This document is the accepted manuscript version of the following article:
Frei, D., De-Kayne, R., Selz, O. M., Seehausen, O., & Feulner, P. G. D. (2022).
Genomic variation from an extinct species is retained in the extant radiation
following speciation reversal. Nature Ecology & Evolution.
https://doi.org/10.1038/s41559-022-01665-7

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

David Frei[1,2], Rishi De-Kayne[1,2], Oliver M. Selz[1,2], Ole Seehausen[1,2], Philine G.D. Feulner[1,2*]

[1] Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland

[2] Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

* Corresponding author (philine.feulner@eawag.ch)

Ecosystem degradation and biodiversity loss are major global challenges. When reproductive isolation between species is contingent upon 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.

Introduction

A mechanistic understanding of species extinction is critical for a better understanding of contemporary patterns of biodiversity as well as for predicting its future[1]. Extinction can result from demographic decline, from loss of reproductive isolation or from a combination of both[2,3]. When extinction involves the loss of reproductive isolation[4,5], the extinction process can leave a lasting legacy in the genomes of surviving species through introgressive hybridization[2,6], potentially even influencing species that will only emerge in the future[7]. 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[8-11]. However, apart from these few examples, genomic information for extinct species is still rare[12]. As a result, the extent and the evolutionary significance 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[15,16], is an important process in the evolution of a substantial proportion of contemporary eukaryotic species diversity[13,14]. 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[13,14,16-18]. Ecologically-mediated reproductive isolation, both pre- and postzygotic, results from performance trade-offs between, and adaptation to, alternative fitness optima[15,16]. 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 break-down of reproductive isolation might culminate in the collapse of sympatric species into hybrid populations[19,20], a process called speciation reversal, potentially resulting in the sudden and rapid extinction of species through introgressive hybridization[4,5].

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[5]. Whilst the potentially widespread impacts of speciation reversal on contemporary biodiversity loss are still underappreciated in conservation[4], its genomic consequences are still underappreciated in evolutionary biology. Genetically admixed hybrid populations that emerged from speciation reversal might have enhanced evolvability[21]. In the future, such populations may adapt in new and unexpected ways[22,23], expand their ranges[24], and even seed further species diversification[25]. 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[26-28]. Across the large pre-Alpine lakes of Switzerland more than 30 endemic whitefish species have evolved since the end of the last glacial maximum[27,29-32]. 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[31]. Therefore, Alpine whitefish species are highly sensitive to speciation reversal when habitat diversity and suitability along the lacustrine water depth

gradient changes[31,33,34]. Anthropogenic eutrophication during the 20th 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[3]. 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[3,26].

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 gutturosus* now comprises only three extant species (see Fig. 1, 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[3]. Additionally, five private microsatellite alleles of the extinct species were discovered in all extant species after eutrophication, consistent with speciation reversal through introgressive hybridization[3]. We here provide a novel 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 prior to 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[3,19,20]. Prior to the ecosystem changes during the 20th century (Fig. 1a), the four Lake Constance whitefish species, including the now extinct *C. gutturosus*, formed well defined species clusters within a multidimensional genotype space[35] (Fig. 1b) based on genotype likelihoods of 222'017 polymorphic sites, despite complete sympatry. An analysis of population structure[35] (see 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 10 microsatellite markers (Extended Data Table 1) that demonstrated a rapid reduction of genetic differentiation (global F_{ST} decreased over twofold[3]) between these whitefish species by comparing samples collected more than 40 years apart and separated by a period of anthropogenic lake eutrophication[3]. 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; Extended Data Table 1), 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 (i.e. see Fig. 2c for a generic topology) using an extended version of D-statistics that allows to include multiple individuals per population[36]. We found that significant introgression had occurred from deeper into shallower spawning species, but not in the opposite direction (Fig. 2a and Extended Data Table 2). 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 Table 2). 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[37,38]. Whereas low oxygen conditions in deeper benthic areas probably prevented successful reproduction of *C. gutturosus* and contributed to its extinction[37,39], 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. arenicolus* to survive[39]. 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[37,39]. At the same time, increased productivity during eutrophication led to an increase of zooplankton density[40] and decreased zoobenthos densities[41-43], possibly relaxing divergent selection between feeding niches[3]. Our data therefore uncovers 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[44] to explore evolutionary relationships in 50 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). Based on a rarefaction analysis[45] on windows indicating signals consistent with introgression, we estimated that ~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. arenicolus*; Extended Data Fig. 3). Alternative approaches resulted in very similar approximations of *C. gutturosus* admixture proportions in the three contemporary species (Extended Data Table 3). Windows showing an introgression signature were more frequently shared between individuals of the same species (Fig. 3) than between individuals of different species ($t=57.18$; $p<0.01$; $df=29.34$). 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. 4) 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[26]. We performed a selection scan using the haplotype-based statistic nSL[46] 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[47] in this top 1% of windows is significantly different from the rest of the genome; see Extended Data Fig. 5) 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 have even been favoured in their new bearers, although the time span after introgression is likely 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. 6). Both the introgression of potentially adaptive variation and no evidence that introgression is confined to gene-poor regions suggests 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[3,4,26] and suggests that genetic incompatibilities between these species were relatively weak. While those introgressed variants may behave 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

188 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[19]. 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[3,19]. 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 e.g. elephants[11], apes[10] and bears[9], although, the evolutionary processes resulting in the persistence of ancient alleles often remain unclear. We here demonstrate that during extinction by speciation reversal there was substantial and wide-spread 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

214 that had evolved in the extinct species prior to eutrophication, thereby preserving fractions of
215 its evolutionary legacy from being lost forever.

216 Today, oligotrophic conditions of Lake Constance have been largely restored and deep-
217 water habitats are again accessible for fish[48]. Nonetheless, profundal regions remain devoid of
218 whitefish[49]. Theoretical work has suggested that when disturbance of reproductive isolation is
219 short and transient, species pairs that collapsed may re-emerge after restoration of
220 environmental conditions favourable of speciation[50]. However, re-emergence appears less
221 likely the more species that are involved in hybridization during the collapse of reproductive
222 isolation, and the timescale in which re-emergence might happen is orders of magnitudes
223 larger than it takes to collapse species into hybrid populations during disturbances. In terms of
224 whitefish generations, the eutrophic phase of Lake Constance was of relatively short duration
225 (~30 years or ~6 whitefish generations[51]) and thus, the re-emergence of a deep water
226 ecomorph in the distant future is not to be ruled out, highlighting that the conservation of
227 hybrid populations can be important.

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

239 requiring us to adjust our approaches in conservation biology.

Methods

Sample collection

Historical whitefish scale samples, assembled by David Bittner (see Vonlanthen et al.[3] for details) and collected before the onset of eutrophication in the upper basin of Lake Constance (Fig. 1a), were used to extract DNA from two to eleven individuals of each of four species (*C. arenicolus* (n=3), *C. gutturosus* (n=11), *C. macrophthalmus* (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. arenicolus* (n=5), *C. macrophthalmus* (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.[30] (Extended Data Fig. 4). The phosphorus data (yearly averaged total phosphorus) was retrieved from © BOWIS – Data from the Lake Constance Water Information System ("Bodensee-Wasserinformationssystem") of the International Commission of Lake Constance Water Conservation ("Internationale Gewässerschutzkommission für den Bodensee, IGKB").

DNA extraction

DNA extraction of both historical scale samples and recent fin-clip samples was done using the Qiagen DNeasy blood and tissue kit (Qiagen AG, CH). 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 4M urea[54] and elongated incubation time (overnight) at 37°C were used for lysis of five scales per individual prior to 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 1.12-70.2 ng/µl. 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, CH) 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, CH) using the manufacturer's high sensitivity assay kit. Contamination of DNA samples was measured on a NanoDrop 1000 (Thermo Fisher Scientific AG, CH).

Library preparation and sequencing

For each individual whitefish scale sample, one Illumina paired-end TruSeq DNA Nano library (Illumina GmbH, CH) was produced, while an Illumina paired-end TruSeq DNA PCR-Free library (Illumina GmbH, CH) was prepared for each contemporary fin-clip sample. Library preparation was done by the 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 adapter 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 adapter dimers.

Libraries from historical scale samples and contemporary fin clip samples were prepared according to Extended Data Table 4 and sequenced 2x150 paired-end on either HiSeq 3000 or on Novaseq 6000.

Mapping and filtering of sequencing reads

Poly-G strings at the end of the reads were removed using fastp[55] (Version 0.20.0). Overlapping paired end reads with total length longer than 25 bp were merged using SeqPrep version 1.0 (https://github.com/jstjohn/SeqPrep). Raw reads were aligned to the Alpine whitefish genome assembly[56] (ENA accession: GCA_902810595.1) with bwa mem[57] version 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 (Version 2.20.2; http://broadinstitute.github.io/picard/).

Population genomic analysis

Due to differences in sequencing depth (mean coverage of 6.3x for historical samples and mean coverage of 22.1x for contemporary samples at polymorphic sites included in downstream analyses; see Extended Data Table 4) and to account for possible sequencing errors, we avoided genotype calling whenever possible and only analysed whitefish chromosomes without any potentially collapsed duplicated regions[56]. Instead of hard genotyping, we calculated genotype likelihoods[58] and minor allele frequencies[59,60] at polymorphic sites applying the samtools genotype likelihood model[58] implemented in angsd[61] version 0.925. Only sites covered with at least two reads from every individual (no missing data), passing a p-value cut-off of 10E-6 for being variable[59] and having not more than two

different alleles were included. Reads that did not map uniquely to the reference and had a mapping quality below 30, as well as bases with quality score below 20 were not considered for calculation of genotype likelihoods in the following analyses. We used the following p-value cut-offs for SNP filters implemented in angsd version 0.925: -sb_pval 0.05 -qscore_pval 0.05 -edge_pval 0.05 -mapq_pval 0.05, resulting in a total of 477'981 sites.

We performed a PCA on all polymorphic sites with a minor allele frequency above 5% (222'017 sites) and including all individuals and estimated population structure based on the three most important eigenvectors with PCAngsd[35] version 0.98 and default parameters (see Extended Data Fig. 2 for log-likelihoods of K=1-7). Typically, ancient samples are shifted towards the center of the PC space in relation to modern samples of the same populations[62]. Also, samples sequenced at lower depths[35] or having increased missing data[63] tend to be shifted to the center of PC axes. We here 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.[3], which are based on 10 microsatellite markers (see Extended Data Table 1). We used beagle[64] 4.1 to infer genotypes from the genotype likelihoods at the 477'981 polymorphic sites produced in angsd[61] 0.925 from above and calculated F_{ST} estimates across all three contemporary species pre- and post-eutrophication with the R package hierfstat[65] version 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[61] (version 0.925) option "doAbbababa2", using multiple individuals per population[36]. 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: SSR3669756) from Kjaerner-Semb *et al.*[66] served as outgroup, which defines the ancestral allele (A). We used a block-jackknife approach implemented in angsd[61] 0.925 with a block size of 5 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[67] version 8.2.12. We first calculated genotype likelihoods of the *S. salar* outgroup at all 477'981 polymorphic sites with angsd[61] 0.925, and then inferred genotypes of all individuals including the outgroup and phased these using beagle[64] 4.1. We then thinned this dataset using VCFtools[68] 0.1.16 so that all SNPs were at least 500 bp apart from each other, and then filtered the resulting data set with bcftools 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/edgardomortiz/vcf2phylip). Finally, we used RAxML[67] version 8.2.12 to produce the phylogeny with the ASC_GTRGAMMA substitution model and

100 bootstrap replicates. The resulting phylogeny was plotted with Figtree 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 sub-trees (TWISST)[44]. First, we calculated genotype likelihoods in angsd[61] (0.925), using the same thresholds and filtering parameters as above, but allowing for missing reads in two individuals of the whole data set to increase resolution. Additionally, we genotyped the same *S. salar* individual as used in the ABBA BABA test (see above) at the positions identified to be polymorphic in our dataset. We then inferred genotypes from the likelihoods and phased these genotypes with beagle[64] 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[69] that are relevant for our TWISST approach, and TWISST has been reported to be robust to within-taxon phasing errors[70]. We assessed coverage of each sample at these polymorphic sites with angsd[61] (0.925), and calculated average coverage at across all these polymorphic sites (see Extended Data Table 4). For each discrete 50 kb window across the genome, we computed a maximum likelihood tree including all genotyped samples using PhyML[71] version 3.0 and the script phyml_sliding_windows.py (https://github.com/simonhmartin/genomics_general/blob/master/phylo). 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 (see 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/simonhmartin/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 pre-speciation reversal individuals (p1) of the same species should only be supported within windows that were introgressed by *C. gutturosus* ("introgression topology"; see Fig. 2c). Following Meier *et al.*[72], 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 hetero- and conspecific individuals and the R package iNEXT[45] version 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[73] 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[35] admixture analysis. Second, we used the script ABBABABAwindows.py (https://github.com/simonhmartin/genomics_general/blob/master/ABBABABAwindows.py) to calculate admixture proportions with f_d[74] 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 more than 700 SNPs.

We performed a selection scan using the statistic nSL[46]. nSL is a haplotype based-

410 statistic inferring signatures of selection by combining information on the distribution of
411 fragment lengths defined by pairwise differences with the distribution of the number of
412 segregating sites between all pairs of chromosomes. We first subsetted our data set of
413 genotype likelihoods obtained from angsd[61] (0.925) to only *C. gutturosus* individuals, and
414 then inferred genotypes and phased these using beagle[64] 4.1. We then calculated the
415 unstandardized nSL statistic with the software selscan[75] (version 1.3.0). Because the sample
416 size consisted of 11 individuals, we included low frequency variants. We then used norm[75]
417 (version 1.3.0) to normalize the unstandardized nSL calculations with default parameters in
418 50 kb windows along the genome. We considered windows with more than 51.1% of variable
419 sites (top 1 percentile) with a normalized nSL score above 2 (default) to be under selection.
420 As our sample size was low for such an approach relying on statistical phasing, we
421 additionally calculated Tajima's D[47] in angsd[61] (0.925) based on genotype likelihoods in 50
422 kb windows along the genome, to ensure that the pattern is not heavily impacted by phasing
423 errors. First, we estimated the site allele frequency likelihood in angsd[61] (0.925) and then
424 calculated the maximum likelihood estimate of the folded site allele frequency spectrum using
425 realSFS of angsd[61] (0.925). We used the global site allele frequency spectrum to calculate
426 theta per site in realSFS of angsd[61] (0.925), and then calculated Tajima's D in 50 kb windows
427 using thetaStat of angsd[61] (0.925). We then compared the Tajima's D values of the top 1
428 percentile of 50 kb windows identified to be under selection by nSL to the rest of the genome
429 (Extended Data Fig. 5). Finally, we showed that Tajima's D in the top 1 percentile of 50 kb
430 windows identified to be under selection by nSL differed significantly from the rest of the
431 genome using a two-sided Wilcoxon rank sum test in R 'wilcox.test'[73] ($p<0.01$; $W=$
432 8352543). We assessed how many of these regions under selection introgressed into other
433 whitefish species with a custom R-script. We tested if introgressed regions were enriched for
434 windows under selection by permutation: We randomly sampled the number of windows that
435 were under selection from all windows along the genome and counted the number of overlaps

436 of these randomly sampled windows with the observed introgressed windows. We then
437 compared the expected counts of overlaps of 10'000 permutations with the observed count of
438 overlaps to calculate a p-value.

439 Regions identified as under selection in *C. gutturosus* were further investigated to
440 identify which genes fall within these selected regions. Gene annotations (from the Alpine
441 whitefish genome[56]; ENA accession: GCA_902810595.1) that overlap in their position with
442 the identified windows under selection were identified using bedtools[76] v.2.28.0. Gene
443 enrichment for specific gene ontology (GO) terms (from
444 https://datadryad.org/stash/dataset/doi:10.5061/dryad.xd2547ddf) within these windows was
445 then tested using the R package topGO[77] v.2.38.1 separately for each of the three ontology
446 classes cellular component (CC), biological processes (BP), and molecular function (MF). We
447 used Fisher's exact test applying both the 'weight' and 'elim' algorithms to each ontology
448 class (with no fdr multiple testing correction in accordance to the topGO manual). GO terms
449 that were enriched ($p<0.05$) from both the 'elim' and 'weight' algorithms were reported.

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

456 **Data availability**

457 The raw sequencing files are accessible on SRA (PRJEB43605). Additional
458 supporting data (genotype and genotype likelihood files, morphological raw data, data

underlying Fig. 3, full output table of GO enrichment analysis) is deposited on the eawag research data institutional collections (doi:10.25678/0005AP).

The Alpine whitefish reference genome[56] used was downloaded from ENA and is accessible with accession GCA_902810595.1. The *S. salar* outgroup sample[66] used was downloaded from SRA and is accessible with accession SSR3669756. 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).

Acknowledgements

We thank all professional fisherman for providing specimens, David Bittner for compiling the historical whitefish scale collection, Michael 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 the Internationale Gewässerschutzkomission für den Bodensee (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 center (GDC) at ETH Zurich for bioinformatic support. Further, we are very thankful to the Fish Ecology and Evolution department of Eawag, but especially to Carmela Dönz, Blake Matthews, David Marques, Kotaro Kagawa, Joana Meier and Johanna Ten Brink for feedback, comments and ideas on this manuscript. Also, thanks to Owen 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

OS conceived of the study, DF, OS and PGDF designed and conceptualized it. PGDF managed and supervised the study. OMS collected contemporary specimens and collected and analysed morphological data. RDK contributed to DNA extraction and genomic analysis. DF analysed genomic data and visualized the results. DF wrote the original manuscript draft with input from OS and PGDF. All authors edited and reviewed the final manuscript.

Competing interest declaration

The authors declare no competing interests.

Figure Legends

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). The four whitefish species are indicated by distinct colours. **b)** In genomic PCA space, the same species post-eutrophication (circles) are less distinct than pre-eutrophication (crosses), and one species is completely lost (i.e. *C. gutturosus*). **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.

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. Significant tests for introgression during speciation reversal are indicated as black arrows, and non-significant tests as dashed grey arrows. Severity of niche collapse is indicated by yellow (highest) to blue (lowest) shading of the water. **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; see Extended Data Table 2). c**)** The topology shows the grouping of species for the underlying D-statistic test.

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* and orange *C. arenicolus*) 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 introgressed windows between individuals (pairwise comparison yellow to red colour scale) and the absolute count of introgressed windows for each individual (blue colour scale).

References

1 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, doi:10.1146/annurev-arplant-042817-040348 (2018).

2 Rhymer, J. M. & Simberloff, D. Extinction by hybridization and introgression. *Annu. Rev. Ecol. Syst.* **27**, 83-109, doi:10.1146/annurev.ecolsys.27.1.83 (1996).

3 Vonlanthen, P. *et al.* Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482**, 357-U1500, doi:10.1038/nature10824 (2012).

4 Seehausen, O. Conservation: Losing biodiversity by reverse speciation. *Curr. Biol.* **16**, R334-R337, doi:10.1016/j.cub.2006.03.080 (2006).

5 Seehausen, O., Takimoto, G., Roy, D. & Jokela, J. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol. Ecol.* **17**, 30-44, doi:10.1111/j.1365-294X.2007.03529.x (2008).

6 Kearns, A. M. *et al.* Genomic evidence of speciation reversal in ravens. *Nat. Commun.* **9**, doi:10.1038/s41467-018-03294-w (2018).

7 Meier, J. I. *et al.* Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 11, doi:10.1038/ncomms14363 (2017).

8 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

9 Barlow, A. *et al.* Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**, 1563-1570, doi:10.1038/s41559-018-0654-8 (2018).

10 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, doi:10.1038/s41559-019-0881-7 (2019).

11 Palkopoulou, E. *et al.* A comprehensive genomic history of extinct and living elephants. *Proc. Natl. Acad. Sci.U.S.A.* **115**, E2566-E2574, doi:10.1073/pnas.1720554115 (2018).

12 Ottenburghs, J. Ghost introgression: Spooky gene flow in the distant past. *BioEssays* **42**, 2000012, doi:10.1002/bies.202000012 (2020).

13 Schluter, D. Evidence for ecological speciation and its alternative. *Science* **323**, 737-741, doi:10.1126/science.1160006 (2009).

14 Nosil, P. *Ecological Speciation*. (Oxford University Press, Oxford, 2012).

15 Schluter, D. *The Ecology of Adaptive Radiation*. (Oxford University Press, Oxford, 2000).

16 Rundle, H. D. & Nosil, P. Ecological speciation. *Ecol. Lett.* **8**, 336-352, doi:10.1111/j.1461-0248.2004.00715.x (2005).

17 Yeaman, S. & Whitlock, M. C. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897-1911, doi:10.1111/j.1558-5646.2011.01269.x (2011).

18 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, doi:10.1002/ece3.413 (2012).

19 Seehausen, O., van Alphen, J. J. M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808-1811, doi:10.1126/science.277.5333.1808 (1997).

20 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, doi:10.1111/j.1365-294X.2005.02794.x (2006).

21 Grant, P. R. & Grant, B. R. Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci.U.S.A.* **116**, 23216-23224, doi:10.1073/pnas.1913534116 (2019).

22 Kagawa, K. & Takimoto, G. Hybridization can promote adaptive radiation by means of transgressive segregation. *Ecol. Lett.* **21**, 264-274, doi:10.1111/ele.12891 (2018).

23 Feller, A. F. *et al.* Rapid generation of ecologically relevant behavioral novelty in experimental cichlid hybrids. *Ecol. and Evol.* **10**, 7445-7462, doi:10.1002/ece3.6471 (2020).

24 Pfennig, K. S., Kelly, A. L. & Pierce, A. A. Hybridization as a facilitator of species range expansion. *Proc. R. Soc. B* **283**, doi:10.1098/rspb.2016.1329 (2016).

25 Lamichhaney, S. *et al.* Rapid hybrid speciation in Darwin's finches. *Science* **359**, 224-227, doi:10.1126/science.aao4593 (2018).

26 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, doi:10.1186/1471-2148-13-108 (2013).

27 Hudson, A. G., Vonlanthen, P. & Seehausen, O. Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proc. R. Soc. B* **278**, 58-66, doi:10.1098/rspb.2010.0925 (2011).

28 Jacobs, A. *et al.* Rapid niche expansion by selection on functional genomic variation after ecosystem recovery. *Nat. Ecol. Evol.* **3**, 77-86, doi:10.1038/s41559-018-0742-9 (2019).

29 Steinmann, P. Monographie der schweizerischen Koregonen. Beitrag zum Problem der Entstehung neuer Arten. Spezieller Teil. *Schweiz. Z. Hydrologie* **12**, 340-491, doi:doi:10.1007/BF02486030 (1950).

30 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). *Zookeys* **989**, 79-162, doi:10.3897/zookeys.989.32822 (2020).

31 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, doi:10.1111/eva.12446 (2016).

32 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, doi:10.1002/ece3.4375 (2018).

33 Vonlanthen, P. *et al.* Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *J. Evol. Biol.* **22**, 498-514, doi:10.1111/j.1420-9101.2008.01670.x (2009).

34 Feulner, P. G. D. & Seehausen, O. Genomic insights into the vulnerability of sympatric whitefish species flocks. *Mol. Ecol.* **28**, 615-629, doi:10.1111/mec.14977 (2019).

35 Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719-731, doi:10.1534/genetics.118.301336 (2018).

606 36 Soraggi, S., Wiuf, C. & Albrechtsen, A. Powerful inference with the d-statistic on low-
607 coverage whole-genome data. *G3-Genes Genomes Genet.* **8**, 551-566,
608 doi:10.1534/g3.117.300192 (2018).
609 37 Wahl, B. & Loffler, H. Influences on the natural reproduction of whitefish (*Coregonus*
610 *lavaretus*) in Lake Constance. *Can. J. Fish. Aquat. Sci.* **66**, 547-556, doi:10.1139/f09-019
611 (2009).
612 38 Nümann, W. The Bodensee: Effects of exploitation and eutrophication on the Salmonid
613 community. *J. Fish. Res. Board of Can.* **29**, 833-884, doi:10.1139/f72-127 (1972).
614 39 Deufel, J., Löffler, H. & Wagner, B. Auswirkungen der Eutrophierung und anderer
615 anthropogener Einflüsse auf die Laichplätze einiger Bodensee-Fischarten. *Österreichs*
616 *Fischerei* **39**, 325-336 (1986).
617 40 Straile, D. & Geller, W. Crustacean zooplankton in Lake Constance from 1920 to 1995:
618 response to eutrophication and re-oligotrophication. *Adv. Limnol.* **58**, 255-274, doi:http://nbn-
619 resolving.de/urn:nbn:de:bsz:352-opus-39858 (1998).
620 41 Eby, L. A., Crowder, L. B., McClellan, C. M., Peterson, C. H. & Powers, M. J. Habitat
621 degradation from intermittent hypoxia: impacts on demersal fishes. *Mar. Ecol. Prog. Ser.* **291**,
622 249-261, doi:10.3354/meps291249 (2005).
623 42 Powers, S. P. *et al.* Effects of eutrophication on bottom habitat and prey resources of demersal
624 fishes. *Mar. Ecol. Prog. Ser.* **302**, 233-243, doi:10.3354/meps302233 (2005).
625 43 Caires, A. M., Chandra, S., Hayford, B. L. & Wittmann, M. E. Four decades of change:
626 dramatic loss of zoobenthos in an oligotrophic lake exhibiting gradual eutrophication. *Freshw.*
627 *Sci.* **32**, 692-705, doi:10.1899/12-064.1 (2013).
628 44 Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the
629 genome using topology weighting. *Genetics* **206**, 429-438, doi:10.1534/genetics.116.194720
630 (2017).
631 45 Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of
632 species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451-1456, doi:10.1111/2041-
633 210x.12613 (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, doi:10.1093/molbev/msu077 (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, doi:10.1186/1471-2105-14-289 (2013).

48 Doenz, C. J. & Seehausen, O. Rediscovery of a presumed extinct species, *Salvelinus profundus*, after re-oligotrophication. *Ecology* **101**, doi:10.1002/ecy.3065 (2020).

49 Alexander, T. & Seehausen, O. *Diversity, distribution and community composition of fish in perialpine lakes - "Projet Lac" synthesis report.* (Bundesamt für Umwelt BAFU, in press).

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, doi:10.1111/j.1558-5646.2011.01320.x (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, doi:10.1111/j.1752-4571.2008.00054.x (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, doi:10.1016/j.tree.2017.12.008 (2018).

53 Seehausen, O. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198-207, doi:10.1016/j.tree.2004.01.003 (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, doi:10.1034/j.1601-5223.2003.01503.x (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, doi:10.1093/bioinformatics/bty560 (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, doi:10.1111/1755-0998.13187 (2020).

57 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (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, doi:10.1093/bioinformatics/btr509 (2011).

59 Kim, S. Y. *et al.* Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinform.* **12**, doi:10.1186/1471-2105-12-231 (2011).

60 Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Association testing for next-generation sequencing data using score statistics. *Genet. Epidemiol.* **36**, 430-437, doi:10.1002/gepi.21636 (2012).

61 Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of next generation sequencing data. *BMC Bioinform.* **15**, doi:10.1186/s12859-014-0356-4 (2014).

62 Duforet-Frebourg, N. & Slatkin, M. Isolation-by-distance-and-time in a stepping-stone model. *Theor. Popul. Biol.* **108**, 24-35, doi:10.1016/j.tpb.2015.11.003 (2016).

63 Meisner, J., Liu, S., Huang, M. & Albrechtsen, A. Large-scale inference of population structure in presence of missingness using PCA. *Bioinformatics*, doi: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, doi:10.1086/521987 (2007).

65 Goudet, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184-186, doi:10.1111/j.1471-8286.2004.00828.x (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**, doi:10.1186/s12864-016-2867-z (2016).

67 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

68 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158, doi:10.1093/bioinformatics/btr330 (2011).

69 Bukowicki, 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, doi:10.1111/1755-0998.12516 (2016).

70 Marburger, S. *et al.* Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* **10**, doi:10.1038/s41467-019-13159-5 (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, doi:10.1093/sysbio/syq010 (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, doi:10.1093/molbev/msy051 (2018).

73 *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, Vienna, 2018).

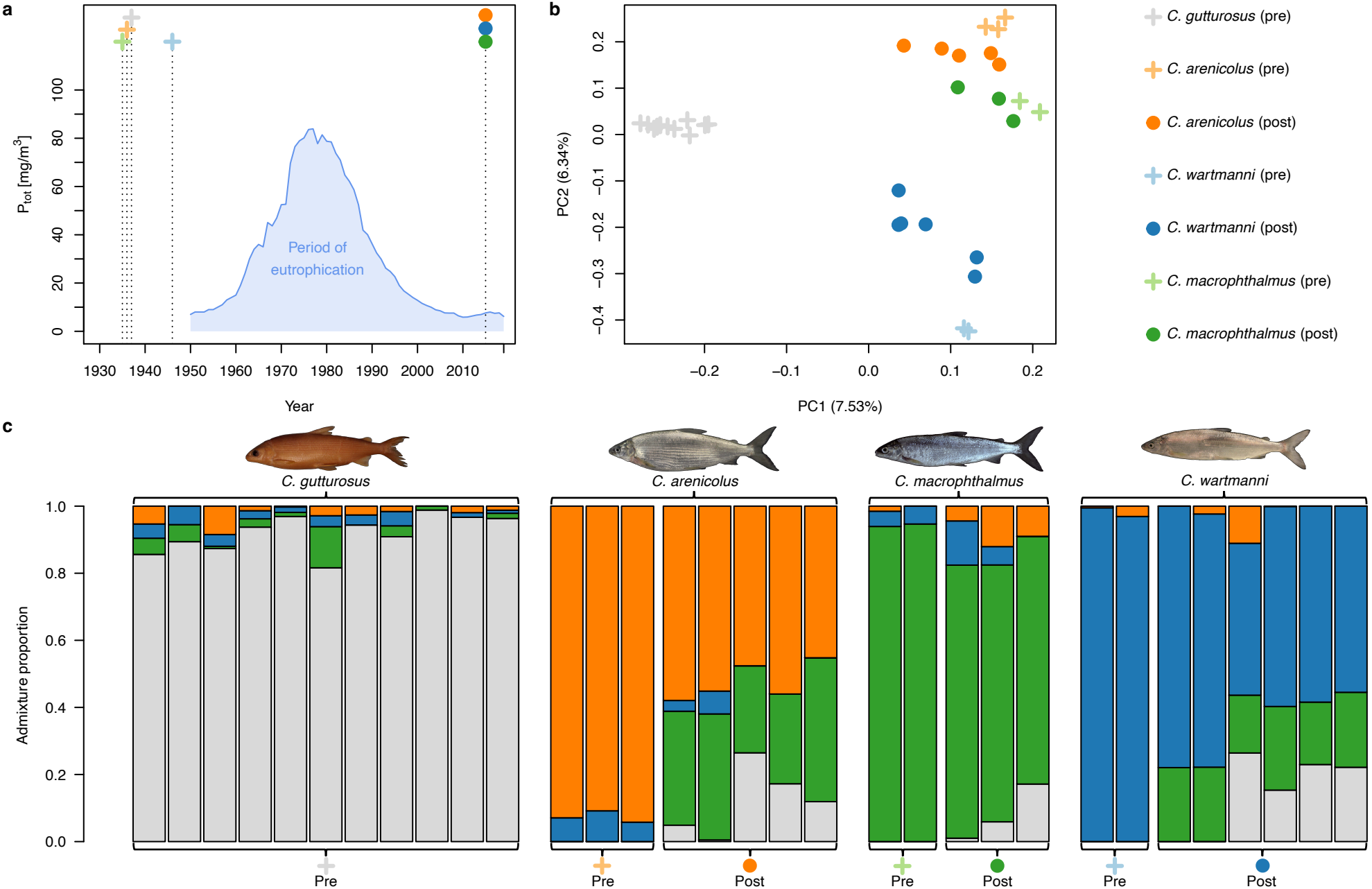
74 Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244-257, doi:10.1093/molbev/msu269 (2015).

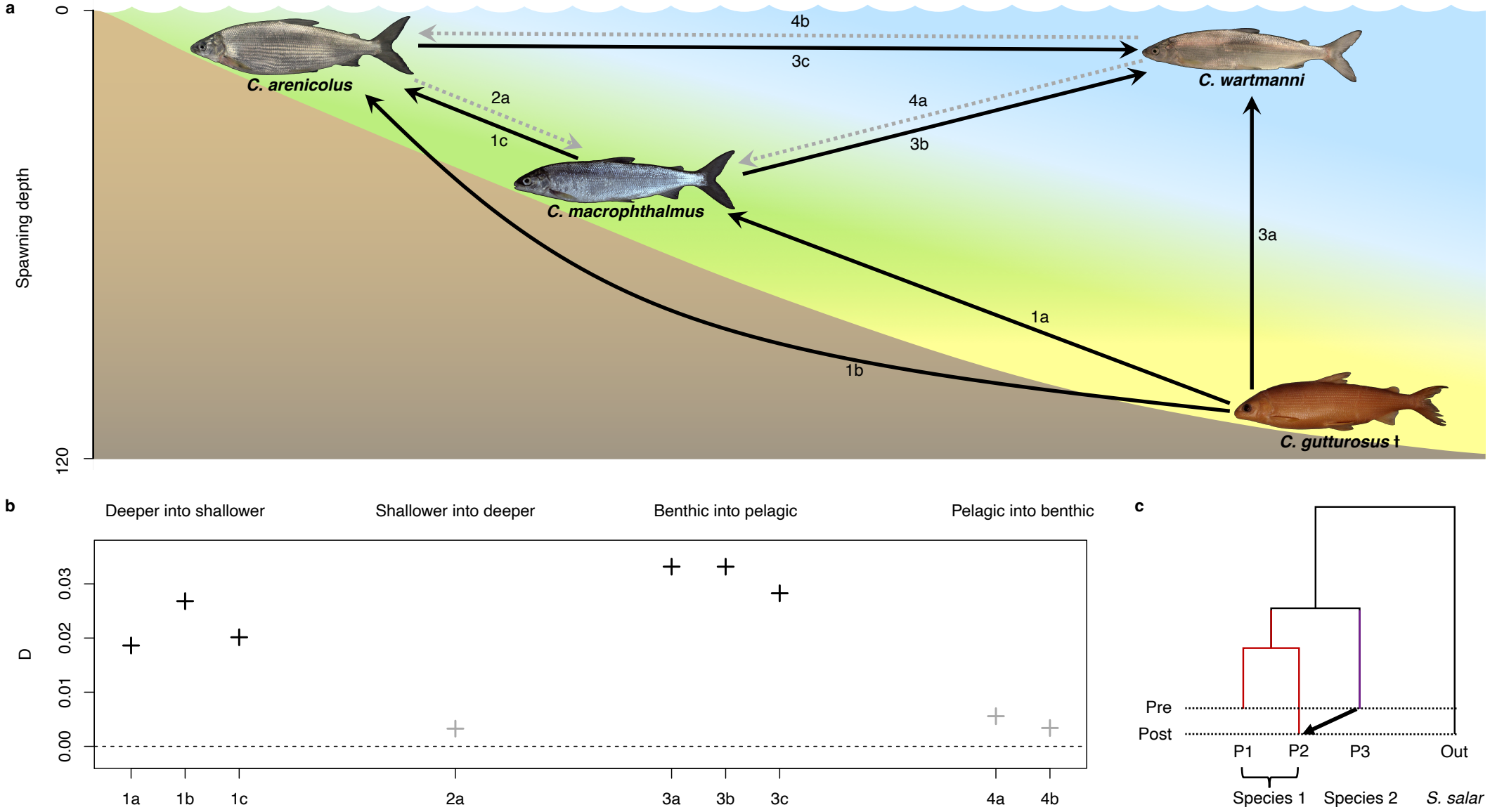
75 Szpiech, Z. A. & Hernandez, R. D. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824-2827, doi:10.1093/molbev/msu211 (2014).

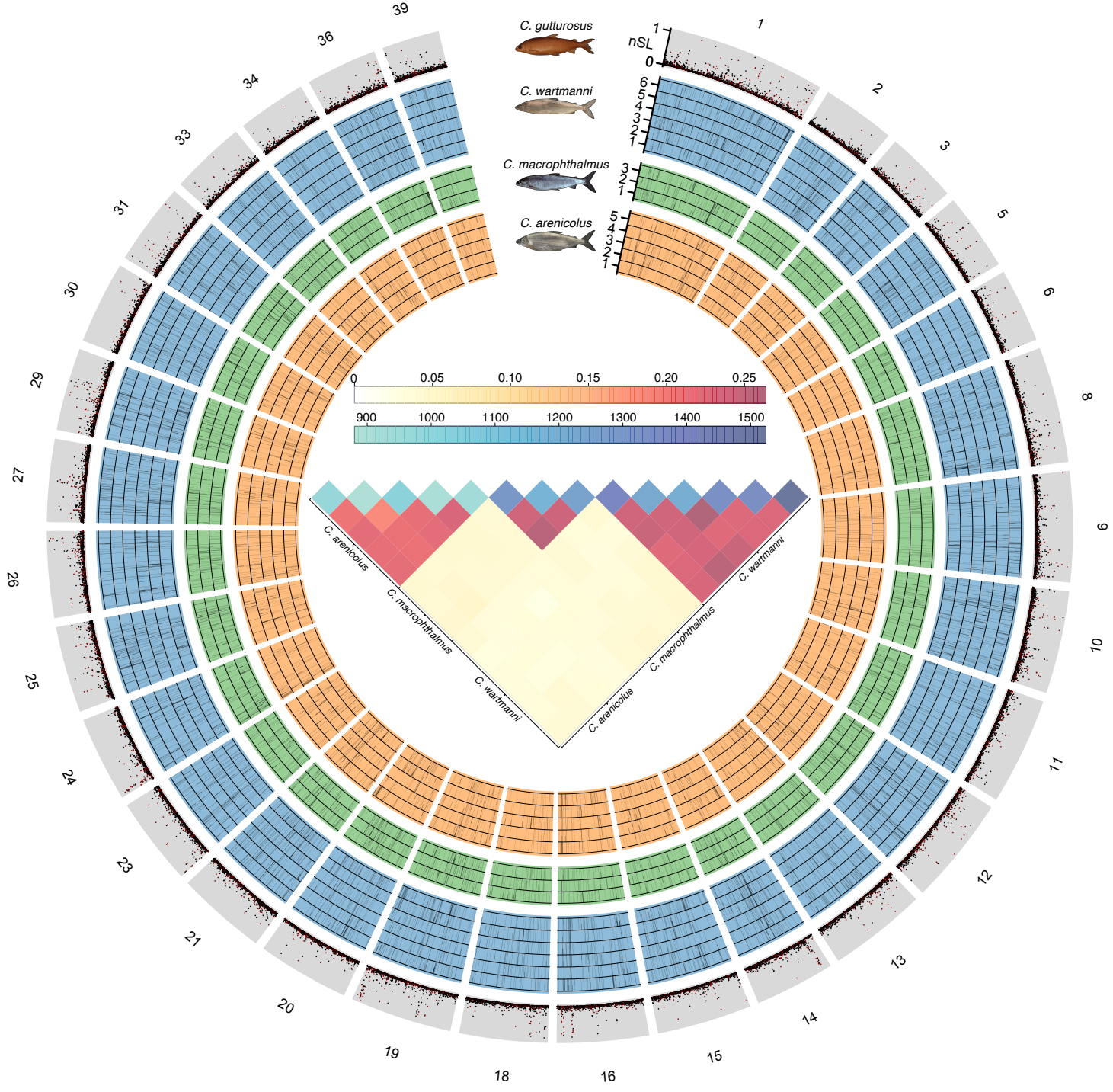
76 Quinlan, A. R. BEDTools: The Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1-11.12.34, doi:https://doi.org/10.1002/0471250953.bi1112s47 (2014).

77 Alexa, A. & **Rahnenfuhrer**, J. *TopGO: Enrichment Analysis for Gene Ontology.* R package version 2.42.0. doi: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*. doi:10.25678/0005AP (2021).


a
P_{tot} [mg/m^3]
0
20
40
60
80
100
1930
1940
1950
1960
1970
1980
1990
2000
2010
Year
Period of
eutrophication
b
PC2 (6.34%)
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.2
−0.1
0.0
0.1
0.2
PC1 (7.53%)
C. gutturosus (pre)
C. arenicolus (pre)
C. arenicolus (post)
C. wartmanni (pre)
C. wartmanni (post)
C. macrophthalmus (pre)
C. macrophthalmus (post)
c
C. gutturosus
C. arenicolus
C. macrophthalmus
C. wartmanni
Admixture proportion
1.0
0.8
0.6
0.4
0.2
0.0
Pre
Pre
Post
Pre
Post
Pre
Post


a
Spawning depth
0
120
C. arenicolus
C. wartmanni
C. macrophthalmus
C. gutturosus †
4b
3c
2a
1c
4a
3b
3a
1a
1b
b
Deeper into shallower
Shallower into deeper
Benthic into pelagic
Pelagic into benthic
D
0.00
0.01
0.02
0.03
1a
1b
1c
2a
3a
3b
3c
4a
4b
c
Pre
Post
P1
P2
P3
Out
Species 1
Species 2
S. salar


C. gutturosus
C. wartmanni
C. macrophthalmus
C. arenicolus
nSL
1
0
6
5
4
3
2
1
3
2
1
5
4
3
2
1
0
0.05
0.10
0.15
0.20
0.25
900
1000
1100
1200
1300
1400
1500
C. arenicolus
C. macrophthalmus
C. wartmanni
C. wartmanni
C. macrophthalmus
C. arenicolus
1
2
3
5
6
8
9
10
11
12
13
14
15
16
18
19
20
21
23
24
25
26
27
29
30
31
33
34
36
39