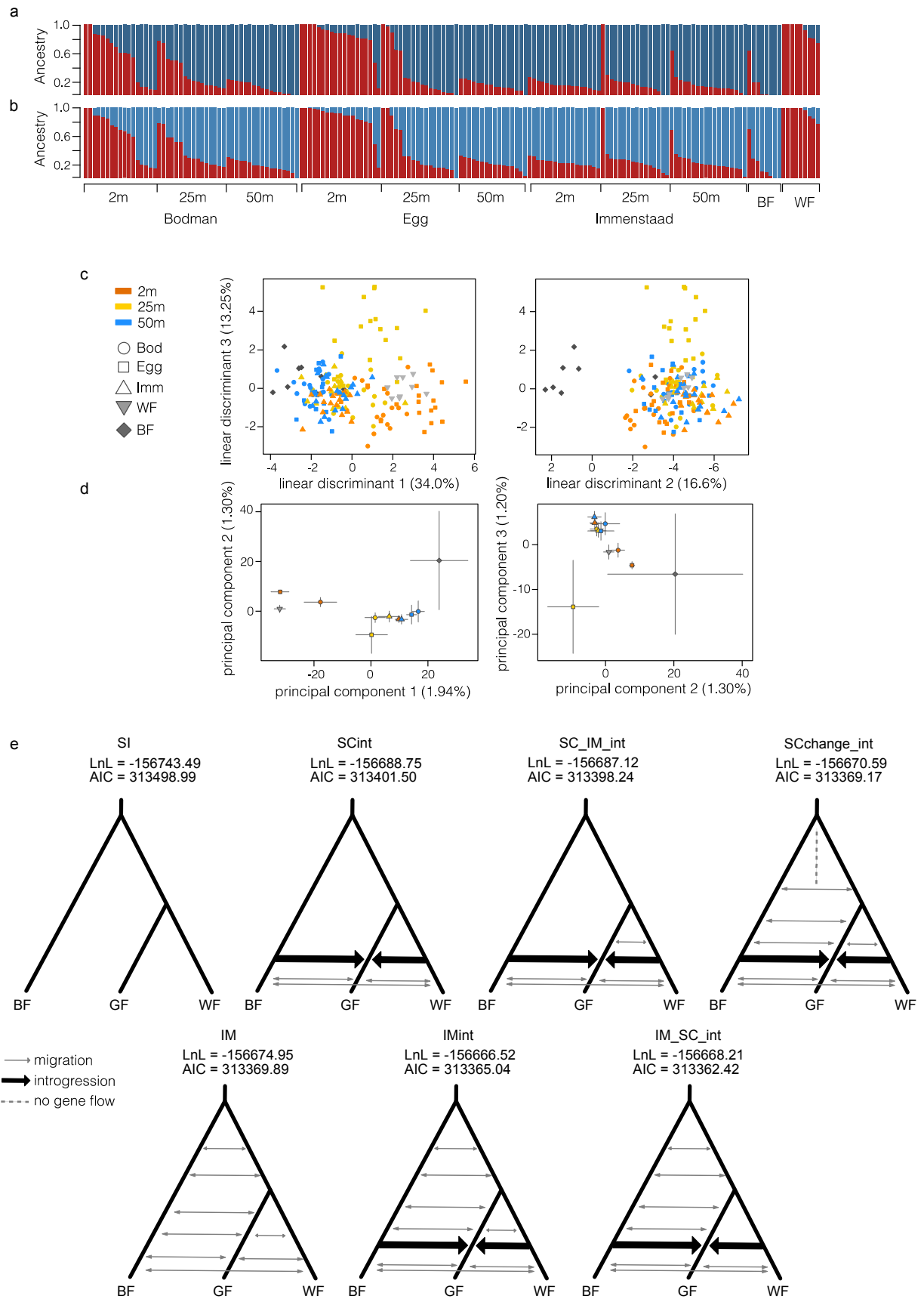


Supplementary Figure 2 | Phenotypic variation across spawning depths. **a**, Positions of landmarks used for geometric morphometric analysis. Red dots = standing landmarks, grey dots = sliding semi-landmarks, orange dots = ‘helper’ landmarks used for bending correction (not included in shape analyses). Landmark descriptions: 1) tip of the snout, 2) end of the maxilla, 3 - 6) most posterior, anterior, ventral and dorsal parts of the eye, 7) edge of cranium directly above the eye, 8) posterior edge of the cranium, 9) fusion point of pre-operculum and operculum, 10) ventral end of operculum, 11) posterior end of pre-operculum, 12) posterior end of operculum, 13) anterior base of the dorsal fin, 14) anterior base of the adipose fin, 15) anterior edge of the most dorsal caudal fin spine, 16) posterior end of lateral line, 17) anterior end of the most ventral caudal fin spine, 18) anterior base of the anal fin, 19) dorsal insertion point of pectoral fin, 20) posterior tip of pectoral fin, 21 - 32) ventral curve of body shape, 33 - 36) curve of anterior head shape, 37) point on lateral line below anterior base of dorsal fin, 38) point on lateral line below anterior end of adipose fin. **b**, Canonical variate analysis (CVA) showing variation in body shape between gangfisch caught at three different spawning depths (shallow in orange, intermediate in yellow and deep in blue) from Bodman (circles), Egg (triangles) and Immenstaad (squares) (N=270, 30 per depth, per site), and the extant benthic ecotype, weissfelchen (light grey diamonds, N=20) and extant pelagic ecotype, blaufelchen (dark grey diamonds, N=20). The top figure in **b** represents shape change along CV1 on the x-axis and CV2 on the y-axis (explaining 40.6% and 38.4% of the total variation, respectively), and the bottom figure in **b** represents shape change along CV1 on the x-axis and CV3 on the y-axis (explaining 40.6% and 12.7% of the total variation, respectively). Wireframes depict body shape at outer most point of each axis. **c**, Principal component analysis (PCA) showing variation in body shape in gangfisch (N=270, 30 per depth, per site). Gangfisch caught at shallow spawning depths are shown in orange, intermediate in yellow and deep in blue, for all the three sites (Bodman, Egg and Immenstaad). Data are shown as

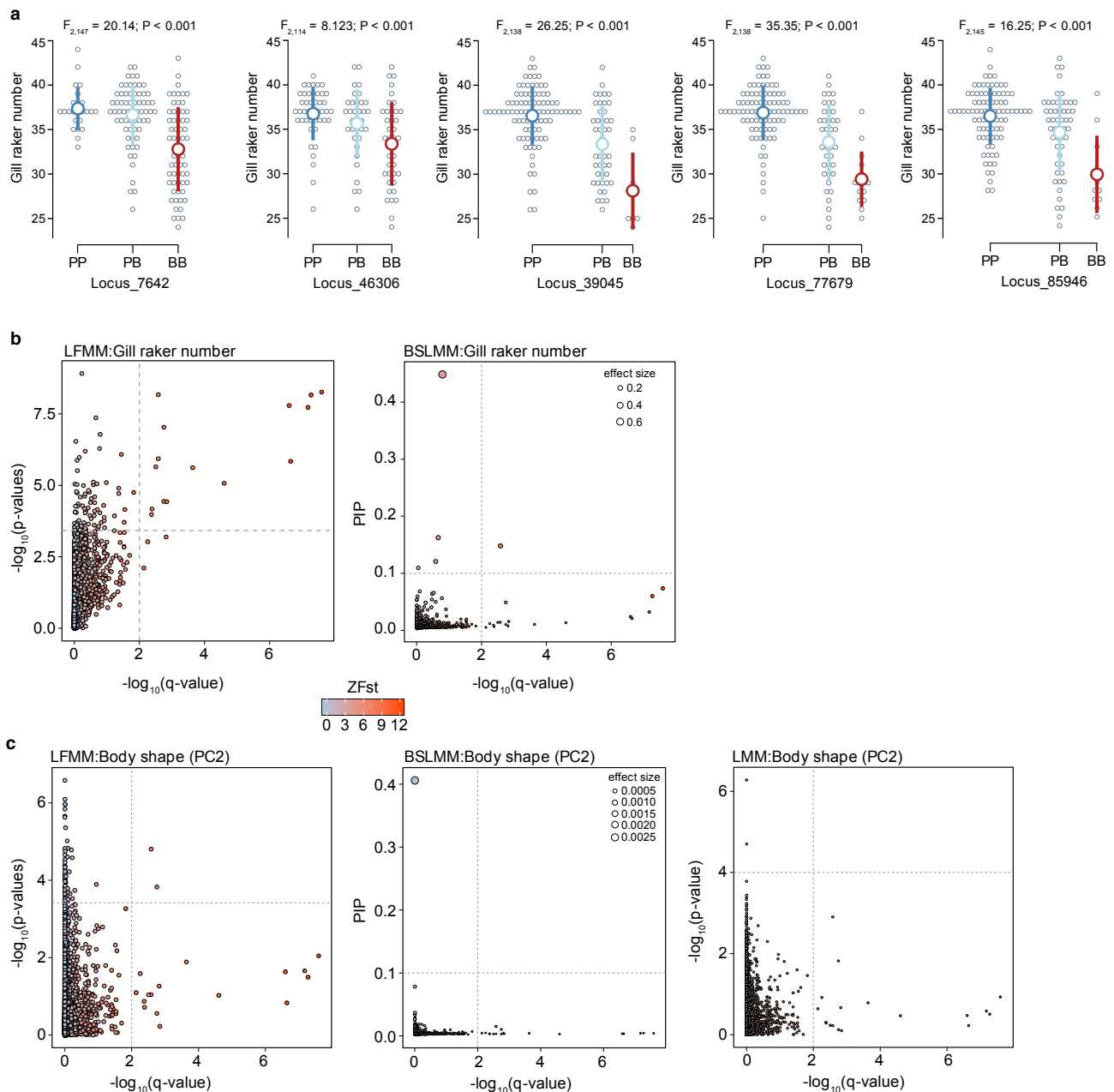
mean (\pm SEM) for each of the nine sampling locations; the top figure in **c** shows shape variation along the first principal component (PC1), explaining 53.4% of the total variation, and the bottom figure in **c** shows shape variation along the second principal component (PC2), explaining 8.1% of the total variation. Wireframes depict body shape at outer most point of each axis. **d**, Variation of eco-morphological traits across samplings sites, depths and subspecies for standard length (cm), weight (g) and Fulton's condition factor (K) and SL-corrected gill raker length (GR/SL). **e**, Age distribution of gangfish caught at different sites and spawning depths (shallow in orange, intermediate in yellow, and deep in blue; N=270).



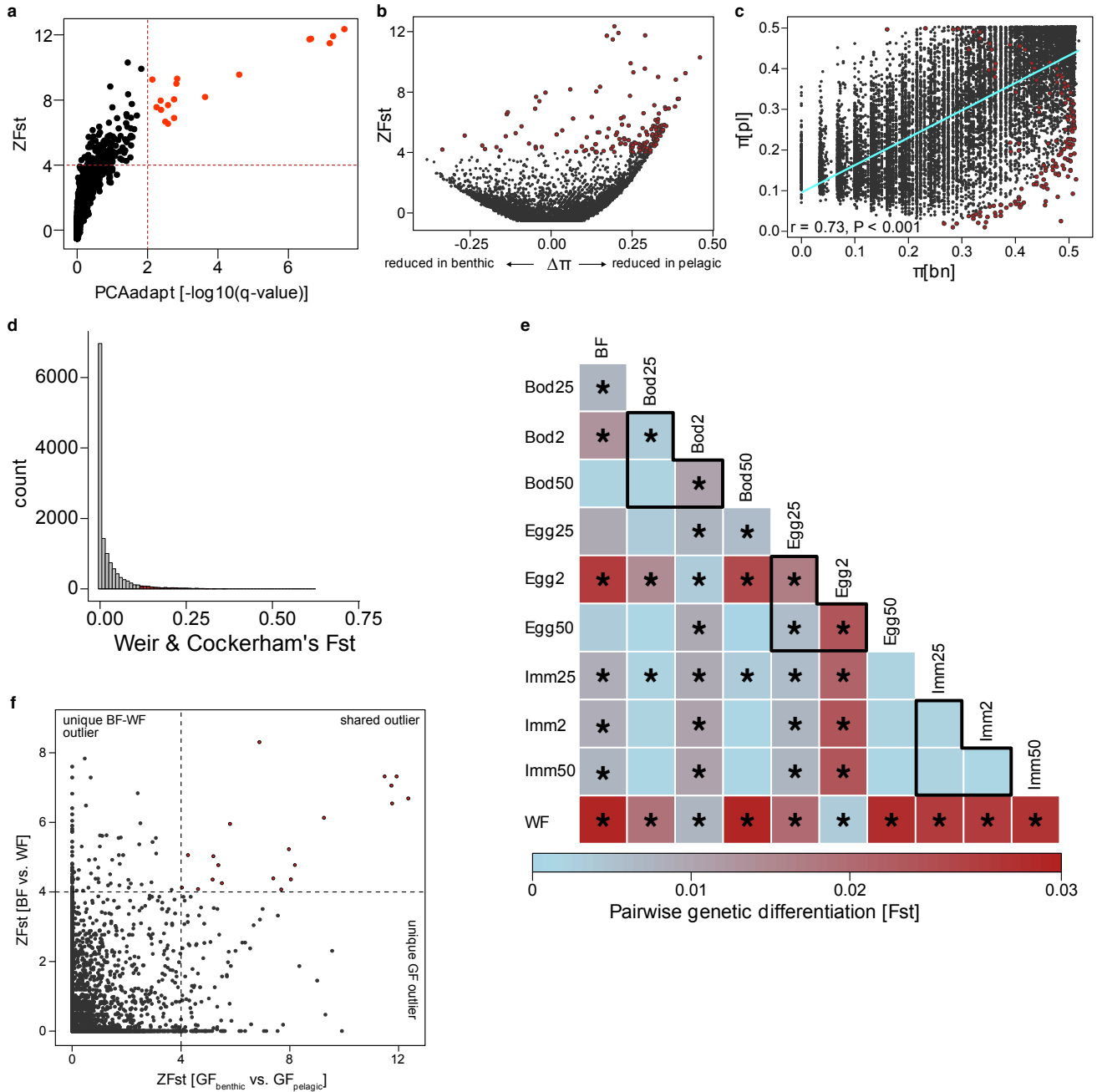
Supplementary Figure 3 | Testing for correlations of gill raker number with ecological variables and morphological traits. Of the additional morphological traits analysed, gill raker length (corrected for body length: GR/SL; **a**; N=269), body shape PC1 (**b**; N=268) and standard length (**d**, N=269) were significantly correlated with gill raker number (GRN). **c**, body shape PC3 was not significantly correlated with gill raker number. **d**, The nitrogen stable isotope signature was not correlated with the number of gill rakers (N=122) in gangfisch. **e**, Furthermore, the total length of individuals at age one, two and three (N=260), which was back-calculated from scales, was only correlated with gill raker number at age three, indicating an increase in growth in gangfisch with lower gill raker numbers at this age. R^2 -values and p-values from individual linear models are given within each plot. R^2 -values and p-values for the correlation of gill raker number and total length at age one and two are: One: $R^2=0.0114$, $p=0.055$; Two: $R^2=0.0022$, $p=0.210$.



Supplementary Figure 4 | Population structure of whitefish in lake Constance. **a,b**, Genetic ancestry of gangfisch by site and spawning depth, weissfelchen (WF) and blaufelchen (BF), highlighting the different proportions of genetic ancestry by depth and site, based on the full SNP dataset (**a**) and a ‘neutral’ SNP dataset without differentiated or phenotype-associated SNPs (**b**). **c**, Linear discriminant plots showing the first three axes of genetic variation, with individuals clustered by depth and site, showing the intermediate placement of gangfisch in between blaufelchen and weissfelchen. **d**, Principal component plots for PC1 to PC3 showing the mean \pm SEM for individuals by sampling site and spawning depth, demonstrating a similar populations structure as in **c** but without prior groupings **e**, Demographic models compared in *fastsimcoal2*: (SI) Strict isolation; (SCint) secondary contact and introgression; (SC_IM_int) secondary contact and introgression after isolation-with-migration of weissfelchen (WF) and gangfisch (GF) and introgression; (SCchange_int) ancient secondary contact without introgression and isolation with migration followed by introgression; (IM) isolation-with-migration; (IMint) isolation-with-migration and introgression; (IM_SC_int) IM with secondary contact between WF and GF and introgression. Results of model selection (natural logarithm of the maximum estimated likelihood (LnL) and the AIC) based on the full SNP dataset are given above each model.

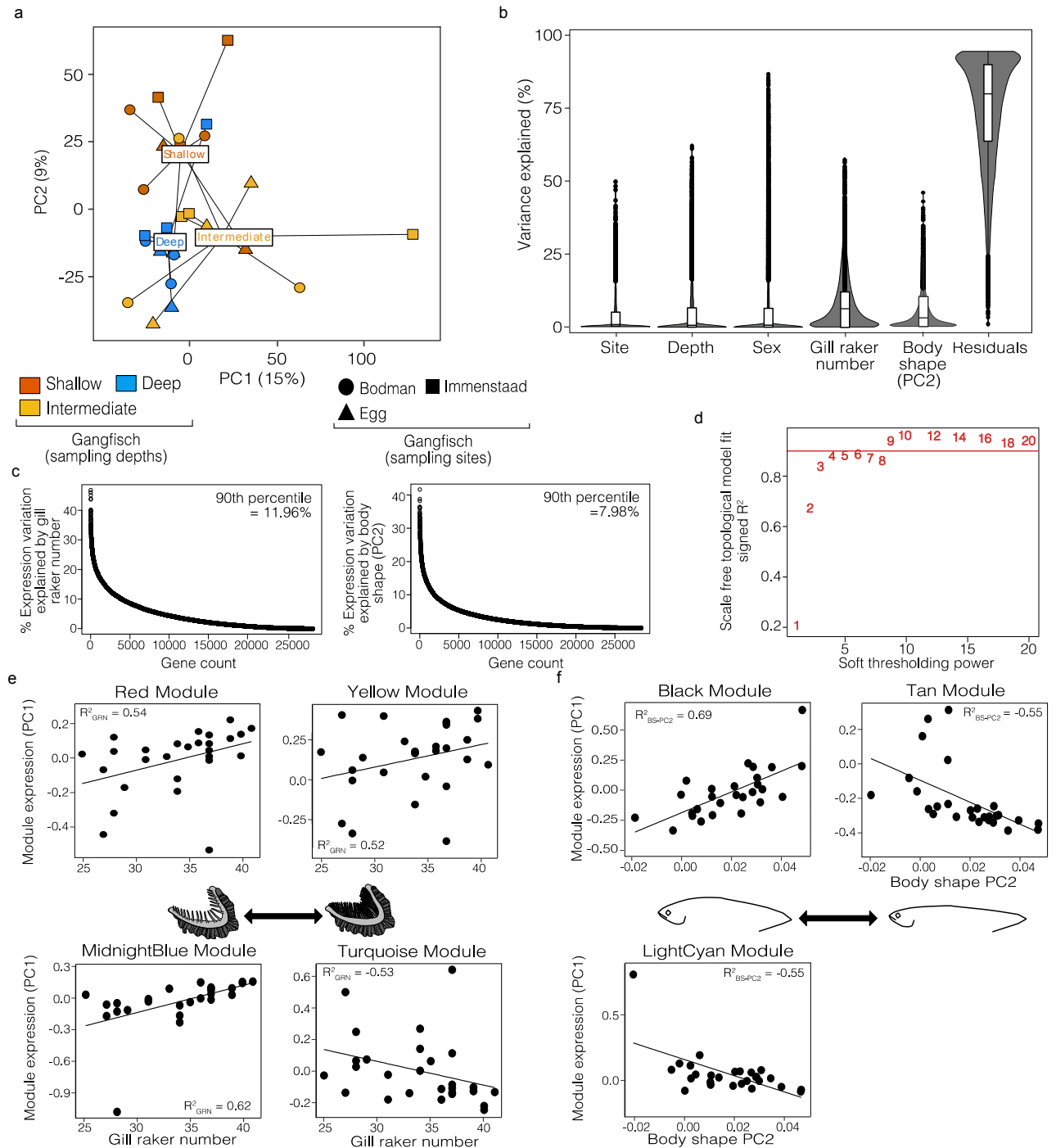


Supplementary Figure 5 | Genotype-phenotype association. **a**, Association of alleles at candidate loci with different gill raker numbers in gangfish. Individuals homozygous for opposite alleles (PP – homozygous pelagic allele; BB – Homozygous benthic allele) are on opposite ends of the gill raker spectrum, while heterozygous individuals are intermediate. ANOVA results for testing differences in gill raker numbers for each locus are shown above each plot. Candidate loci were determined based on genome-wide association with gill raker number. **b**, Signals of selection (as detected using pcadapt [$-\log_{10}(q\text{-value})$]) and association with gill raker number determined using *LFMM* and *BSLMM* in *GEMMA*. Effect sizes of loci are given by circle sizes (see legend) and the standardised Fst (ZFst) between benthic and pelagic gangfish (determined by genetic ancestry) are given by point colour (see legend). Significance thresholds for the different approaches are displayed by dashed lines. **c**, Signals of selection and association with body shape (PC2) for all loci.



Supplementary Figure 6 | Selection and genome scan results. **a**, Loci that were detected to be under selection (pcadapt) were also strongly differentiated between benthic and pelagic gangfish (highlighted in red). **b-d**, Genome scan between benthic and pelagic gangfish (pure genetic ancestry). **b**, Relationship between standardised Fst (ZFst) and delta nucleotide diversity ($\Delta\pi$). $\Delta\pi$ shows the difference in nucleotide diversity between benthic and pelagic gangfish, with negative $\Delta\pi$ meaning that nucleotide diversity is reduced at a particular locus in benthic gangfish and *vice versa*. Loci showing increased differentiation also show mostly reduced diversity in the trophic specialists, consistent with recent selective sweeps. **c**, Patterns of nucleotide diversity are highly correlated between benthic and pelagic gangfish, with most outlier loci (highlighted in red) showing reduced diversity in pelagic gangfish. **d**, Genome-wide distribution of Fst values, showing a subtle skew in the distribution. Overall, these results suggest that parts of the genome are differentiated due to

positive selection, mostly acting in pelagic individuals. **e**, Pairwise genetic differentiation between species (weissfelchen – WF; blaufelchen – BF) and gangfisch from different sites and depth (Bod – Bodman, Egg, Imm – Immenstaad), with spawning depths given by numbers behind site name. Significant F_{st} values ($P < 0.05$) are highlighted with an asterisk and comparisons between depths within sites are highlighted by black boxes. **f**, Shared and unique outlier loci between the ‘gangfisch benthic vs pelagic genome scan’ and ‘blaufelchen vs weissfelchen genome scan’. 20 of 165 outlier loci between pure genetic gangfisch clusters (benthic vs. pelagic) were also differentiated between weissfelchen and blaufelchen, suggesting shared selection pressures between the different species or differential sorting of shared polymorphisms within gangfisch. Unique outlier loci indicate unique responses to selection in the different comparisons.



Supplementary Figure 7 | Gene expression variation in gangfish. **a**, Transcriptional profiles of gangfish collected from all three sampling sites (Bodman represented as circles, Egg as triangles and Immenstaad as squares), and spawning depths in Lake Constance. Gangfish are coloured by spawning depth individuals were caught at (shallow in orange, intermediate in yellow, and deep in blue). Each point represents an individual (N=27, three per depth, per site), and the centroids for each depth group are plotted and labelled accordingly. Analysis was based on the subset of 31,872 genes with >20 read counts across all samples. **b**, Violin plots of percent expression variation explained by each trait specified in the LMM (site, depth, sex, gill raker number and body shape (PC2)) and the

residual variation. Gill raker number explained the greatest proportion of the overall gene expression variation with an average of 5.7%, followed by body shape (2.9%), with only negligible amounts of variation explained by site (1.3%), depth (1.8) and sex (0.7%). The majority of gene expression variation was attributed to residual factors (86.9%), as expected given high levels of individual variation and background differences present within natural environments, as well as the dynamic nature of gene expression (Ranz & Machado 2006; Gibson 2008). **c**, Distribution of expression variation explained by gill raker number (left) and body shape (right). The 90th percentile cut-off of expression variation for both traits are given in the corresponding figures. **d**, Choice of soft power for scale free network construction. The soft power was chosen according to the biologically motivated scale free topology criterion (Zhang and Horvath 2005): selecting the lowest integer above the model fit $R^2=0.9$. **e**, Association of module eigengene expressions with gill raker number (GRN) for the remaining four significant modules (N=27). See also Fig. 5b-c in the main text. R^2 -values for significant module-trait relationships are given in the figures. **f**, Association of module eigengene expressions with body shape (BS-PC2) for the remaining four significant modules (N=27). See also Fig. 5b,d in the main text. R^2 -values for significant module-trait relationships are given in the figures.

Supplementary Table 1 | Information on extinct and extant species of whitefish in Upper Lake Constance.

Subspecies	Niche	Historical GRN (mean) 1967 ^{14,15}	Contemp. GRN (mean) 2010/2014 ¹⁵	Spawning Time	Spawning Location
Kilch (<i>C. gutturosus</i>) *	Deep water, benthic feeder	17-25 (21)	No record	July - November	Deep water, >70m ¹⁴
Sandfelchen (<i>C. arenicolus</i>) **	Shallow water, benthic feeder	25-29 (-)	No record	Late January	Near shoreline in very shallow water ¹⁴
Gangfisch (<i>C. macrophthalmus</i>)	Pelagic, near shore, mostly zooplankton	33-43 (38)	24-44 (35) [†]	December – January	Near the shore, close to the ground ¹⁴
Blaufelchen (<i>C. wartmanni</i>)	Pelagic, open water, planktivore	30-40 (36)	31–36 (33)	Beginning of December	Middle of lake above deep water, 2-20m ¹⁴

Note: * Extinct or **very rare benthic species, [†]This study.

Supplementary Table 2 | Sampling sites, depths and sample sizes for different datasets.

Species	Sampling site	Spawning depth	Lat. (N)	Lon. (E)	Year	N (phen)	N (gen)	N (GEx)
Gangfisch	Bodman	2m	47°49'	09°02'	2014	30	17	3
	Bodman	25m			2014	30	16	3
	Bodman	50m			2014	30	17	3
	Egg	2m	47°41'	09°11'	2014	30	19	3
	Egg	25m			2014	30	18	3
	Egg	50m			2014	30	16	3
	Immenstaad	2m	47°39'	09°22'	2014	30	17	3
	Immenstaad	25m			2014	30	16	3
	Immenstaad	50m			2014	30	18	3
Blaufelchen	Romanshorn	<100m	47°59'	09°43'	2015	20	8	-
Weissfelchen	Untersee	10m	47°71'	09°02'	2015	20	9	-

Note: N (phen) gives the number of samples used for the phenotype analysis. N (gen) gives the number of samples that were genotypes using ddRADseq. N (GEx) gives the number of samples used for RNAseq.

Supplementary Table 3 | Inferred parameters (mean and 95% CI) of the most likely demographic model for the full SNP dataset.

Parameter	mean	2.5% CI	97.5% CI
Ancestral (Ne)	22452	14362	30814
BF (Ne)	2028	1768	2293
GF (Ne)	10968	9632	12189
WF (Ne)	1674	1491	1847
Adm (BF-GF)	0.78	0.64	0.90
Adm (WF - GF)	0.45	0.19	0.72
mig (BF-WF/GF)	1.18E-03	4.84E-04	2.29E-03
mig(BF-GF)	7.84E-03	6.63E-03	9.15E-03
mig(GF-WF)	7.71E-03	6.85E-03	8.73E-03
mig(BF-WF)	1.59E-05	3.35E-07	4.06E-05
Tint	139	83	207
Tdiv1	177	142	226
Tdiv2	6150	4275	8521

Note: Means and 95%-CI are given for each inferred model parameter. Ne – effective population size of ancestral population, blaufelchen (BF), gangfisch (GF) and weissfelchen (WF). Admixture proportion (Adm) from BF into GF and WF into GF. Bi-directional migration rates (mig) from BF into ancestral population of GF and WF, and all possible combinations. Tint – Timing of introgression from WF and BF into GF. Tdiv1 – Divergence time between GF and WF. Tdiv2 – Divergence time between BF and the common ancestor of GF and WF.

Supplementary Table 4 | Model selection and inferred mean parameter values of the most likely demographic model (Isolation-with-migration) for the filtered SNP dataset.

a.

Model	ln lhood	N	AIC	dAIC
parameters				
IM	-131189.44	10	262398.881	0
SCchange_int	-131188.19	14	262404.38	5.49939259
IMint	-131190.84	13	262407.671	8.79073313
IM_SC_int	-131190.59	16	262413.188	14.3071903
SC_IM_int	-131195.24	12	262414.476	15.5958185
SC_int	-131197.44	12	262418.884	20.0029664
SI	-131256.41	6	262524.813	125.932165

Note: IM – Isolation-with-migration; Int – Introgression; IMint – IM with introgression, IM_SC_int – IM with secondary contact between weissfelchen (WF) and gangfisch (GF) and introgression; SC_IM_int – secondary contact and introgression after isolation-with-migration of weissfelchen (WF) and gangfisch (GF) and introgression; SC_int – secondary contact and introgression; SI – strict isolation. Illustrations of all tested models can be found in Supplementary Fig. 4e.

b.

Parameter	mean
Ancestral (Ne)	62275
BF (Ne)	3089
GF (Ne)	155884
WF (Ne)	3713
mig (BF-WF/GF)	2.83E-04
mig(BF-GF)	1.43E-02
mig(GF-WF)	8.62E-03
mig(BF-WF)	2.43E-06
TDiv(GF-WF)	142
TDIV(BF-WF/GF)	6976

Note: Means are given for each inferred model parameter. Ne – effective population size of ancestral population, blaufelchen (BF), gangfisch (GF) and weissfelchen (WF). Admixture proportion (Adm) from BF into GF and WF into GF. Bi-directional migration rates (mig) from BF into ancestral population of GF and WF, and all possible combinations.

Supplementary Table 5 | Parameters describing the genetic architecture underlying variation in gill raker number and body shape.

Trait	h (95%-CI)	PVE (95%-CI)	PGE (95%-CI)	n_gamma (95%-CI)
GRN	0.99915 (0.9966-1.0)	0.99913 (0.9969-1.0)	0.97176 (0.9174-0.9995)	91 (68-117)
BS-PC2	0.8688 (0.4205-0.9999)	0.8636 (0.4216-0.9998)	0.2296 (0.0-0.8247)	47 (0-264)

Note: Parameters describing the genetic architecture underlying variation in gill raker number (GRN) and body shape PC2 (BS-PC2) were estimated using the BSLMM model implemented in GEMMA using a set of 13,589 SNPs (N=152). PVE – proportion of phenotypic variance explained by genotype; PGE – proportion of genetic variance explained by sparse (major) effects; n_gamma – the expected number of sparse (major) effect loci; h – approximate parameter used for the prior specification of the expected PVE.

Supplementary Table 6 | Summary statistics for loci associated with gill raker number and body shape.

Trait	SNP	p_wald	PIP	LFMM	pcadapt	Fst	ZFst	π (bn)	π (pl)
GRN	39045	1.46E-06	0.449	2.13E-04	no	0.221	4.037	0.503	0.250
	56834	4.55E-05	0.148	2.25E-05	yes	0.398	7.685	0.350	0.397
	77679	7.09E-05	0.073	2.13E-04	yes	0.624	12.36	0.391	0.196
	7642	8.39E-05	0.162	8.03E-05	no	0.228	4.194	0.160	0.497
	85946	3.16E-04	0.121	4.45E-03	no	0.207	3.751	0.509	0.294
	46306	3.45E-04	0.109	3.43E-04	no	0.083	1.205	0.475	0.484
BS	76778	5.25E-07	0.406	1.51E-01	no	0.0	-0.50	0.246	0.175
	59884	1.98E-05	0.037	1.18E-01	no	0.011	-0.28	0.068	0.167

Note: Summary statistics for all six loci associated with gill raker number, determined using a LMM (p_wald) and BSLMM (PIP) in GEMMA, and LFMM (q-value). Summary statistics for both loci associated with body shape (PC2) using a LMM (p_wald) and BSLMM (PIP) in GEMMA. Locus 56834 was also associated with body shape using LFMM (q-value). The remaining SNPs only associated with body shape and gill raker number in the LFMM analysis are not shown. Strength of association with body shape and GRN in the LFMM are shown in Figure 5c. Pcadapt indicates if a locus was found to be differentiate along PC1 in the pcadapt analysis. ZFst is the standardised Fst (Weir and Cockerham), π the nucleotide diversity of benthic (bn) and pelagic (pl) gangfish.

Supplementary Table 7 | Summary of significantly over-represented pathways identified for the nine modules in the WGCNA generated network that were significantly correlated with one or more of the measured traits (site, depth, genetic ancestry coefficient (GAC), gill raker number (GRN) and body shape (BS-PC2)). No significantly enriched pathways were identified for gene set comprising the black module. All P-values are Bonferroni corrected (FDR < 0.05). Asterisks indicate PANTHER pathways that were also associated with candidate loci (see Supplementary Table 8).

WGCNA Module	Significantly associated trait	Ecotype associated with	PANTHER Pathway	Pathway ID	P-value (FDR < 0.05)
Red	GRN	Pelagic	General transcription by RNA polymerase I	P00022	3.28E-02
Red	GRN	Pelagic	Heme biosynthesis	P02746	4.43E-02
Yellow	GRN	Pelagic	Cysteine biosynthesis	P02737	3.40E-02
Yellow	GRN	Pelagic	5HT3 type receptor mediated signaling pathway	P04375	3.18E-02
Yellow	GRN	Pelagic	5HT4 type receptor mediated signaling pathway	P04376	1.87E-02
Yellow	GRN	Pelagic	B cell activation*	P00010	3.71E-03
Yellow	GRN	Pelagic	T cell activation*	P00053	6.92E-03
Yellow	GRN	Pelagic	Apoptosis signaling pathway	P00006	3.16E-02
Yellow	GRN	Pelagic	CCKR signaling map*	P06959	9.56E-03
Brown	GAC, GRN	Pelagic	Integrin signalling pathway*	P00034	4.12E-03
MidnightBlue	GRN	Pelagic	Metabotropic glutamate receptor group I pathway	P00041	4.56E-02
MidnightBlue	GRN	Pelagic	Histamine H1 receptor mediated signaling pathway	P04385	5.41E-03
MidnightBlue	GRN	Pelagic	Thyrotropin-releasing hormone receptor signaling pathway	P04394	7.43E-03
MidnightBlue	GRN	Pelagic	Oxytocin receptor mediated signaling pathway*	P04391	7.43E-03
MidnightBlue	GRN	Pelagic	5HT2 type receptor mediated signaling pathway	P04374	8.78E-03
MidnightBlue	GRN	Pelagic	Endothelin signaling pathway*	P00019	2.74E-02
MidnightBlue	GRN	Pelagic	Apoptosis signaling pathway	P00006	4.43E-02
Turquoise	GRN	Benthic	Gonadotropin-releasing hormone receptor pathway*	P06664	1.25E-04
Turquoise	GRN	Benthic	TGF-beta signaling pathway*	P00052	1.10E-02
Turquoise	GRN	Benthic	SCW signaling pathway	P06216	1.46E-02
Turquoise	GRN	Benthic	GBB signaling pathway	P06214	1.46E-02
Turquoise	GRN	Benthic	DPP-SCW signaling pathway	P06212	1.46E-02
Tan	BS-PC2	Benthic	MYO signaling pathway	P06215	8.98E-03

Tan	BS-PC2	Benthic	ALP23B signaling pathway	P06209	1.79E-02
Tan	BS-PC2	Benthic	Activin beta signaling pathway	P06210	2.67E-02
Tan	BS-PC2	Benthic	SCW signaling pathway	P06216	3.54E-02
Tan	BS-PC2	Benthic	GBB signaling pathway	P06214	3.54E-02
Tan	BS-PC2	Benthic	DPP-SCW signaling pathway	P06212	3.54E-02
Tan	BS-PC2	Benthic	DPP signaling pathway	P06213	4.41E-02
Tan	BS-PC2	Benthic	BMP/activin signaling pathway	P06211	4.41E-02
Tan	BS-PC2	Benthic	Coenzyme A biosynthesis	P02736	4.41E-02
Tan	BS-PC2	Benthic	Cholesterol biosynthesis	P00014	5.41E-03
Tan	BS-PC2	Benthic	FAS signaling pathway	P00020	3.04E-02
LightCyan	BS-PC2	Benthic	T cell activation*	P00053	3.07E-02
LightCyan	BS-PC2	Benthic	Angiogenesis	P00005	2.96E-03
LightCyan	BS-PC2	Benthic	Integrin signalling pathway*	P00034	4.37E-03
Blue	GAC, BS-PC2	Benthic	P53 pathway feedback loops 1	P04392	4.18E-02
Blue	GAC, BS-PC2	Benthic	General transcription regulation	P00023	3.40E-02
Blue	GAC, BS-PC2	Benthic	EGF receptor signaling pathway*	P00018	2.24E-02

Supplementary Table 8 | Functional pathways (PANTHER) associated with candidate genes.

None of the pathways were significantly enriched.

PANTHER pathways associated with candidate loci	N.genes	p-value
Gonadotropin-releasing hormone receptor pathway	4	0.17
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	4	0.06
PDGF signaling pathway	4	0.05
Wnt signaling pathway	4	0.37
Alzheimer disease-amyloid secretase pathway	3	0.03
Cadherin signaling pathway	3	0.18
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	3	0.11
Inflammation mediated by chemokine and cytokine signaling pathway	3	0.50
Nicotinic acetylcholine receptor signaling pathway	3	0.07
EGF receptor signaling pathway	2	0.36
FGF signaling pathway	2	0.31
Integrin signalling pathway	2	0.69
Interleukin signaling pathway	2	0.20
TGF-beta signaling pathway	2	0.22
Alpha adrenergic receptor signaling pathway	1	0.21
Alzheimer disease-presenilin pathway	1	1.00
Axon guidance mediated by netrin	1	0.28
Axon guidance mediated by Slit/Robo	1	0.21
B cell activation	1	0.47
CCKR signaling map	1	1.00
De novo purine biosynthesis	1	0.25
DNA replication	1	0.23
Endothelin signaling pathway	1	0.54
Formyltetrahydroformate biosynthesis	1	0.08
Insulin/IGF pathway-protein kinase B signaling cascade	1	0.32
Metabotropic glutamate receptor group III pathway	1	0.47
Muscarinic acetylcholine receptor 1 and 3 signaling pathway	1	0.43
Notch signaling pathway	1	0.32
Oxytocin receptor mediated signaling pathway	1	0.41
p53 pathway	1	0.56
p53 pathway feedback loops 2	1	0.37
PI3 kinase pathway	1	0.40
Pyruvate metabolism	1	0.10
Ras Pathway	1	0.49
Salvage pyrimidine ribonucleotides	1	0.12
T cell activation	1	0.56

Note: candidate genes were selected based on gill raker number and/or body shape associated loci (N=117), detected to be under selection in PCAdapt analysis (N=19) and/or being significantly differentiated (N=171). Pathways were inferred using Panther. Pathways that were identified as over-represented in co-expression modules are in bold. N.genes – Number of genes from the respective pathway. Raw uncorrected p-value.

Supplementary references.

83. Gunter H.M., Fan S., Xiong F., Franchini P., Fruciano C. & Meyer A. Shaping development through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a cichlid fish. *Mol Ecol* **22**, 4516-4531 (2013).
84. Scythe. <https://github.com/vsbuffalo/scythe/>. Accessed January 2016.
85. Yip, A. M., and Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* **8**,1–14. (2007)
86. Seibt, C., Peeters, F., Graf, M., Sprenger, M., & Hofmann, H. Modeling wind waves and wave exposure of nearshore zones in medium-sized lakes. *Limnol. Oceanogr.* **58**, 23-36. (2013)