



# Genomic variation from an extinct species is retained in the extant radiation following speciation reversal

David Frei<sup>1,2</sup>, Rishi De-Kayne<sup>1,2</sup>, Oliver M. Selz<sup>1,2</sup>, Ole Seehausen<sup>1,2</sup> and Philine G. D. Feulner<sup>1,2</sup>✉

**Ecosystem degradation and biodiversity loss are major global challenges. When reproductive isolation between species is contingent on the interaction of intrinsic lineage traits with features of the environment, environmental change can weaken reproductive isolation and result in extinction through hybridization. By this process called speciation reversal, extinct species can leave traces in genomes of extant species through introgressive hybridization. Using historical and contemporary samples, we sequenced all four species of an Alpine whitefish radiation before and after anthropogenic lake eutrophication and the associated loss of one species through speciation reversal. Despite the extinction of this taxon, substantial fractions of its genome, including regions shaped by positive selection before eutrophication, persist within surviving species as a consequence of introgressive hybridization during eutrophication. Given the prevalence of environmental change, studying speciation reversal and its genomic consequences provides fundamental insights into evolutionary processes and informs biodiversity conservation.**

A mechanistic understanding of species extinction is critical for a better understanding of contemporary patterns of biodiversity as well as for predicting its future<sup>1</sup>. Extinction can result from demographic decline, from loss of reproductive isolation or from a combination of both<sup>2,3</sup>. When extinction involves the loss of reproductive isolation<sup>4,5</sup>, the extinction process can leave a lasting legacy in the genomes of surviving species through introgressive hybridization<sup>2,6</sup>, potentially even influencing species that will only emerge in the future<sup>7</sup>. When the loss of reproductive isolation contributes to extinction and some of the taxa involved in introgressive hybridization survive, parts of the evolutionary history of extinct species persist and might affect future dynamics, although species extinction is functionally complete. Previous studies have identified examples of genomic variation in extant species that originated from extinct species<sup>8–11</sup>. However, apart from these few examples, genomic information for extinct species is still rare<sup>12</sup>. As a result, the extent and the evolutionary importance of genetic transfer from extinct to extant species could be underestimated.

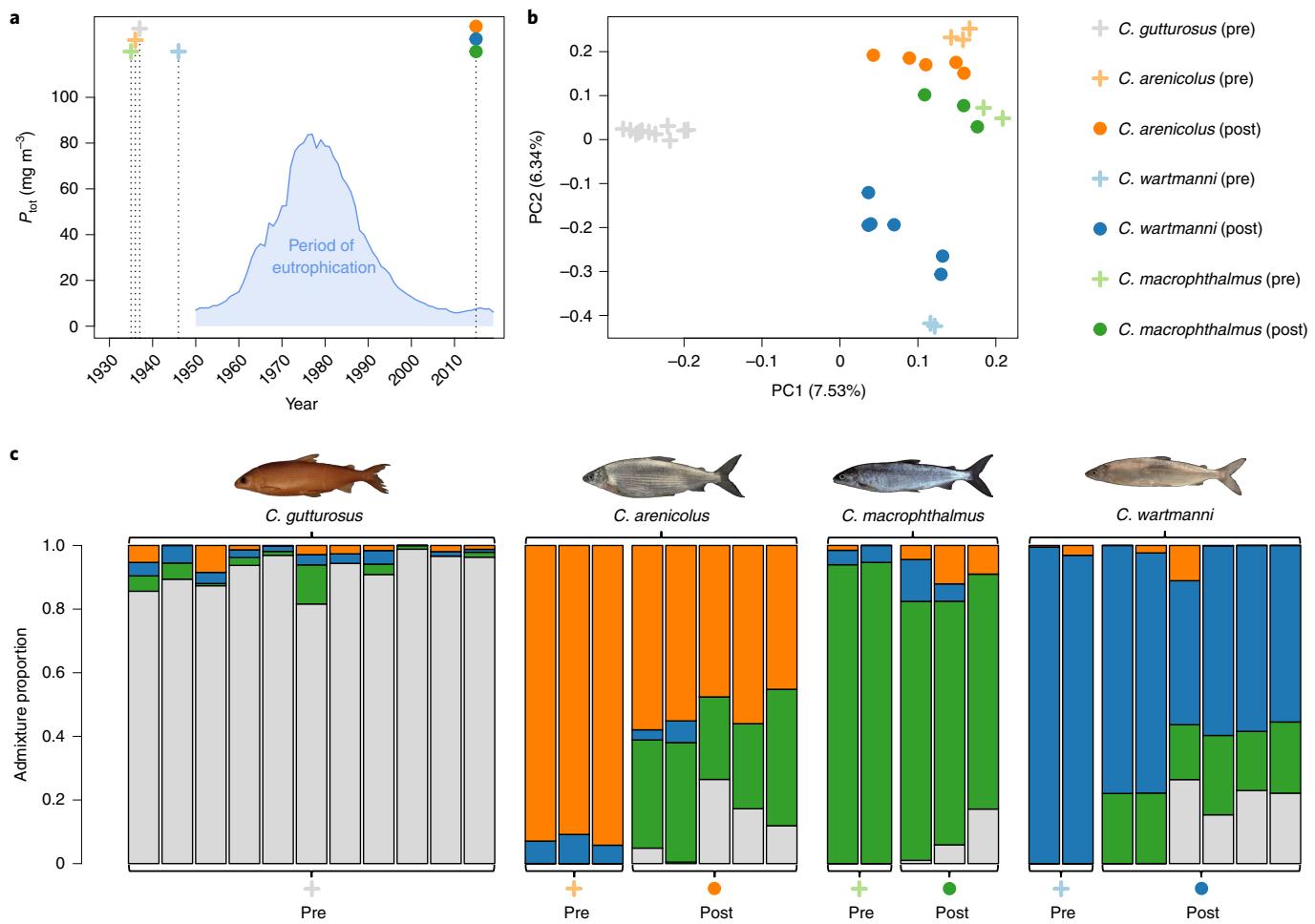
Ecological speciation, the process by which reproductive isolation evolves in response to divergent ecological selection or ecologically mediated divergent sexual selection<sup>13,14</sup>, is an important process in the evolution of a substantial proportion of contemporary eukaryotic species diversity<sup>15,16</sup>. In early stages of ecological speciation, species differentiation is maintained by prezygotic and/or extrinsic postzygotic reproductive isolation mechanisms, both mediated by ecology, while genetic incompatibilities remain weak or absent<sup>14–18</sup>. Ecologically mediated reproductive isolation, both prezygotic and postzygotic, results from performance trade-offs between, and adaptation to, alternative fitness optima<sup>13,14</sup>. When environments change, fitness optima shift and may converge. This can lead to a weakening or complete loss of prezygotic reproductive isolation between species and a relaxation of divergent selection,

weakening extrinsic postzygotic isolation. The breakdown of reproductive isolation might culminate in the collapse of sympatric species into hybrid populations<sup>19,20</sup>, a process called speciation reversal, potentially resulting in the sudden and rapid extinction of species through introgressive hybridization<sup>4,5</sup>.

Concerningly, contemporary extinction rates caused by speciation reversal through anthropogenic homogenization of environments are likely to be faster than rates of extinction by demographic decline alone<sup>5</sup>. Whilst the potentially widespread impacts of speciation reversal on contemporary biodiversity loss are still underappreciated in conservation<sup>4</sup>, its genomic consequences are still underappreciated in evolutionary biology. Genetically admixed hybrid populations that emerged from speciation reversal might have enhanced evolvability<sup>21</sup>. In the future, such populations may adapt in new and unexpected ways<sup>22,23</sup>, expand their ranges<sup>24</sup> and even seed further species diversification<sup>25</sup>. A deeper understanding of causes and consequences of extinction by speciation reversal is therefore needed to determine the immediate as well as the long-term influence of anthropogenic environmental change on biodiversity, to enhance nature conservation measures and improve policy (hybrid populations are in some countries still considered unworthy of protection) and to advance our comprehension of evolutionary dynamics in changing environments.

The evolutionarily young Alpine whitefish radiation provides an excellent system in which to study ecological speciation and the consequences of its reversal<sup>26–28</sup>. Across the large pre-Alpine lakes of Switzerland more than 30 endemic whitefish species have evolved since the end of the last glacial maximum<sup>27,29–32</sup>. As the water depth of spawning grounds represents one important axis of Alpine whitefish species differentiation, reproductive isolation among sympatric species may often depend on the persistence of fine-scale depth-related differences between spawning habitats<sup>31</sup>. Therefore,

<sup>1</sup>Department of Fish Ecology and Evolution, Center for Ecology, Evolution and Biogeochemistry, Eawag—Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland. <sup>2</sup>Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland. ✉e-mail: [philine.feulner@eawag.ch](mailto:philine.feulner@eawag.ch)



**Fig. 1 | Partial loss of genetic differentiation between Lake Constance whitefish species during eutrophication-induced speciation reversal.** **a**, Total phosphorus concentrations in Lake Constance over time as a proxy for severity of eutrophication. Time points when whitefish were sampled are indicated by dotted lines with crosses (pre-eutrophication) and circles (post-eutrophication). **b**, In genomic PCA space, the same species post-eutrophication (circles) are less distinct than pre-eutrophication (crosses) and one species is completely lost (*C. guttatus*). **c**, Estimated admixture proportions grouped by species and whether collected pre- or post-eutrophication. Post-eutrophication samples show consistently more admixture than pre-eutrophication samples. The four whitefish species are indicated by colours: grey, *C. guttatus*; orange, *C. arenicola*; blue, *C. wartmanni*; green, *C. macrophthalmus*.

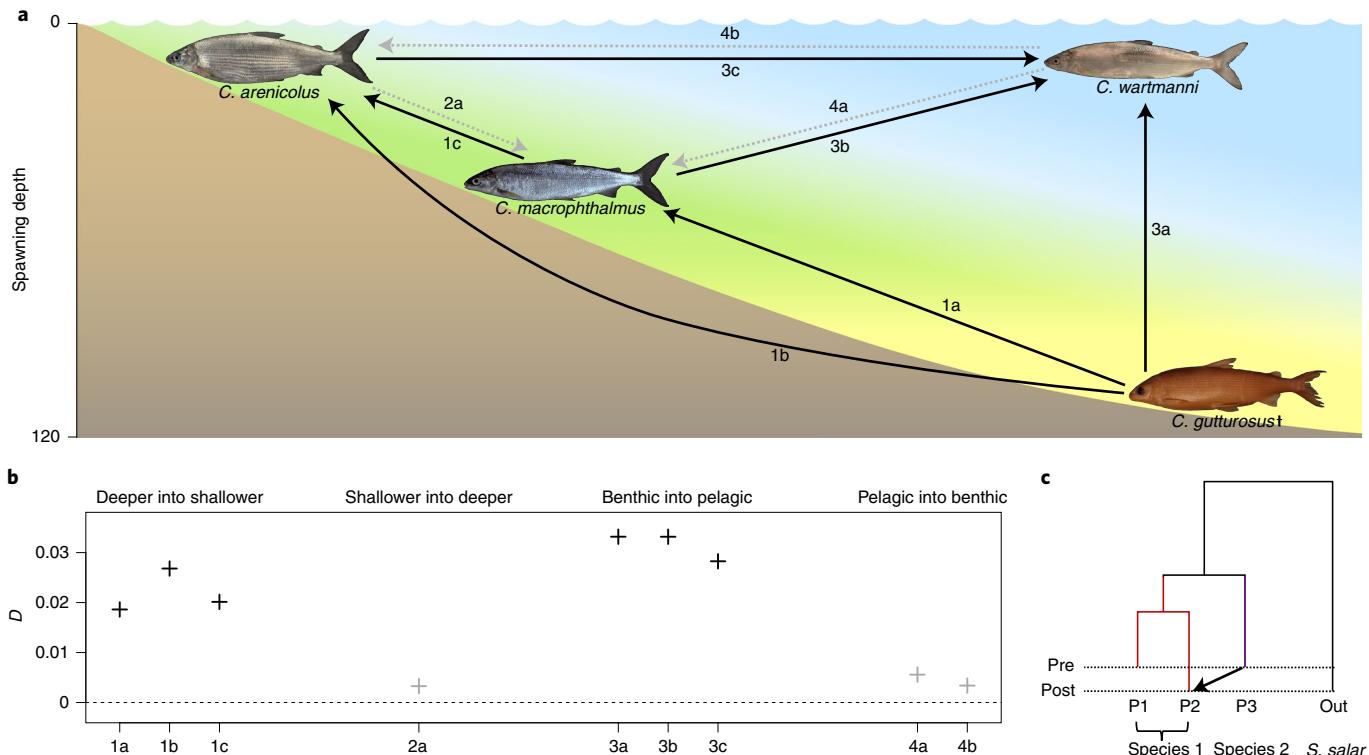
Alpine whitefish species are highly sensitive to speciation reversal when habitat diversity and suitability along the lacustrine water depth gradient changes<sup>31,33,34</sup>. Anthropogenic eutrophication during the twentieth century led to the loss of deep-water spawning habitats, reducing prezygotic isolation between sympatric whitefish species. At the same time, eutrophication changed the abundance ratios between prey types, possibly resulting in the loss of extrinsic postzygotic isolation through relaxed divergent selection between feeding niches<sup>3</sup>. The combination of reduced prezygotic reproductive isolation and weakened divergent selection between niches led to speciation reversal through introgressive hybridization and, in combination with demographic decline of those species whose niches shrank, resulted in dramatic losses of Alpine whitefish diversity<sup>3,26</sup>.

Speciation reversal is most comprehensively documented in the Lake Constance whitefish radiation, which originally consisted of four endemic sympatric species but, with the extinction of the profundal *Coregonus guttatus*, now comprises only three extant species (Fig. 1 and Extended Data Fig. 1). Previous work showed a substantial decline in both neutral genetic and functional morphological differentiation between all three extant whitefish species, indicating a partial breakdown of reproductive isolation<sup>3</sup>. Additionally, five private microsatellite alleles of the extinct species

were discovered in all extant species after eutrophication, consistent with speciation reversal through introgressive hybridization<sup>3</sup>. We here provide a genome-wide perspective of environmental change-induced speciation reversal that affected an entire whitefish radiation by comparing whole-genome resequencing data of pre- and post-speciation reversal populations of all species in the radiation. We reveal the radiation-wide pattern of introgression during speciation reversal and demonstrate that the extinct species introgressed into all extant species. Introgression from the extinct species included genomic variation shaped by positive selection before eutrophication, indicating that these regions were potentially adaptive in the extinct species before speciation reversal.

## Results

**Genomic differentiation weakened across the entire radiation.** Speciation reversal may cause sudden and rapid collapses of entire radiations within only few generations<sup>3,19,20</sup>. Before the ecosystem changes during the twentieth century (Fig. 1a), the four Lake Constance whitefish species, including the now extinct *C. guttatus*, formed well-defined species clusters within a multidimensional genotype space<sup>35</sup> (Fig. 1b) based on genotype likelihoods of 222,017 polymorphic sites, despite complete sympatry. An analysis



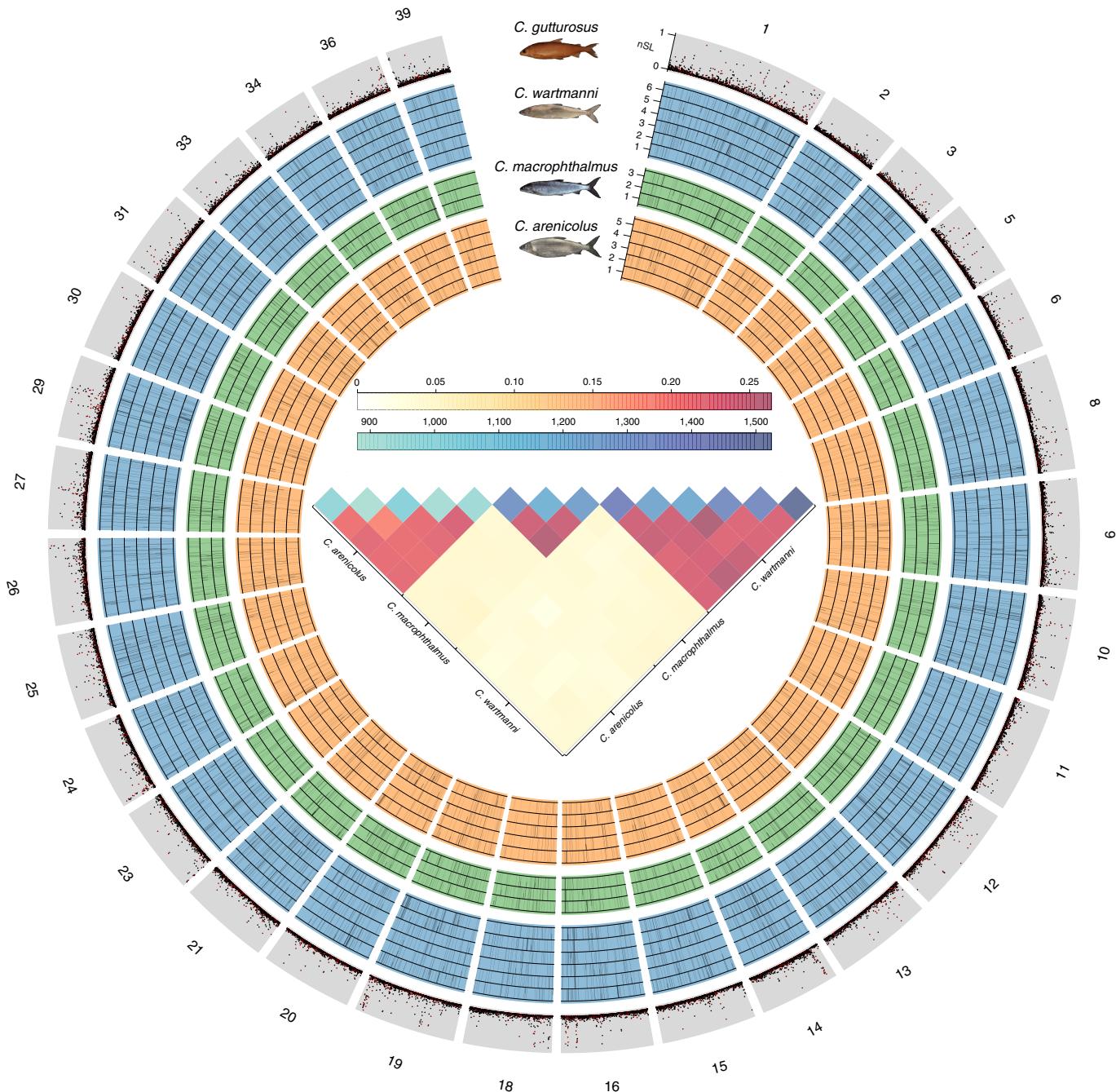
**Fig. 2 | Directionality of introgression during speciation reversal.** **a**, Schematic representation of spawning habitat (water depth and benthic or pelagic habitat) of four Lake Constance whitefish species and the directionality of introgression. Spawning depth is shown in metres. Significant tests for introgression during speciation reversal are indicated as black arrows and non-significant tests as dashed grey arrows. The labelling of the arrows (1a, 1b, 1c, 2a, 3a, 3b, 3c, 4a, 4b) corresponds to the tests for introgression shown in **b**. Severity of niche collapse is indicated by yellow (highest) to blue (lowest) shading of the water. The extinct species is denoted with †. **b**, *D* values for each test for introgression, grouped by contrasts among reproductive ecology (significant values are shown as black crosses, non-significant values as grey crosses; Extended Data Fig. 4). **c**, The topology shows the grouping of species for the underlying *D*-statistic test.

of population structure<sup>35</sup> (Fig. 1c and Extended Data Fig. 2) confirms four distinct genetic clusters for pre-eutrophication samples (Fig. 1c ‘pre’) but reveals that post-eutrophication individuals of all three extant species (Fig. 1c ‘post’) are strongly admixed. Our results are in line with previous work based on ten microsatellite markers (Extended Data Fig. 3) that demonstrated a rapid reduction of genetic differentiation (global  $F_{ST}$  decreased over twofold<sup>3</sup>) between these whitefish species by comparing samples collected more than 40 years apart and separated by a period of anthropogenic lake eutrophication<sup>3</sup>. Our new results based on whole-genome resequencing data demonstrate a dramatic genome-wide reduction of genetic differentiation amongst all species following the period of eutrophication (Fig. 1 and Extended Data Fig. 3), matching the prediction of relaxed reproductive isolation during speciation reversal.

**Directionality of introgression mirrors niche collapse.** By comparing whole-genome sequence information obtained from historical samples with that from contemporary samples, we were able to formally test whether introgression had occurred during the eutrophic phase and identify the specific direction of such introgression (see Fig. 2c for a generic topology) using an extended version of *D*-statistics that allows us to include multiple individuals per population<sup>36</sup>. We found that significant introgression had occurred from deeper into shallower spawning species but not in the opposite direction (Fig. 2a and Extended Data Fig. 4). Further, we identified introgression from benthic species into one species occupying a pelagic reproductive niche but no introgression was detected from the pelagic into either of the benthic spawning species (Fig. 2a and Extended Data Fig. 4). Eutrophication of Lake Constance resulted

in the loss of deep-water spawning habitats as consequence of decreased oxygen concentrations at the water–sediment interface, the location of whitefish egg development<sup>37,38</sup>. Whereas low oxygen conditions in deeper benthic areas probably prevented successful reproduction of *C. gutturosus* and contributed to its extinction<sup>37,39</sup>, shallower benthic spawning habitats might have been less severely affected and recovered quickly enough after restoration of oligotrophic conditions to allow *C. macrophthalmus* and *C. arenicola* to survive<sup>39</sup>. Although its recruitment was affected by low oxygen conditions, the pelagic spawning *C. wartmanni* expanded its spawning grounds during the eutrophic phase of the lake<sup>37,39</sup>. At the same time, increased productivity during eutrophication led to an increase of zooplankton density<sup>40</sup> and decreased zoobenthos densities<sup>41–43</sup>, possibly relaxing divergent selection between feeding niches<sup>3</sup>. Our data therefore uncover a directionality of introgression that mirrors the severity of reproductive niche collapse caused by anthropogenic lake eutrophication and is consistent with major changes in the selective regime during the eutrophic phase of the lake.

Quantification of the extent of introgression in individual genomes is needed to assess the fraction of an extinct species’ genome that persists in extant species as a consequence of speciation reversal. We used topology weighting by iterative sampling of subtrees<sup>44</sup> to explore evolutionary relationships in 50 kilobase (kb) windows along the genome to find regions where an introgression topology was most supported. Across all 14 contemporary individuals combined, ~22% of the evaluated 31,476 windows along the genome showed signatures of introgression from the extinct *C. gutturosus* (Fig. 3). On the basis of a rarefaction analysis<sup>45</sup> on windows indicating signals consistent with introgression, we estimated that



**Fig. 3 | Genomic distribution and characterization of introgression derived from extinct *C. gutturosus*.** Each of the three inner tracks corresponds to a species (blue, *C. wartmanni*; green, *C. macrophthalmus*; orange, *C. arenicola*) and each track is subdivided into individual genomes. Each black bar corresponds to one introgressed window in one individual. The outermost track summarizes a selection scan with nSL in the extinct *C. gutturosus* (windows that introgressed are shown as red dots, non-introgressed windows as black dots), indicating that regions that were under positive selection in *C. gutturosus* have often introgressed into contemporary species. The heatmap in the centre shows the proportion of shared introgessed windows between individuals (pairwise comparison yellow to red colour scale) and the absolute count of introgressed windows for each individual (blue colour scale).

~28% of the total genome of *C. gutturosus* is still maintained and segregating within the three extant species, with different subsets of windows introgressed by *C. gutturosus* in each species (~14% in *C. wartmanni*, ~12% in *C. macrophthalmus* and ~11% in *C. arenicola*; Extended Data Fig. 5). Alternative approaches resulted in very similar approximations of *C. gutturosus* admixture proportions in the three contemporary species (Extended Data Fig. 6). Windows showing an introgression signature were more frequently shared

between individuals of the same species (Fig. 3) than between individuals of different species ( $t=57.18$ ,  $P<0.01$ , d.f. 29.34). This distribution pattern suggests that some reproductive isolation between the three extant whitefish species has persisted during speciation reversal, in agreement with diminished but sustained genetic (Fig. 1) and morphological (Extended Data Fig. 7) differentiation. The distribution of introgressed genomic windows along genomes of the three surviving species implies that introgression occurred directly

from the extinct species into each extant species, as the potentially introgressed windows in these species do not form subsets of each other (Fig. 3). Independent introgression from *C. gutturosus* into all other members of the radiation highlights the sensitivity of reproductive isolation to environmental change in adaptive radiations.

**Exchange of adaptive variation during speciation reversal.** Speciation reversal might transfer entire chromosomal segments containing intact regions shaped by selection between hybridizing species<sup>26</sup>. We performed a selection scan using the haplotype-based statistic nSL<sup>46</sup> to determine whether genomic regions with signatures of positive selection in the now extinct profundal *C. gutturosus* introgressed into extant whitefish species during speciation reversal. We considered the highest 1% fraction (315 50-kb windows) of regions showing signals of positive selection (Tajima's *D* based on genotype likelihoods<sup>47</sup>; in this top 1% of windows is significantly different from the rest of the genome; Extended Data Fig. 8) as potentially having conferred adaptation to profundal habitats in *C. gutturosus* (see Supplementary Table 1 for functional enrichment of genes in those regions, which revealed a link to the regulation of platelet aggregation and the organization of the photoreceptor cell outer segment amongst various others functions). Of these putatively selected regions, 53.3% have introgressed from *C. gutturosus* into extant whitefish species (Fig. 3). Across all individuals of the extant species, introgressed regions were enriched for genomic windows that carry signatures indicative of positive selection in the extinct *C. gutturosus* ( $P < 0.01$  with 10,000 permutations). This suggests that, after introgression, such regions have not been under negative selection and some might even have been favoured in their new bearers, although the time span after introgression is probably too short to yet leave any distinct signatures of selection. We observed no difference in gene density between introgressed and non-introgressed regions (Extended Data Fig. 9). Both the introgression of potentially adaptive variation and no evidence that introgression is confined to gene-poor regions suggest that there was no strong selection against introgressed variants from *C. gutturosus*. This pattern is consistent with the hypothesis of relaxed divergent selection during speciation reversal<sup>3,4,26</sup> and suggests that genetic incompatibilities between these species were relatively weak. While those introgressed variants may behave as neutral in the niches of the other species although they have been under positive selection in the extinct species before eutrophication, the resulting polymorphisms may fuel the extant species with evolutionary potential to recolonize the lost niche after ecosystem restoration.

## Discussion

Since species diversity can evolve in response to heterogenous environments, the homogenization of environments can drive species extinction<sup>19</sup>. Conservation biology traditionally relies on understanding the demographic consequences of such habitat change. However, species diversity collapse can be greatly accelerated when changes to natural habitats lead to shifts in evolutionary forces such that ecologically mediated reproductive isolation between otherwise coexisting species is lost. In such situations, entire adaptive radiations may collapse into hybrid populations, resulting in dramatic losses of biodiversity within very few generations through speciation reversal<sup>3,19</sup>. Relaxation of reproductive isolation between all four species in the radiation of Lake Constance whitefish has led to such speciation reversal, with the extinction of one species and diminished genetic differentiation among all others. Our data reveal evidence for introgression between all species of the radiation, including introgression from the extinct *C. gutturosus* into all extant species, associated with a transient period of eutrophication and associated degradation of habitat niches.

Speciation reversal resulted in the persistence of considerable fractions of genomic variation derived from the extinct *C. gutturosus*

within extant species. Partial genomic survival of taxa despite being functionally extinct as species has been recently described as well in, for example, elephants<sup>11</sup>, apes<sup>10</sup> and bears<sup>9</sup>, although the evolutionary processes resulting in the persistence of ancient alleles often remain unclear. Here, we demonstrate that during extinction by speciation reversal there was substantial and widespread introgression of potentially adaptive variation from the extinct *C. gutturosus* into all three extant species, resulting in the persistence of a considerable fraction of its gene pool. If extinction occurred by demographic decline alone, all alleles characteristic of the extinct species would have been completely lost. However, speciation reversal culminated in the rescue of genomic variation that had evolved in the extinct species before eutrophication, thereby preserving fractions of its evolutionary legacy from being lost forever.

Today, oligotrophic conditions of Lake Constance have been largely restored and deep-water habitats are again accessible for fish<sup>48</sup>. Nonetheless, profundal regions remain devoid of whitefish<sup>49</sup>. Theoretical work has suggested that when disturbance of reproductive isolation is short and transient, species pairs that collapsed may re-emerge after restoration of environmental conditions favourable for speciation<sup>50</sup>. However, re-emergence appears less likely the more species are involved in hybridization during the collapse of reproductive isolation and the timescale in which re-emergence might happen is orders of magnitudes larger than it takes to collapse species into hybrid populations during disturbances. In terms of whitefish generations, the eutrophic phase of Lake Constance was of relatively short duration (~30 yr or about six whitefish generations<sup>51</sup>) and, thus, the re-emergence of a deep-water ecomorph in the distant future is not to be ruled out, highlighting that the conservation of hybrid populations can be important.

As most environments have continuously changed, even via natural processes (albeit the rate of change has massively accelerated under recent anthropogenic impact) and since many species are sensitive to hybridization-mediated evolutionary dynamics<sup>5,52</sup>, speciation reversal might be an important but underappreciated evolutionary pathway when environments change. In the context of adaptive radiations, reassembling of genomic variation derived from admixture between distinct parental lineages into new adaptive combinations of genotypes can accelerate adaptation and speciation<sup>53</sup>. Therefore, speciation reversal could potentially facilitate adaptation and diversification in response to changing or even entirely new environments in the future. Thus, our increasingly detailed understanding of both short- and long-term consequences of speciation reversal will advance our understanding of the evolution of biodiversity, especially its dynamics under environmental change, whilst also requiring us to adjust our approaches in conservation biology.

## Methods

**Sample collection.** Historical whitefish scale samples, assembled by D. Bittner (see Vonlanthen et al.<sup>3</sup> for details) and collected before the onset of eutrophication in the upper basin of Lake Constance (Fig. 1a), were used to extract DNA from 2–11 individuals of each of four species (*C. areniculus* ( $n=3$ ), *C. gutturosus* ( $n=11$ ), *C. macropthalmus* ( $n=2$ ) and *C. wartmanni* ( $n=2$ )). The contemporary individuals used were caught by local fishermen during the spawning season of 2015 on known whitefish spawning grounds (*C. areniculus* ( $n=5$ ), *C. macropthalmus* ( $n=3$ ) and *C. wartmanni* ( $n=6$ )), using gill-nets with varying mesh sizes. Individuals were anaesthetized and subsequently euthanized using appropriate concentrations of tricaine methane sulfonate solutions (MS-222) according to the permit issued by the cantons of Zurich and St.Gallen (ZH128/15). Fin-clips were taken and stored in 100% analytical ethanol until extraction of DNA. Contemporary samples were phenotypically assigned to species by external morphology and assignments were confirmed by morphometrics, using morphological measurements following Selz et al.<sup>30</sup> (Extended Data Fig. 7). The phosphorus data (yearly averaged total phosphorus) were retrieved from Bodensee-Wasserinformationsystem (BOWIS)—data from the Lake Constance Water Information System of the International Commission of Lake Constance Water Conservation (Internationale Gewässerschutzkommission für den Bodensee, IGKB).

**DNA extraction.** DNA extractions of both historical scale samples and recent fin-clip samples were done using the Qiagen DNeasy blood and tissue kit (Qiagen AG). For scale samples, we followed the manufacturer's supplementary protocol for crude lysates (<https://www.qiagen.com/at/resources/resourcedetail?id=ad5ef878-8327-4344-94ad-a8e703e62b49&lang=en>) with the following minor adjustments: an alternative lysis buffer containing 4 M urea<sup>54</sup> and elongated incubation time (overnight) at 37°C were used for lysis of five scales per individual before the DNA extraction. To ensure that no contamination with external sources of DNA was present, we included a negative control in each batch of scale extractions. Negative controls always resulted in no detectable DNA concentrations, while the historical scale extractions resulted in DNA concentrations ranging between 1.12 and 70.2 ng  $\mu$ l<sup>-1</sup>. Fin-clips of contemporary individuals were extracted following the standard protocol supplied by the manufacturer.

After extraction, we measured DNA fragmentation on an Agilent TapeStation 2200 (Agilent Technologies AG) on either D5000 (historical scale samples) or genomic DNA (recent fin-clip samples) screen tapes. DNA concentration was quantified on a Qubit 2 fluorometer (Thermo Fisher Scientific AG) using the manufacturer's high-sensitivity assay kit. Contamination of DNA samples was measured on a NanoDrop 1000 (Thermo Fisher Scientific AG).

**Library preparation and sequencing.** For each individual whitefish scale sample, one Illumina paired-end TruSeq DNA Nano library (Illumina GmbH) was produced, while an Illumina paired-end TruSeq DNA PCR-Free library (Illumina GmbH) was prepared for each contemporary fin-clip sample. Library preparation was done by the next-generation sequencing (NGS) platform of the University of Bern following the manufacturer's instructions. Three of the historical scale samples failed in the first round of library preparation, indicated by a high amount of adaptor dimers relative to the DNA template concentration. For these samples, the standard library preparation protocol was repeated without the shearing step, decreasing the amounts of adaptor dimers.

Libraries from historical scale samples and contemporary fin-clip samples were prepared according to Extended Data Fig. 10 and sequenced 2  $\times$  150 paired end on either HiSeq 3000 or Novaseq 6000.

**Mapping and filtering of sequencing reads.** Poly-G strings at the end of the reads were removed using fastp<sup>55</sup> v.0.20.0. Overlapping paired-end reads with total length  $>25$  base pairs (bp) were merged using SeqPrep v.1.0 (<https://github.com/jstjohn/SeqPrep>). Raw reads were aligned to the Alpine whitefish genome assembly<sup>56</sup> (ENA accession [GCA\\_902810595.1](https://www.ncbi.nlm.nih.gov/assembly/GCA_902810595.1)) with bwa mem<sup>57</sup> v.0.7.12 and adjusting the 'r' parameter to 1 (increasing accuracy of alignment but reducing computational speed). Duplicated reads were marked with MarkDuplicates, mate information was fixed with FixMateInformation and read groups were replaced with AddOrReplaceReadGroups from picard-tools v.2.20.2 (<http://broadinstitute.github.io/picard/>).

**Population genomic analysis.** Due to differences in sequencing depth (mean coverage of 6.3 $\times$  for historical samples and mean coverage of 22.1 $\times$  for contemporary samples at polymorphic sites included in downstream analyses; Extended Data Fig. 10) and, to account for possible sequencing errors, we avoided genotype calling whenever possible and only analysed whitefish chromosomes without any potentially collapsed duplicated regions<sup>56</sup>. Instead of hard genotyping, we calculated genotype likelihoods<sup>58</sup> and minor allele frequencies<sup>59,60</sup> at polymorphic sites applying the samtools genotype-likelihood model<sup>58</sup> implemented in angsd<sup>61</sup> v.0.925. Only sites covered with at least two reads from every individual (no missing data), passing a P-value cut-off of  $\times 10^{-6}$  for being variable<sup>59</sup> and having not more than two different alleles were included. Reads that did not map uniquely to the reference and had a mapping quality  $<30$ , as well as bases with quality score  $<20$ , were not considered for calculation of genotype likelihoods in the following analyses. We used the following P-value cut-offs for single-nucleotide polymorphism (SNP) filters implemented in angsd v.0.925: -sb\_pval 0.05 -qscore 0.05 -edge\_pval 0.05 -mapq\_pval 0.05, resulting in 477,981 sites.

We performed a principle component analysis (PCA) on all polymorphic sites with a minor allele frequency  $>5\%$  (222,017 sites) and including all individuals and estimated population structure based on the three most important eigenvectors with PCAngsd<sup>35</sup> v.0.98 and default parameters (see Extended Data Fig. 2 for log-likelihoods of K = 1–7). Typically, ancient samples are shifted towards the centre of the PC space in relation to modern samples of the same populations<sup>62</sup>. Also, samples sequenced at lower depths<sup>35</sup> or having increased missing data<sup>63</sup> tend to be shifted to the centre of PC axes. Here, we observe the opposite pattern, since our historical samples are shifted towards the extremes of the PC space compared to our contemporary samples (as we would expect when these species have recently hybridized), increasing our confidence that we can draw robust and biologically meaningful conclusions from the PCA analysis and from our data.

We assessed the change in genetic differentiation across all species of the radiation during the eutrophication period and then compared the obtained values from our SNP data to the global  $F_{ST}$  estimates from Vonlanthen et al.<sup>3</sup>, which are based on ten microsatellite markers (Extended Data Fig. 3). We used beagle<sup>64</sup> v.4.1 to infer genotypes from the genotype likelihoods at the 477,981 polymorphic sites produced in angsd<sup>61</sup> v.0.925 from above and calculated  $F_{ST}$  estimates across all three contemporary species pre- and post-eutrophication with the R package hierfstat<sup>65</sup>

v.0.5–7 and additionally calculated the same estimate including our sample of the extinct *C. gutturosus* population collected pre-eutrophication.

To formally test for introgression between all species of the Lake Constance whitefish radiation during eutrophication, we performed ABBA BABA tests based on genotype likelihoods at all 477,981 sites inferred to be polymorphic within our whitefish dataset with the angsd<sup>61</sup> v.0.925 option 'doAbbababa2', using multiple individuals per population<sup>36</sup>. The ABBA BABA test requires four populations in the following order: ((P1,P2)P3)O. We used the pre- and post-eutrophication populations of one extant species as focal test populations (P1 and P2; Fig. 2c) and then tested for introgression into this species from all possible donor species (P3; Fig. 2c). By this assignment of populations to P1, P2 and P3 we could test for introgression that must have happened during eutrophication, respectively, during speciation reversal, as well as assess the directionality of introgression within the whole radiation. A *Salmo salar* individual (short read archive accession number SRR3669756) from Kjaerner-Semb et al.<sup>66</sup> served as outgroup, which defines the ancestral allele (A). We used a block-jackknife approach implemented in angsd<sup>61</sup> v.0.925 with a block size of 5 megabases (Mb) to assess the significance of potential excesses of ABBA or BABA sites.

To visualize general relationships among the four studied species, we produced a maximum likelihood phylogeny using RAxML<sup>67</sup> v.8.2.12. We first calculated genotype likelihoods of the *S. salar* outgroup at all 477,981 polymorphic sites with angsd<sup>61</sup> v.0.925 and then inferred genotypes of all individuals including the outgroup and phased these using beagle<sup>64</sup> v.4.1. We then thinned this dataset using VCFtools<sup>68</sup> v.0.1.16 so that all SNPs were at least 500 bp apart from each other, and then filtered the resulting dataset with bcftools v.1.10.2 (<https://github.com/samtools/bcftools>) to contain only sites that are homozygous for the reference and homozygous for the alternative allele in at least one individual, resulting in a total of 58,831 SNPs. We then converted the VCF- to a phylip file using the python script vcf2phylip.py (<https://github.com/edgarmortiz/vcf2phylip>). Finally, we used RAxML<sup>67</sup> v.8.2.12 to produce the phylogeny with the ASC\_GTRGAMMA substitution model and 100 bootstrap replicates. The resulting phylogeny was plotted with Figtree v.1.4.4 (<https://github.com/rambaut/figtree>).

To identify regions introgressed by the extinct *C. gutturosus* within individual genomes of all sequenced post-eutrophication samples, we used topology weighting by iterative sampling of subtrees (TWISST)<sup>44</sup>. First, we calculated genotype likelihoods in angsd<sup>61</sup> v.0.925, using the same thresholds and filtering parameters as above but allowing for missing reads in two individuals of the whole dataset to increase resolution. Additionally, we genotyped the same *S. salar* individual as used in the ABBA BABA test (above) at the positions identified to be polymorphic in our dataset. We then inferred genotypes from the likelihoods and phased these genotypes with beagle<sup>64</sup> v.4.1, resulting in a total of 2,676,591 polymorphic sites for further analysis. We acknowledge that our samples size is low for statistical phasing. However, statistical phasing is reasonably accurate at the short genomic ranges<sup>69</sup> that are relevant for our TWISST approach and TWISST has been reported to be robust to within-taxon phasing errors<sup>70</sup>. We assessed coverage of each sample at these polymorphic sites with angsd<sup>61</sup> v.0.925 and calculated average coverage across all these polymorphic sites (Extended Data Fig. 10). For each discrete 50 kb window across the genome, we computed a maximum likelihood tree including all genotyped samples using PhyML<sup>71</sup> v.3.0 and the script phym\_sliding\_windows.py ([https://github.com/simonhamartin/genomics\\_general/blob/master/phyllo](https://github.com/simonhamartin/genomics_general/blob/master/phyllo)). TWISST was performed separately for each post-eutrophication sample, using the same four-taxon topology in the same ordering as for the ABBA BABA tests (Fig. 2c), except that the potential recipient population P2 consisted of only one focal individual (all available pre-eutrophication samples of one extant species (P1), focal post-eutrophication sample of the same extant species (P2), all available *C. gutturosus* samples (P3) and *S. salar* as outgroup). With four populations, three different (unrooted) topologies are possible. Using the script twisst.py (<https://github.com/simonhamartin/twisst>), we computed the proportion of subtrees matching each possible topology (option 'complete'). The topology in which the focal post-eutrophication individual (P2) is more closely related to all available *C. gutturosus* individuals (P3) compared to all available prespeciation reversal individuals (P1) of the same species should only be supported within windows that were introgressed by *C. gutturosus* ('introgression topology'; Fig. 2c). Following Meier et al.<sup>72</sup>, we considered a window as introgressed if the weighting of the introgression topology exceeded a value of 66.6% (introgression topology received at least twice the statistical support of any other topology). We used a custom R script to assess the sharing of introgressed windows between heterospecific and conspecific individuals and the R package iNEXT<sup>45</sup> v.2.0.20 to estimate the total number of windows introgressed from *C. gutturosus* in all three extant species combined with the Chao estimator for species richness based on incidence data, as well as the total number of introgressed windows in each extant species separately. We performed a two-sided t-test<sup>73</sup> to evaluate whether the sharing of windows introgressed from *C. gutturosus* was significantly higher between conspecific individuals compared to heterospecifics. To verify our estimation of the amount of *C. gutturosus* variation retained in post-eutrophication populations of Lake Constance whitefish, we first calculated the mean *C. gutturosus* admixture proportions of all post-eutrophication populations of the PCAngsd<sup>35</sup> admixture analysis. Second, we used the script ABBABABAwindows.py ([https://github.com/simonhamartin/genomics\\_general/blob/master/ABBABABAwindows.py](https://github.com/simonhamartin/genomics_general/blob/master/ABBABABAwindows.py)) to calculate admixture proportions with  $f_d$  (ref. <sup>74</sup>) in discrete 500 kb windows

across the genome. We then calculated the genome-wide average. We used 500 kb windows to increase the number of SNPs per window, as we only included windows in the analysis that contained >700 SNPs.

We performed a selection scan using the statistic nSL<sup>46</sup>. nSL is a haplotype-based statistic inferring signatures of selection by combining information on the distribution of fragment lengths defined by pairwise differences with the distribution of the number of segregating sites between all pairs of chromosomes. We first subsetted our dataset of genotype likelihoods obtained from *angsd*<sup>61</sup> v.0.925 to only *C. gutturosus* individuals and then inferred genotypes and phased these using *beagle*<sup>64</sup> v.4.1. We then calculated the unstandardized nSL statistic with the software *selscan*<sup>75</sup> v.1.3.0. Because the sample size consisted of 11 individuals, we included low-frequency variants. We then used *norm*<sup>75</sup> v.1.3.0 to normalize the unstandardized nSL calculations with default parameters in 50 kb windows along the genome. We considered windows with >51.1% of variable sites (top 1 percentile) with a normalized nSL score >2 (default) to be under selection. As our sample size was low for such an approach relying on statistical phasing, we additionally calculated Tajima's *D* (ref. <sup>47</sup>) in *angsd*<sup>61</sup> v.0.925 on the basis of genotype likelihoods in 50 kb windows along the genome, to ensure that the pattern is not heavily impacted by phasing errors. First, we estimated the site allele frequency likelihood in *angsd*<sup>61</sup> v.0.925 and then calculated the maximum likelihood estimate of the folded site allele frequency spectrum using *realSFS* of *angsd*<sup>61</sup> v.0.925. We used the global site allele frequency spectrum to calculate theta per site in *realSFS* of *angsd*<sup>61</sup> v.0.925 and then calculated Tajima's *D* in 50 kb windows using *thetaStat* of *angsd*<sup>61</sup> v.0.925. We then compared the Tajima's *D* values of the top 1 percentile of 50 kb windows identified to be under selection by nSL to the rest of the genome (Extended Data Fig. 8). Finally, we showed that Tajima's *D* in the top 1 percentile of 50 kb windows identified to be under selection by nSL differed significantly from the rest of the genome using a two-sided Wilcoxon rank sum test in R 'wilcox.test'<sup>73</sup> ( $P < 0.01$ ,  $W = 8,352,543$ ). We assessed how many of these regions under selection introgressed into other whitefish species with a custom R script. We tested if introgressed regions were enriched for windows under selection by permutation: we randomly sampled the number of windows that were under selection from all windows along the genome and counted the number of overlaps of these randomly sampled windows with the observed introgressed windows. We then compared the expected counts of overlaps of 10,000 permutations with the observed count of overlaps to calculate a *P* value.

Regions identified as under selection in *C. gutturosus* were further investigated to identify which genes fall within these selected regions. Gene annotations (from the Alpine whitefish genome<sup>56</sup>; ENA accession *GCA\_902810595.1*) that overlap in their position with the identified windows under selection were identified using *bedtools*<sup>76</sup> v.2.28.0. Gene enrichment for specific gene ontology (GO) terms (from <https://datadryad.org/stash/dataset/doi:10.5061/dryad.xd2547ddf>) within these windows was then tested using the R package *topGO*<sup>77</sup> v.2.38.1 separately for each of the three ontology classes cellular component, biological processes and molecular function. We used Fisher's exact test applying both the 'weight' and 'elim' algorithms to each ontology class (with no fdr multiple testing correction in accordance to the *topGO* manual). GO terms that were enriched ( $P < 0.05$ ) from both the 'elim' and 'weight' algorithms were reported.

To determine whether introgressed and non-introgressed regions of the genome varied in gene density we repeated the above overlap analysis and calculated the base-pair overlap of genes from the Alpine whitefish genome with each of the introgressed and non-introgressed sets of windows. The difference in gene overlap between introgressed and non-introgressed windows was tested using a two-sided Wilcoxon rank sum test in R<sup>73</sup> 'wilcox.test' and showed that there was no significant difference between the two sets of windows.

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

## Data availability

The raw sequencing files are accessible on SRA (PRJEB43605). Additional supporting data (genotype and genotype-likelihood files, morphological raw data, data underlying Fig. 3, full output table of GO enrichment analysis) are deposited on the eawag research data institutional collections (<https://doi.org/10.25678/0005AP>)<sup>78</sup>. The Alpine whitefish reference genome<sup>56</sup> used was downloaded from ENA and is accessible with accession *GCA\_902810595.1*. The *S. salar* outgroup sample<sup>66</sup> used was downloaded from SRA and is accessible with accession *SRR3669756*. Gene ontology (GO) terms were downloaded from <https://datadryad.org/stash/dataset/doi:10.5061/dryad.xd2547ddf>.

## Code availability

Scripts used for data analysis are available on GitHub (<https://github.com/freidavid/Genomic-Consequences-of-Speciation-Reversal>).

Received: 28 June 2021; Accepted: 10 January 2022;

Published online: 24 February 2022

## References

- Vamosi, J. C., Magallon, S., Mayrose, I., Otto, S. P. & Sauquet, H. Macroevolutionary patterns of flowering plant speciation and extinction. *Annu. Rev. Plant Biol.* **69**, 685–706 (2018).
- Rhymer, J. M. & Simberloff, D. Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* **27**, 83–109 (1996).
- Vonlanthen, P. et al. Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482**, 357–362 (2012).
- Seehausen, O. Conservation: losing biodiversity by reverse speciation. *Curr. Biol.* **16**, R334–R337 (2006).
- Seehausen, O., Takimoto, G., Roy, D. & Jokela, J. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol. Ecol.* **17**, 30–44 (2008).
- Kearns, A. M. et al. Genomic evidence of speciation reversal in ravens. *Nat. Commun.* **9**, 906 (2018).
- Meier, J. I. et al. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 11 (2017).
- Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
- Barlow, A. et al. Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**, 1563–1570 (2018).
- Kuhlwilm, M., Han, S., Sousa, V. C., Excoffier, L. & Marques-Bonet, T. Ancient admixture from an extinct ape lineage into bonobos. *Nat. Ecol. Evol.* **3**, 957–965 (2019).
- Palkopoulou, E. et al. A comprehensive genomic history of extinct and living elephants. *Proc. Natl. Acad. Sci. USA* **115**, E2566–E2574 (2018).
- Ottenburghs, J. Ghost introgression: spooky gene flow in the distant past. *BioEssays* **42**, 2000012 (2020).
- Schlüter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
- Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8**, 336–352 (2005).
- Schlüter, D. Evidence for ecological speciation and its alternative. *Science* **323**, 737–741 (2009).
- Nosil, P. *Ecological Speciation* (Oxford Univ. Press, 2012).
- Yeaman, S. & Whitlock, M. C. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897–1911 (2011).
- Ghosh, S. M. & Joshi, A. Evolution of reproductive isolation as a by-product of divergent life-history evolution in laboratory populations of *Drosophila melanogaster*. *Ecol. Evol.* **2**, 3214–3226 (2012).
- Seehausen, O., van Alphen, J. J. M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811 (1997).
- 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).
- Grant, P. R. & Grant, B. R. Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci. USA* **116**, 23216–23224 (2019).
- Kagawa, K. & Takimoto, G. Hybridization can promote adaptive radiation by means of transgressive segregation. *Ecol. Lett.* **21**, 264–274 (2018).
- Feller, A. F. et al. Rapid generation of ecologically relevant behavioral novelty in experimental cichlid hybrids. *Ecol. Evol.* **10**, 7445–7462 (2020).
- Pfennig, K. S., Kelly, A. L. & Pierce, A. A. Hybridization as a facilitator of species range expansion. *Proc. R. Soc. B* <https://doi.org/10.1098/rspb.2016.1329> (2016).
- Lamichhaney, S. et al. Rapid hybrid speciation in Darwin's finches. *Science* **359**, 224–227 (2018).
- Hudson, A. G., Vonlanthen, P., Bezault, E. & Seehausen, O. Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish. *BMC Evol. Biol.* **13**, 17 (2013).
- Hudson, A. G., Vonlanthen, P. & Seehausen, O. Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proc. R. Soc. B* **278**, 58–66 (2011).
- Jacobs, A. et al. Rapid niche expansion by selection on functional genomic variation after ecosystem recovery. *Nat. Ecol. Evol.* **3**, 77–86 (2019).
- Steinmann, P. Monographie der schweizerischen Koregonen. Beitrag zum problem der entstehung neuer Arten. *Schweiz. Z. Hydrol.* **12**, 340–491 (1950).
- Selz, O. M., Donz, C. J., Vonlanthen, P. & Seehausen, O. A taxonomic revision of the whitefish of lakes Brienz and Thun, Switzerland, with descriptions of four new species (Teleostei, Coregonidae). *Zoobios* **989**, 79–162 (2020).
- Hudson, A. G., Lundsgaard-Hansen, B., Lucek, K., Vonlanthen, P. & Seehausen, O. Managing cryptic biodiversity: fine-scale intralacustrine speciation along a benthic gradient in Alpine whitefish (*Coregonus* spp.). *Evol. Appl.* **10**, 251–266 (2016).
- Doenz, C. J., Bittner, D., Vonlanthen, P., Wagner, C. E. & Seehausen, O. Rapid buildup of sympatric species diversity in Alpine whitefish. *Ecol. Evol.* **8**, 9398–9412 (2018).
- Vonlanthen, P. et al. Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *J. Evol. Biol.* **22**, 498–514 (2009).
- Feulner, P. G. D. & Seehausen, O. Genomic insights into the vulnerability of sympatric whitefish species flocks. *Mol. Ecol.* **28**, 615–629 (2019).
- Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
- Soraggi, S., Wiuf, C. & Albrechtsen, A. Powerful inference with the D-statistic on low-coverage whole-genome data. *Genes Genomes Genet.* **8**, 551–566 (2018).

37. Wahl, B. & Löffler, H. Influences on the natural reproduction of whitefish (*Coregonus lavaretus*) in Lake Constance. *Can. J. Fish. Aquat. Sci.* **66**, 547–556 (2009).
38. Nümann, W. The Bodensee: effects of exploitation and eutrophication on the Salmonid community. *J. Fish. Res. Board Can.* **29**, 833–884 (1972).
39. Deufel, J., Löffler, H. & Wagner, B. Auswirkungen der Eutrophierung und anderer anthropogener einflüsse auf die laichplätze einiger Bodensee-Fischarten. *Österr. Fisch.* **39**, 325–336 (1986).
40. Straile, D. & Geller, W. Crustacean zooplankton in Lake Constance from 1920 to 1995: response to eutrophication and re-oligotrophication. *Adv. Limnol.* **58**, 255–274 (1998).
41. Eby, L. A., Crowder, L. B., McClellan, C. M., Peterson, C. H. & Powers, M. J. Habitat degradation from intermittent hypoxia: impacts on demersal fishes. *Mar. Ecol. Prog. Ser.* **291**, 249–261 (2005).
42. Powers, S. P. et al. Effects of eutrophication on bottom habitat and prey resources of demersal fishes. *Mar. Ecol. Prog. Ser.* **302**, 233–243 (2005).
43. Caires, A. M., Chandra, S., Hayford, B. L. & Wittmann, M. E. Four decades of change: dramatic loss of zoobenthos in an oligotrophic lake exhibiting gradual eutrophication. *Freshw. Sci.* **32**, 692–705 (2013).
44. Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
45. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).
46. Ferrer-Admetlla, A., Liang, M., Korneliussen, T. & Nielsen, R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* **31**, 1275–1291 (2014).
47. Korneliussen, T. S., Moltke, I., Albrechtsen, A. & Nielsen, R. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinform.* **14**, 289 (2013).
48. Doenz, C. J. & Seehausen, O. Rediscovery of a presumed extinct species, *Salvelinus profundus*, after re-oligotrophication. *Ecology* **101**, e03065 (2020).
49. Alexander, T. & Seehausen, O. *Diversity, Distribution and Community Composition of Fish in Perialpine Lakes—‘Projec Lac’ Synthesis Report* (Eawag, 2021).
50. 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).
51. Nussle, S., Bornand, C. N. & Wedekind, C. Fishery-induced selection on an Alpine whitefish: quantifying genetic and environmental effects on individual growth rate. *Evol. Appl.* **2**, 200–208 (2009).
52. Grabenstein, K. C. & Taylor, S. A. Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends Ecol. Evol.* **33**, 198–212 (2018).
53. Seehausen, O. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198–207 (2004).
54. Wasko, A. P., Martins, C., Oliveira, C. & Foresti, F. Non-destructive genetic sampling in fish. An improved method for DNA extraction from fish fins and scales. *Hereditas* **138**, 161–165 (2003).
55. Chen, S. F., Zhou, Y. Q., Chen, Y. R. & Gu, J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, 884–890 (2018).
56. De-Kayne, R., Zoller, S. & Feulner, P. G. D. A *de novo* chromosome-level genome assembly of *Coregonus* sp. ‘Balchen’: one representative of the Swiss Alpine whitefish radiation. *Mol. Ecol. Resour.* **20**, 1093–1109 (2020).
57. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
58. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
59. Kim, S. Y. et al. Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinform.* **12**, 231 (2011).
60. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Association testing for next-generation sequencing data using score statistics. *Genet. Epidemiol.* **36**, 430–437 (2012).
61. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
62. Duforet-Frebbourg, N. & Slatkin, M. Isolation-by-distance-and-time in a stepping-stone model. *Theor. Popul. Biol.* **108**, 24–35 (2016).
63. Meisner, J., Liu, S., Huang, M. & Albrechtsen, A. Large-scale inference of population structure in presence of missingness using PCA. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btab027> (2021).
64. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
65. Goudet, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
66. Kjaerner-Semb, E. et al. Atlantic salmon populations reveal adaptive divergence of immune related genes—a duplicated genome under selection. *BMC Genom.* **17**, 610 (2016).
67. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
68. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
69. Bokowicki, M., Franssen, S. U. & Schlotterer, C. High rates of phasing errors in highly polymorphic species with low levels of linkage disequilibrium. *Mol. Ecol. Resour.* **16**, 874–882 (2016).
70. Marburger, S. et al. Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* **10**, 5218 (2019).
71. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
72. Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L. & Seehausen, O. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**, 1489–1506 (2018).
73. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018).
74. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
75. Szpiech, Z. A. & Hernandez, R. D. SelScan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014).
76. Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–11.12.34 (2014).
77. Alexa, A. & Rahnenfuhrer, J. TopGO: enrichment analysis for gene ontology. R package version 2.42.0. <https://doi.org/10.18129/B9.bioc.topGO> (2020).
78. Frei, D., De-Kayne, R., Selz, O. M., Seehausen, O. & Feulner, P. G. D. *Data for: Genomic Variation From an Extinct Species is Retained in the Extant Radiation Following Speciation Reversal* (Eawag, 2021); <https://doi.org/10.25678/0005AP>

## Acknowledgements

We thank all professional fisherman for providing specimens, D. Bittner for compiling the historical whitefish scale collection, M. Kugler from the Amt für Natur, Jagd und Fischerei, St.Gallen and the Institute of Seenforschung and Fischereiwesen Langenargen for providing historical whitefish scales from Lake Constance and IGKB for providing the yearly averaged total phosphorus data. We thank the NGS facility of the University of Bern for sequencing support and the Genetic Diversity Centre at ETH Zurich for bioinformatic support. Further, we are very thankful to the Fish Ecology and Evolution Department of Eawag, especially to C. Dönz, B. Matthews, D. Marques, K. Kagawa, J. Meier and J.T. Brink for feedback, comments and ideas on this manuscript. Also, thanks to O. Osborne for advice on assessing GO term enrichment. This work received financial support from Eawag (including Eawag Discretionary Funds 2018–2022) and the Swiss Federal Office for the Environment. The work was further supported by the grant ‘SeeWandel: Life in Lake Constance—the past, present and future’ within the framework of the Interreg V programme ‘Alpenrhein-Bodensee-Hochrhein (Germany/Austria/Switzerland/Liechtenstein)’ which funds are provided by the European Regional Development Fund as well as the Swiss Confederation and cantons (to P.F. and O.S.). The funders had no role in study design, decision to publish or preparation of the manuscript.

## Author contributions

O.S. conceived of the study, D.F., O.S. and P.G.D.F. designed and conceptualized it. P.G.D.F. managed and supervised the study. O.M.S. collected contemporary specimens and collected and analysed morphological data. R.D.K. contributed to DNA extraction and genomic analysis. D.F. analysed genomic data and visualized the results. D.F. wrote the original manuscript draft with input from O.S. and P.G.D.F. All authors edited and reviewed the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41559-022-01665-7>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41559-022-01665-7>.

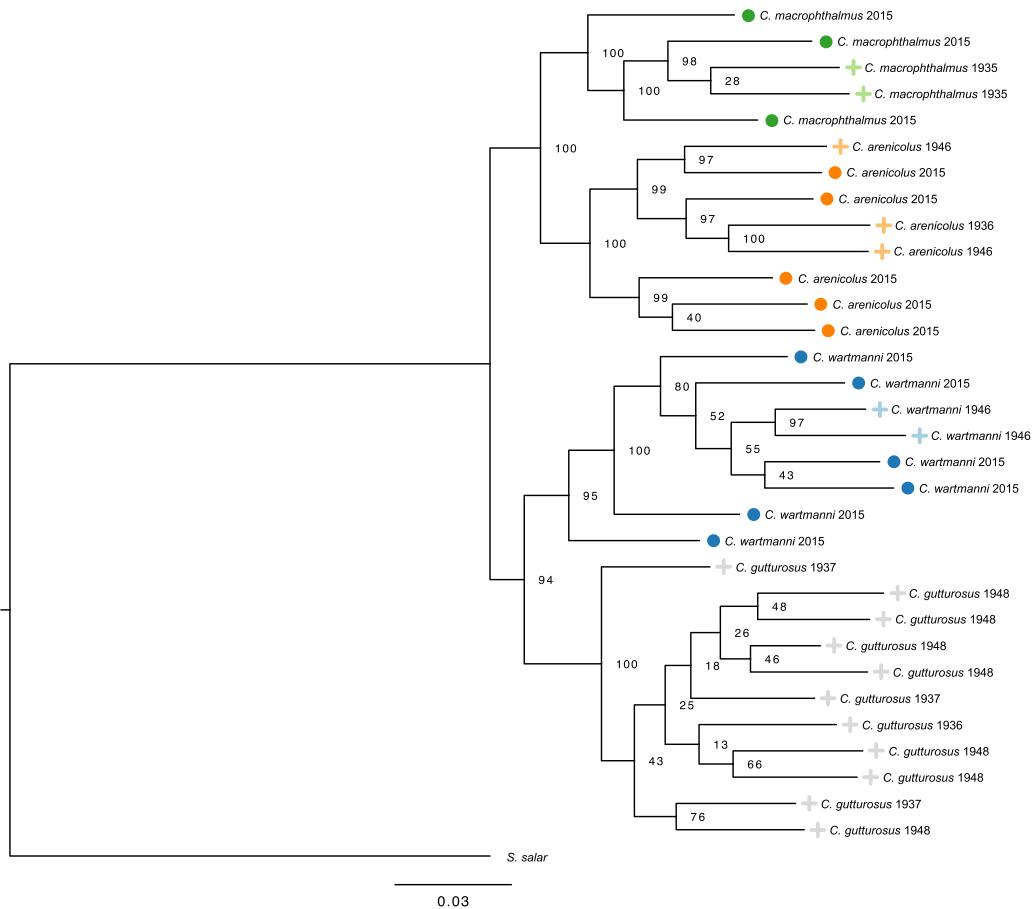
Correspondence and requests for materials should be addressed to Philine G. D. Feulner.

Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work.

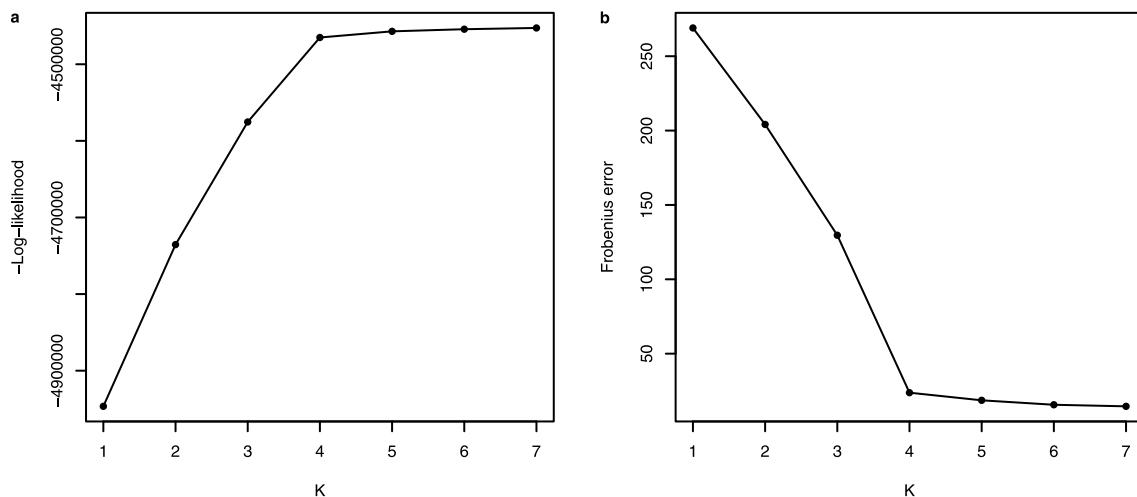
Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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 2022



**Extended Data Fig. 1 | Maximum likelihood phylogeny of all historical and contemporary samples.** Maximum likelihood phylogeny of all pre- (crosses) and post-eutrophication (points) individuals of the four Lake Constance whitefish species based on 58'831 SNPs. Colours correspond to species (see Fig. 1). Support values from 100 bootstrap replicates are shown on each node. Note that the branch length for the *S. salar* outgroup is biased due to the ascertainment towards SNPs segregating within Lake Constance whitefish (see Methods section).



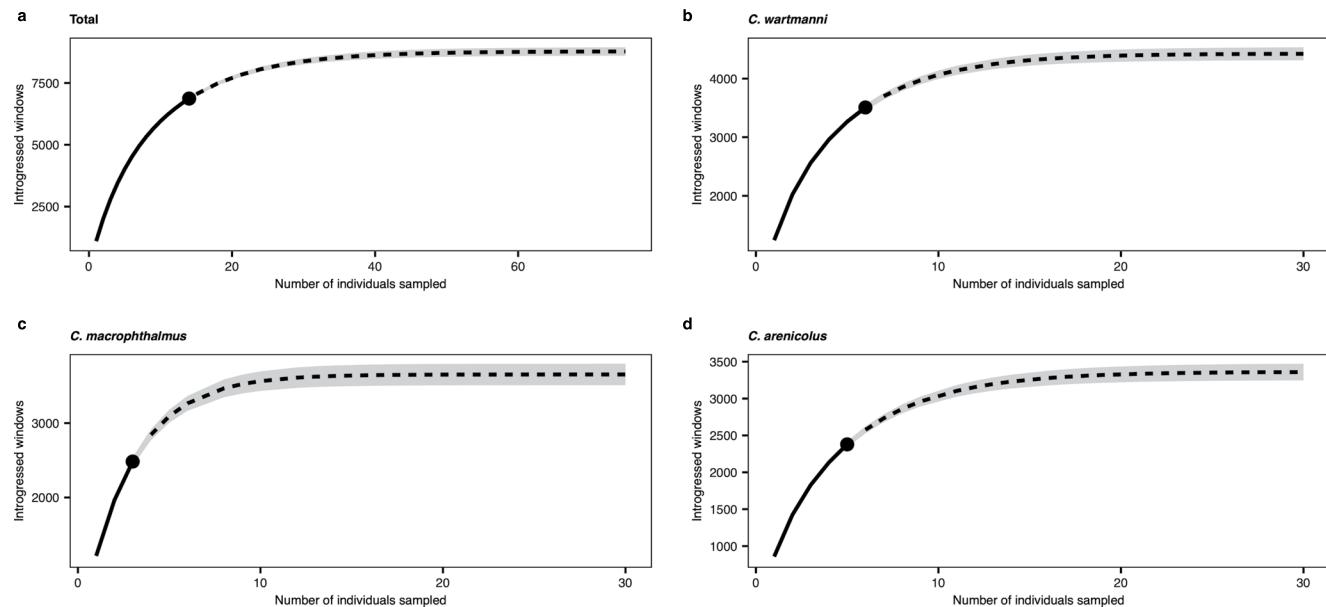
**Extended Data Fig. 2 | Log-likelihood and frobenius error for different K's of the admixture analysis.** Log-likelihood values (a) and frobenius error (b) for different K's of the PCAngsd admixture analysis shown in Fig. 1c. K = 4 turned out to represent the data best, also corresponding to the number of species included.

	Microsatellite data <sup>3</sup>		SNP data	
	$F_{ST}$	n	$F_{ST}$	n
Pre-eutrophication	0.108 (0.165)	68 (133)	0.046 (0.052)	7 (18)
Post-eutrophication	0.046	121	0.022	14

**Extended Data Fig. 3 | Comparison of the change in differentiation across the eutrophication period.** Global  $F_{ST}$  values and sample sizes for pre-eutrophication and post-eutrophication populations of all species of Lake Constance whitefish by Vonlanthen et al. 2012<sup>3</sup> based on 10 microsatellite markers, compared to the genetic differentiation estimates and sample sizes for the same populations based on our whole-genome sequencing approach and 477'981 SNPs. Values in brackets include samples of the now extinct *C. gutturosus*.

P1	P2	P3	P4	D	Z	p
<i>C. macrophthalmus</i> pre	<i>C. macrophthalmus</i> post	<i>C. wartmanni</i> pre	<i>S. salar</i>	0.006	1.273	0.203
<i>C. arenicolus</i> pre	<i>C. arenicolus</i> post	<i>C. wartmanni</i> pre	<i>S. salar</i>	0.003	0.880	0.379
<i>C. wartmanni</i> pre	<i>C. wartmanni</i> post	<i>C. macrophthalmus</i> pre	<i>S. salar</i>	0.033	9.112	< 0.001
<i>C. arenicolus</i> pre	<i>C. arenicolus</i> post	<i>C. macrophthalmus</i> pre	<i>S. salar</i>	0.020	5.150	< 0.001
<i>C. wartmanni</i> pre	<i>C. wartmanni</i> post	<i>C. gutturosus</i> pre	<i>S. salar</i>	0.033	9.074	< 0.001
<i>C. macrophthalmus</i> pre	<i>C. macrophthalmus</i> post	<i>C. gutturosus</i> pre	<i>S. salar</i>	0.019	4.850	< 0.001
<i>C. arenicolus</i> pre	<i>C. arenicolus</i> post	<i>C. gutturosus</i> pre	<i>S. salar</i>	0.027	7.787	< 0.001
<i>C. wartmanni</i> pre	<i>C. wartmanni</i> post	<i>C. arenicolus</i> pre	<i>S. salar</i>	0.028	7.502	< 0.001
<i>C. macrophthalmus</i> pre	<i>C. macrophthalmus</i> post	<i>C. arenicolus</i> pre	<i>S. salar</i>	0.003	0.813	0.416

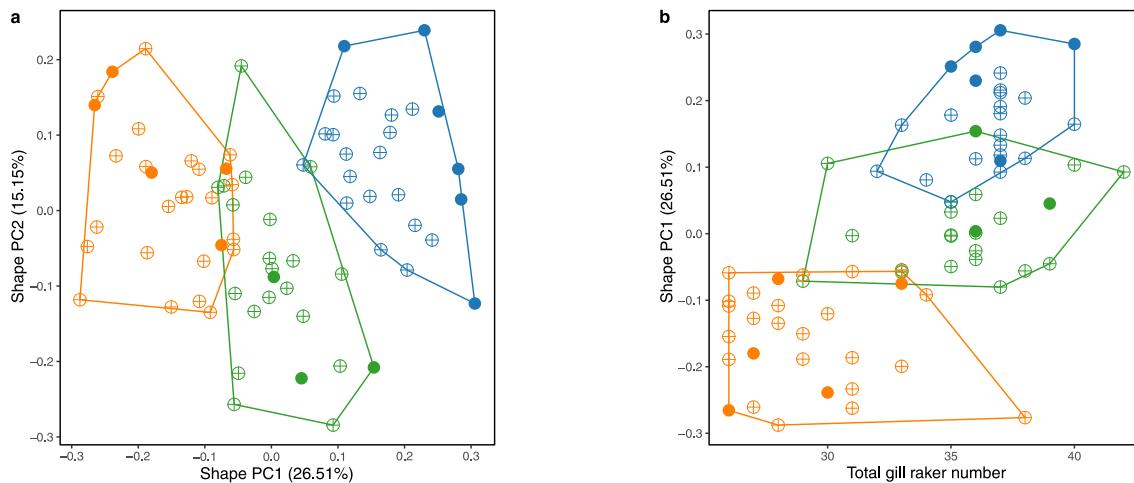
**Extended Data Fig. 4 | D-statistic results for all tests for introgression shown in Fig. 2.** The table includes the ordering of the populations on the four-taxon topology used for the ABBA BABA test, as well as the resulting D values, Z-scores and p-values of the block-jackknife approach in 5 Mb blocks. All sequenced individuals per population have been used for each single test (see Extended Data Fig. 10).



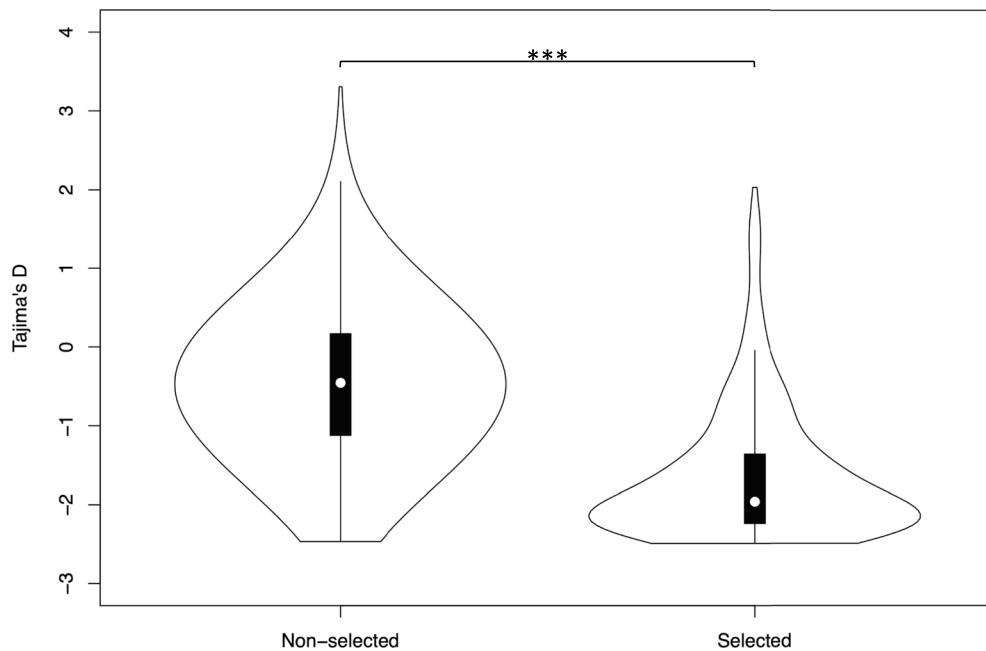
**Extended Data Fig. 5 | Rarefaction analysis of the *C. gutturosus* genome maintained in extant whitefish species.** Rarefaction curves for all species combined showing the estimated number of introgressed windows in contemporary populations of Lake Constance whitefish (a) and for each extant species of the Lake Constance whitefish radiation (b-d). The x axis shows the estimated total number of introgressed 50 kb windows (whole genome corresponds to 31'476 windows) for a given number of sampled individuals. The dashed lines show the sample size-based extrapolation curves and the grey areas around the curves indicate the 95% confidence intervals.

	PCAngsd	TWISST rarefaction	$f_d$
<i>C. arenicolus</i>	0.12	0.11	0.11
<i>C. macrophthalmus</i>	0.08	0.12	0.10
<i>C. wartmanni</i>	0.14	0.14	0.14

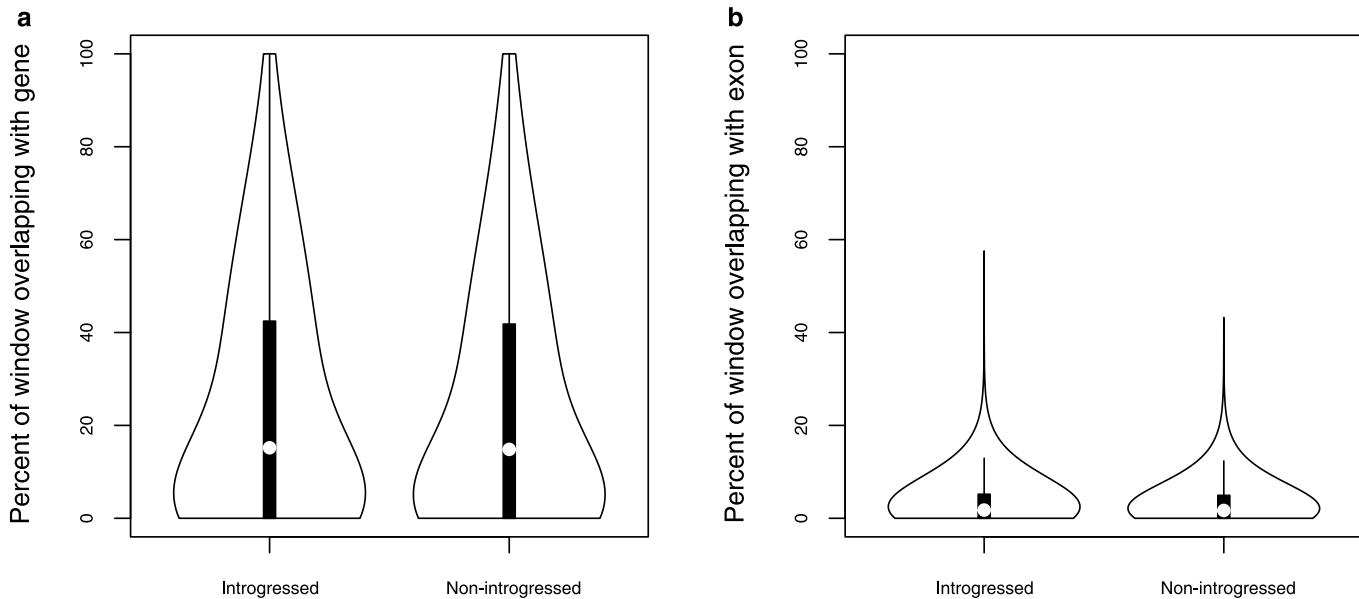
**Extended Data Fig. 6 | *C. gutturosus* admixture proportions in post-eutrophication populations of Lake Constance whitefish.** Table shows mean admixture proportions averaged across all individuals for the PCAngsd approach (see Fig. 1), proportions estimated with the rarefaction analysis for windows showing signals of *C. gutturosus* introgression in the TWISST analysis (Fig. 3 and Extended Data Fig. 5) and genome-wide means of admixture proportions estimated with  $f_d$  (see Methods section).



**Extended Data Fig. 7 | Morphological differentiation of contemporary Lake Constance whitefish.** **a)** Shape PCA of the first two principal components based on body characters (PELVFB, PELVFS, PELVF, PECFB, PECE1, PECE2, DFB, DFAe, DFAd, DFPe, AFB, AFAe, ADFB, CF, CD, CL, PAdC, DHL, PreP, PreA, SL, TL, PreD, BD, PostD; see Table 1 in Selz et al.<sup>30</sup>). Morphological characters were measured and analysed following Selz et al.<sup>30</sup>. **b)** The plot shows shape PC1 of panel a) against the total gill raker count of the individuals. Individuals used for genomic analysis are indicated with filled circles, additional individuals of the contemporary species are indicated with crossed circles. Colours correspond to species (orange *C. areniculus*, green *C. macrophthalmus* and blue *C. wartmanni*).



**Extended Data Fig. 8 | Tajima's D based on genotype likelihoods for windows identified to have been under selection in *C. gutturosus* using nSL.** Violin plots of Tajima's D in *C. gutturosus* ( $n=11$ ) calculated in 50 kb windows comparing the 315 windows identified to be in the top 1 percentile of the nSL analysis to all other windows of the genome. We found a significant difference in Tajima's D between selected and non-selected windows identified by nSL (two-sided Wilcoxon rank sum test,  $W=8352543$ ,  $p < 2.2e-16$ , indicated with bars above the plot '\*\*\*'). Plots show the estimated kernel densities, black boxes show the interquantile range, white dots correspond to medians and spikes are extending to the upper and lower adjacent values.



**Extended Data Fig. 9 | Comparison of gene density in introgressed and non-introgressed windows.** **a)** Comparison of gene density between windows identified to be introgressed and those that did not show evidence for introgression (non-introgressed) from *C. gutturosus* ( $n=11$ ) across all three extant species ( $n=14$ ). There was no significant difference between introgressed and non-introgressed windows (two-sided Wilcoxon rank sum test,  $W=84559580$ ;  $p=0.5458$ ) and thus the test is not represented in the figure. **b)** Comparison of exon density between windows identified to be introgressed and those that did not show evidence for introgression (non-introgressed) from *C. gutturosus* ( $n=11$ ) across all three extant species ( $n=14$ ). There was no significant difference between introgressed and non-introgressed windows (two-sided Wilcoxon rank sum test,  $W=85267215$ ;  $p=0.0906$ ) and thus the test is not represented in the figure. Plots show the estimated kernel densities, black boxes show the interquartile range, white dots correspond to medians and spikes are extending to the upper and lower adjacent values.

Species	Year	Platform	Total reads	Fragment length	Lab ID	Coverage
<i>C. gutturosus</i>	1937	Novaseq & Hiseq	6.77E+08	333	DF5	14.2
<i>C. gutturosus</i>	1937	Novaseq & Hiseq	7.21E+08	348	DF11	14.2
<i>C. gutturosus</i>	1948	Novaseq	2.89E+08	301	DF1	4
<i>C. gutturosus</i>	1948	Novaseq	3.27E+08	329	DF3	6.9
<i>C. gutturosus</i>	1937	Novaseq	4.56E+08	373	DF6	10.2
<i>C. gutturosus</i>	1948	Novaseq	2.85E+08	370	DF7	4.8
<i>C. gutturosus</i>	1948	Novaseq	2.51E+08	338	DF8	5.3
<i>C. gutturosus</i>	1948	Novaseq	3.10E+08	340	DF9	3.4
<i>C. gutturosus</i>	1948	Novaseq	2.68E+08	322	DF12	2.8
<i>C. gutturosus</i>	1948	Novaseq	3.23E+08	337	DF4	6.8
<i>C. gutturosus</i>	1936	Novaseq	3.76E+08	263	DF20	6.1
<i>C. arenicolus</i>	1936	Hiseq	3.67E+08	264	DF19	6.3
<i>C. arenicolus</i>	1946	Hiseq	3.13E+08	311	DF30	4.2
<i>C. arenicolus</i>	1946	Hiseq	2.93E+08	324	DF31	3.9
<i>C. arenicolus</i>	2015	Hiseq	2.21E+08	620	DF123477	11
<i>C. arenicolus</i>	2015	Hiseq	2.58E+08	581	DF123440	13
<i>C. arenicolus</i>	2015	Novaseq	7.33E+08	551	DF126	35.7
<i>C. arenicolus</i>	2015	Novaseq	7.93E+08	528	DF127	31.1
<i>C. arenicolus</i>	2015	Novaseq	6.11E+08	505	DF128	24.9
<i>C. macrophthalmus</i>	1935	Hiseq	3.38E+08	262	DF17	5.2
<i>C. macrophthalmus</i>	1935	Hiseq	4.22E+08	281	DF18	7.6
<i>C. macrophthalmus</i>	2015	Hiseq	1.57E+08	599	DF123458	8.1
<i>C. macrophthalmus</i>	2015	Hiseq	2.86E+08	637	DF123470	14.3
<i>C. macrophthalmus</i>	2015	Novaseq	4.87E+08	518	DF132	22.4
<i>C. wartmanni</i>	1946	Hiseq	4.73E+08	280	DF23	5.8
<i>C. wartmanni</i>	1946	Hiseq	2.13E+08	286	DF24	2.4
<i>C. wartmanni</i>	2015	Hiseq	2.12E+08	560	DF123446	10.7
<i>C. wartmanni</i>	2015	Hiseq	2.07E+08	562	DF123448	10.6
<i>C. wartmanni</i>	2015	Novaseq	6.63E+08	550	DF121	32.7
<i>C. wartmanni</i>	2015	Novaseq	5.60E+08	512	DF122	26.9
<i>C. wartmanni</i>	2015	Novaseq	6.95E+08	523	DF123	33.3
<i>C. wartmanni</i>	2015	Novaseq	7.28E+08	528	DF131	34.9

**Extended Data Fig. 10 | Overview over all sequenced samples.** Year of sampling, sequencing platform used, total yield of reads, mean fragment length of library, lab identification code and mean coverage at polymorphic sites for each individual sequenced. Samples collected before 1950 are scale samples, while samples from 2015 are fin-clip samples.

## 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 our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the 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
- A statement on 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 statistical parameters 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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

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

Data collection No software was used for data collection.

Data analysis Scripts used for analysis and generation of figures are available at <https://github.com/freidavid/Genomic-Consequences-of-Speciation-Reversal>.

Software used for analysis:

Fastp 0.20.0

Seqprep 1.0

Bwa mem 0.7.12

Picard-tools 2.20.2

Angsd 0.925

PCAngsd 0.98

Beagle 4.1

RAxML 8.2.12

VCFtools 0.1.16

Bcftools 1.10.2

Figtree 1.4.4

PhyML 3.0

SelScan 1.3.0

Bedtools 2.28.0

R 4.0.5

Code developed in other studies:

<https://github.com/edgarmortiz/vcf2phylip>

[https://github.com/simonthomas/genomics\\_general/blob/master/phylo/phyml\\_sliding\\_windows.py](https://github.com/simonthomas/genomics_general/blob/master/phylo/phyml_sliding_windows.py)

[https://github.com/simonthomas/genomics\\_general/blob/master/ABBABABAwindows.py](https://github.com/simonthomas/genomics_general/blob/master/ABBABABAwindows.py)  
<https://github.com/simonthomas/twist/blob/master/twist.py>

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 and reviewers. 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

The raw sequencing files are accessible on ENA (PRJEB4360 5). Scripts are available on GitHub (<https://github.com/freidavid/Genomic-Consequences-of-Speciation-Reversal>). Additional supporting data (genotype and genotype likelihood files, morphological raw data, data underlying Figure 3, output of GO enrichment analysis) is deposited on the eawag research data institutional collections (doi:10.25678/0005AP).

The reference genome used was downloaded from ENA and is accessible with accession GCA\_902810595.1.

The S. salar outgroup sample used was downloaded from NCBI SRA and is accessible with accession SRR3669756.

Gene ontology (GO) terms were downloaded from <https://datadryad.org/stash/dataset/doi:10.5061/dryad.xd2547ddf>

## Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

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

Study description	Whole genome sequencing of historical and contemporary whitefish from Lake Constance
Research sample	11 Coregonus guttatus scale samples sampled pre-eutrophication (sampled 1936, 1937 and 1948) 3 Coregonus arenicola scale samples sampled pre-eutrophication (sampled 1936 and 1946) 2 Coregonus macrophthalmus scale samples sampled pre-eutrophication (sampled 1935) 2 Coregonus wartmanni scale samples sampled pre-eutrophication (sampled 1946) 28 Coregonus arenicola (15 females and 13 males, sampled 2015; 5 individuals sequenced, 28 used for morphological analysis) 23 Coregonus macrophthalmus finclip samples (8 females and 15 males, sampled 2015; 3 individuals sequenced, 23 used for morphological analysis) 24 Coregonus wartmanni finclip samples (7 females 17 males, sampled 2015; 6 individuals sequenced, 24 used for morphological analysis)
Sampling strategy	For the historical samples, we focused our sampling effort on the extinct whitefish species (n=11) and represented other co-occurring whitefish species by a minimal sample sizes (n= 2-3). Contemporary samples were caught by professional fishermen during spawning fisheries. For these post-eutrophication populations of contemporary species we sequenced 3 to 6 samples for comparison.
Data collection	Historical scale samples were collected by Swiss cantonal authorities and professional fishermen, and stored in paper bags at room temperature (see Vonlanthen et al. 2012 Nature for detailed information). Contemporary finclip samples were collected from fish caught by professional fishermen during spawning fisheries in 2015. Whole genome sequencing was done at the NGS facility of the university of Bern.
Timing and spatial scale	All samples are from Lake Constance. Historical scale samples are from 1935, 1937, 1946, 1948. The contemporary tissue samples are from 2015 and have been collected during spawning fisheries (see also Figure 1a and Supplementary Table 4).
Data exclusions	No data was excluded from the analysis.
Reproducibility	The work was not experimental in nature, and did not necessitate repeat analysis.
Randomization	The work was not experimental in nature, and did not necessitate random assignment of individuals to treatment groups.
Blinding	The work was not experimental in nature, and did not necessitate blinding.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

- |                                     |                               |
|-------------------------------------|-------------------------------|
| n/a                                 | Involved in the study         |
| <input checked="" type="checkbox"/> | Antibodies                    |
| <input checked="" type="checkbox"/> | Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | Palaeontology and archaeology |
| <input type="checkbox"/>            | Animals and other organisms   |
| <input checked="" type="checkbox"/> | Human research participants   |
| <input checked="" type="checkbox"/> | Clinical data                 |
| <input checked="" type="checkbox"/> | Dual use research of concern  |

## Methods

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

## Animals and other organisms

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

### Laboratory animals

No laboratory animals were used.

### Wild animals

Contemporary individuals used were caught by local fishermen during the spawning season of 2015. Individuals were anaesthetized and subsequently euthanized using appropriate concentrations of tricaine methane sulfonate solutions (MS-222) according to the permit issued by the cantons of Zurich and St.Gallen (ZH128/15).  
Three species were caught: Coregonus arenicola (15 females and 13 males), Coregonus macrourus (8 females and 15 males) and Coregonus wartmanni (7 females 17 males).  
Historical scale samples have been collected by Cantonal authorities, see Vonlanthen et al. (2012) Nature.

### Field-collected samples

No field-collected samples were used in this study.

### Ethics oversight

Permit issued by the cantons of Zurich and St.Gallen (ZH128/15).

Note that full information on the approval of the study protocol must also be provided in the manuscript.