The genomic consequences of extinction by hybridization during eutrophication-induced speciation reversal

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

David Frei

von Diepoldsau-Schmitter

LeiterInnen der Arbeit:
Dr. Philine Feulner & Prof. Dr. Ole Seehausen
Institut für Ökologie und Evolution

The genomic consequences of extinction by hybridization during eutrophication-induced speciation reversal

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

David Frei

von Diepoldsau-Schmitter

LeiterInnen der Arbeit:
Dr. Philine Feulner & Prof. Dr. Ole Seehausen
Institut für Ökologie und Evolution

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, 28.3.2023

Der Dekan
Prof. Dr. Marco Herwegh

Table of Contents

General Introduction

General Introduction

Understanding the evolution of global biodiversity is the ultimate aim of evolutionary biology. Macroevolutionary patterns of biodiversity are shaped by the interaction of speciation and extinction (Vamosi, Magallon, Mayrose, Otto, & Sauquet, 2018). To date, most species that ever existed are extinct (Jablonski, 2004) and species go extinct at an unprecedented rate during the contemporary biodiversity crisis (*IPBES*, 2019). Hence, studying the mechanisms underlying both speciation and extinction is critical for a better understanding of contemporary biodiversity patterns (Vamosi et al., 2018), as well as to mitigate anthropogenic effects on biodiversity (Seehausen, Takimoto, Roy, & Jokela, 2008).

Explaining the heterogeneity in the rate of diversification across the tree of life remains a fundamental question in the field of evolutionary biology (Scholl & Wiens, 2016; Wagner, Harmon, & Seehausen, 2012). Adaptive radiations, characterized by the rapid buildup of sympatric species diversity from a single lineage through adaptation to diverse ecological niches, provide outstanding opportunities to study the influence of intrinsic biological, as well as environmental factors on the evolution of biodiversity (Schluter, 2000; Seehausen, 2004). The evolution of adaptive radiations with numerous sympatric species exhibiting diverse ecological adaptations is dependent on heritable variation in traits related to ecological diversification and reproductive isolation (Meier et al., 2017). Recent empirical work has demonstrated that the evolution of an adaptive radiation from its ancestral lineage requires the coincidence of ecological opportunity with genetic opportunity for hybridization (Meier et al., 2019).

Speciation requires the evolution of reproductive isolation. In evolutionary young adaptive radiations, reproductive isolation is often dependent on extrinsic pre- and postzygotic mechanisms that are based on the interaction of intrinsic lineage traits (such as mate choice, choice of breeding site, or timing of reproduction) with specific features of the environment

(such as temperature, precipitation, wind or currents, or substrate) (Seehausen, 2004, 2006; Seehausen et al., 2008). The evolution of intrinsic postzygotic incompatibilities often requires a period of geographical isolation (Seehausen et al., 2014) and needs long timespans to evolve (Schluter, 1996; Stelkens, Young, & Seehausen, 2010). Before such a phase of geographical isolation, reproductive isolation between species is dependent on environmental characteristics and the maintenance of reproductive isolation requires the persistence of environmental heterogeneity (Ghosh & Joshi, 2012; Seehausen et al., 2014; Yeaman & Whitlock, 2011). As a consequence, reproductive isolation between many evolutionary young species can be weakened when the environment changes, and the resulting hybridization and introgression can lead to the collapse of species into a hybrid swarm within only few generations, a process called speciation reversal (Seehausen, 1997; Seehausen, 2006; Seehausen 2008; Vonlanthen et al., 2012). Hence, the process of speciation remains reversible for millions of years until complete reproductive isolation has evolved (Bolnick & Near, 2005; Mendelson, Imhoff, & Venditti, 2007; Stelkens, Schmid, & Seehausen, 2015; Stelkens et al., 2010).

Anthropogenic environmental change is causing a massive decline of contemporary biodiversity on a global scale (*IPBES*, 2019). Counteracting the current biodiversity crisis resulting from human influence all over the globe requires to understand the effects of changing environments on processes that generate and maintain species diversity (Seehausen et al., 2008). Species can be driven to extinction when the habitat in which they evolved in is affected by environmental change in a way that it does no longer sustain a population size that allows the species to survive (Vonlanthen et al, 2012). The consequence is a declining population and ultimately extinction. Many of the current conservation efforts focus on habitat restoration to prevent species extinctions by demographic decline (Vonlanthen et al, 2012). However, apart from demographic decline, species can go extinct by hybridization as a response to anthropogenic environmental change (Grabenstein & Taylor, 2018; Seehausen et

al., 2008; Taylor et al., 2006; Todesco et al., 2016; Vonlanthen et al., 2012). As long as reproductive isolation between species is dependent on environmental heterogeneity and intrinsic incompatibilities are weak or absent, as it is often the case in evolutionary young adaptive radiations or species evolved through ecological speciation, environmental change can weaken reproductive isolation and result in extinction through hybridization (Grabenstein & Taylor, 2018; Gilman & Behm, 2011; Seehausen, 2006; Todesco et al., 2016).

Adaptive radiations are characterized by the fine-scale partitioning of ecological niches between multiple sympatric species (Schluter, 2000; Seehausen, 2004). When environmental change affects reproductive isolation between multiple members of an adaptive radiation at the same time, it can result in the collapse, respectively in the extinction, of numerous species within a very short time span (Seehausen et al., 2008; Vonlanthen et al., 2012). As a large fraction of contemporary global biodiversity is sensitive to hybridization-driven dynamics, extinction by hybridization induced by environmental change might represent a very relevant, but currently underestimated process of extinction during the current biodiversity crisis (Seehausen, 2006; Seehausen et al., 2008). In turn, efficient conservation measures need to consider such processes of extinction and aim to protect the mechanisms that generate and maintain biodiversity to mitigate the effects of anthropogenic environmental change (Seehausen et al., 2008).

As environments are dynamic and under continuous change, it is well possible that extinction by hybridization is not confined to a context of anthropogenic ecosystem disturbance. Such processes could also be induced by natural environmental change, e.g., through climatic oscillation. Considering the long timespan complete reproductive isolation needs to evolve (Schluter, 1996; Stelkens et al., 2010) and the large fraction of biodiversity that can still hybridize with closely related species, the collapse of species by hybridization as response to environmental change might represent a relevant evolutionary trajectory

(Grabenstein & Taylor, 2018; Seehausen et al., 2008). In theory, hybridization induced by (natural or anthropogenic) environmental change could be responsible for many instances where signals of hybridization between coexisting species have been detected, but the causes and circumstances of the hybridization event remain unknown.

Recent work showed that hybridization can produce new trait combinations through the combination of alleles from both parental lineages, thereby possibly allowing the exploitation of resources that were unavailable to the parental species (Feller et al., 2020; Kagawa & Takimoto, 2018). Hence, hybrid populations can exhibit increased evolvability, because as a consequence of hybridization, their genomic variation is increased (Grant & Grant, 2019; Marques, Meier, & Seehausen, 2019). In consequence, hybridization can facilitate rapid adaptation and speciation (Meier et al., 2017; Meier et al., 2019). To date, there is growing evidence for the involvement of hybridization in adaptation and diversification across various groups of organisms (e.g., fish (Meier et al., 2017), birds (Lamichhaney et al., 2018), butterflies (Dasmahapatra et al., 2012) or humans (Reilly, Tjahjadi, Miller, Akey, & Tucci, 2022)). Recent research demonstrated that hybridization can facilitate the onset of entire adaptive radiations ("hybrid swarm origin" hypothesis of adaptive radiations; Meier et al., 2017; Meier et al., 2019; Seehausen, 2004). However, it has been hypothesized that hybridization can also fuel the continuation of diversification processes beyond the first speciation events of an adaptive radiation ("syngameon" hypothesis of adaptive radiation; Seehausen, 2004).

The increased genomic variation and enhanced adaptive potential of hybrid populations might enable fast evolutionary responses to new selective pressures when the environment is rapidly changing (Grabenstein & Taylor, 2018; Grant & Grant, 2019). Such elevated evolvability and the ability to rapidly adapt to changing environmental conditions might be beneficial in the context of anthropogenic environmental change, potentially

facilitating the survival of species or even entire radiations (Aitken & Whitlock, 2013; Stelkens, Brockhurst, Hurst, & Greig, 2014; Vedder et al., 2022). Resilience is broadly defined as the capability of maintain or regain functioning in the face of disturbance (Cahill, Chandola, & Hager, 2022). Hence, ecological resilience is related to the potential of an ecosystem to maintain its ecological function throughout disturbance (Holling, 1973). As hybridization can increase the adaptive potential of a population or species, it can facilitate the rapid adaptation to the changed environmental conditions. Thereby, hybridization might be able to increase the resilience of a population, a species or even of an entire adaptive radiation during environmental change.

The evaluation of the genomic consequences of environmental change heavily relies on an appropriate baseline from before the onset of environmental change (Jensen & Leigh, 2022). Subjective thresholds for acceptable environmental conditions can be lowered in consequence of increased levels of actual environmental degradation, a process called "shifting baselines syndrome" (Pauly, 1995). In the absence of detailed historical records about the undisturbed state of an ecosystem, new generations might consider the situation in which they have been raised as a suitable baseline level (Soga & Gaston, 2018). Thus, natural history collections and museum samples can be fundamental to assess genomic consequences of anthropogenic environmental change (Jensen & Leigh, 2022). Samples collected before the onset of environmental disturbance can be crucial to estimate the undisturbed state of ecosystem and to evaluate its diversity (Jensen & Leigh, 2022). Such historical samples have the potential to document biodiversity loss against the shifting baseline syndrome and thereby advance the efficient conservation of biodiversity (Jensen & Leigh, 2022).

Within the Alpine whitefish radiation, more than 30 whitefish species have been taxonomically described in nine different Swiss lakes or lake systems, with up to six species per lake (Doenz, Bittner, Vonlanthen, Wagner, & Seehausen, 2018; Selz, Doenz, Vonlanthen,

& Seehausen, 2020; Steinmann, 1950). In some of these lakes, certain similar ecomorphs have evolved independently, while other ecomorphs are lake-specific (Hudson, Vonlanthen, & Seehausen, 2011). In general, small-bodied species with many densely spaced gill-rakers are adapted to feed on zooplankton and typically in deep regions of the lake during summer or winter, whilst large-bodied and winter-spawning species usually have a lower number of more sparsely spaced gill-rakers as adaptation to feeding on benthic macroinvertebrates (Vonlanthen et al., 2012). Furthermore, profundal species can be found in some of the Swiss lakes, typically spawning during summer in profundal regions below the thermocline (Doenz et al., 2018; Selz et al., 2020; Steinmann, 1950; Vonlanthen et al., 2012).

The independent evolution of similar phenotypes in independent lake systems implies a crucial role of divergent selection during the evolution of the Swiss Alpine whitefish radiation (Hudson, Vonlanthen, Bezault, & Seehausen, 2013; Hudson et al., 2011; Ostbye, Bernatchez, Naesje, Himberg, & Hindar, 2005; Praebel et al., 2013) and represents a possible example of species-for-species matching as described in *Anolis* lizards (Mahler, Ingram, Revell, & Losos, 2013). The parallelism in the evolution of phenotypes and the high level of replication makes the Swiss Alpine whitefish radiation an outstanding study system to study questions related to the evolution of adaptation (Jacobs et al., 2019), diversification (Vonlanthen et al., 2009), adaptive radiation and parallel evolution (De-Kayne et al., 2022), but also to investigate the collapse of adaptive radiations in response to environmental change (Feulner & Seehausen, 2019; Vonlanthen et al., 2012).

Reproductive isolation between sympatric Alpine whitefish species is sensitive to changes of habitat characteristics, because spawning niche differentiation, and in turn reproductive isolation, is contingent on the persistence of fine-scale depth-related differences in the specific lacustrine habitat (Hudson, Lundsgaard-Hansen, Lucek, Vonlanthen, & Seehausen, 2016). Hence, a change in environmental conditions can weaken reproductive

isolation and result in hybridization and introgression between sympatric whitefish species (Hudson et al., 2013; Vonlanthen et al., 2012). Many Swiss lakes faced a period of anthropogenic eutrophication during the last century (Vonlanthen et al., 2012). The consequence were dramatic losses of Alpine whitefish diversity in many Swiss lakes (Feulner & Seehausen, 2019; Hudson et al., 2013; Vonlanthen et al., 2012). Largely unnoticed by the public, ~29% of Alpine whitefish species went extinct during the period of anthropogenic eutrophication between ~1950s and ~2000 (Vonlanthen et al., 2012). Anthropogenic eutrophication resulted in the loss of benthic deep-water spawning habitats as consequence of decreased oxygen concentrations at the water-sediment interface (Deufel, Löffler, & Wagner, 1986; Grimaldi & Numann, 1972), the location of whitefish egg development. Thereby, eutrophication reduced the available reproductive niche space (Vonlanthen et al., 2012; Vonlanthen et al., 2009). At the same time, increased productivity during the period of eutrophication resulted in an increase of zooplankton density, but decreased zoo benthos densities, diminishing foraging niche distinctiveness (Alexander, Vonlanthen, & Seehausen, 2017; Hudson et al., 2013; Vonlanthen et al., 2012). In sum, eutrophication had dual consequences for reproductive isolation between sympatric Alpine whitefish species: By decreasing available reproductive niche space, it led to a reduction of extrinsic prezygotic isolation (Vonlanthen et al., 2012; Vonlanthen et al., 2009). By diminishing foraging and reproductive niche distinctiveness, it weakened divergent selection between niches and associated extrinsic postzygotic isolation (Hudson et al., 2013; Vonlanthen et al., 2012). The consequence was hybridization and introgression resulting in speciation reversal (Vonlanthen et al., 2012).

Benthic profundal species, adapted to spawn in deep waters and to feed on benthic macroinvertebrates, were particularly impacted by the loss of deep-water spawning grounds and reduced zoo benthos densities in their deep-water habitats due to the anoxic conditions during the period of eutrophication (Eby, Crowder, McClellan, Peterson, & Powers, 2005;

Powers et al., 2005; Steinmann, 1950; Vonlanthen et al., 2012). As a result, profundal species went extinct in lakes that were exposed to severe eutrophic conditions (e.g. Lake Constance) and can today only be found in Lakes that faced relatively mild eutrophic conditions (e.g., lakes Thun, Brienz, Walen or Lucerne) (Vonlanthen et al., 2012). Thus, understanding profundal adaptation and its genomic basis is important for a better understanding of the diversity loss within the Alpine whitefish radiation as a consequence of anthropogenic eutrophication.

Research questions

The aim of this thesis is to investigate the genomic consequences of eutrophication-induced speciation reversal. The focus is on the extinction of the profundal *Coregonus gutturosus* in Lake Constance, but also on the consequences of speciation reversal for extant species of the Swiss Alpine whitefish radiation. By sequencing genomes of pre-, during- and post-eutrophication populations of all Lake Constance whitefish species, including the extinct *C. gutturosus*, this thesis addresses the following questions:

I. The characterization of introgressed variation: Does introgression during speciation reversal involve adaptive genomic variation?

II. Can genomic variation that is exchanged during the speciation reversal process be re-used to generate novel combinations of genotypes that facilitate the colonization of new niches?

III. Temporal dynamics of the collapse of an adaptive radiation: How does the genomic variation of an adaptive radiation (and all single species) change in response to environmental disturbance?

Profundal habitats of the peri-alpine lakes were affected most severely during eutrophication. In turn, many profundal spawning whitefish species were declining during the

last century, and some of them went extinct. Thus, it is instrumental to understand the genomic basis of adaptation to profundal habitats. Hence, another part of this thesis is concerned with the question:

IV. Is there a shared genomic basis of adaptation to profundal habitats across the extant deep-water species of the Alpine whitefish radiation?

Thesis Overview

I. Alpine whitefish radiation retains adaptive genomic variation from extinct species following speciation reversal

We made use of the historical scale sample collection used by Vonlanthen et al. (2012) to compare whole-genome re-sequencing data of all Lake Constance whitefish species from samples collected before (before 1950) and after (2015) the period of anthropogenic eutrophication. Whole-genome resequencing data of eleven individuals of the extinct *C. gutturosus* allowed us to perform a selection scan in the now extinct profundal species. The results presented in this chapter demonstrate that genomic regions with signatures of positive selection in the now extinct *C. gutturosus* introgressed into extant whitefish species during speciation reversal. Despite the extinction of *C. gutturosus*, substantial fractions of its genome, including regions shaped by positive selection, persist within surviving species as a consequence of introgressive hybridization during eutrophication.

II. Introgression from extinct species facilitates adaptation to its vacated niche

Today, Lake Constance has returned to its original oligotrophic conditions due to restoration efforts. In Chapter I, we demonstrated that speciation reversal transferred alleles that have evolved in the now extinct deep-water species *C. gutturosus* (and thus are potentially adaptive in deep water) into the three surviving whitefish species of the lake. We sampled a depth gradient on a known spawning ground of the extant *C. macrophthalmus* and found that the species has extended its spawning depth range and is currently spawning deeper than historically recorded. We generated whole-genome re-resequencing data of 96 individuals caught at six different depths (4m, 12m, 20m, 40m, 60m and 90m). The results revealed that genomic variation introgressed from the extinct profundal species *C. gutturosus* is potentially facilitating adaptation to deep-water spawning grounds in *C. macrophthalmus*.

III. Ecological disturbance reduces genomic diversity across an Alpine whitefish adaptive radiation

During the period of eutrophication, whitefish species of Lake Constance extensively hybridized, and all three surviving species were exposed to dramatic changes in environmental conditions. The ecological, evolutionary and demographic processes resulting from such ecological disturbance can have contrasting effects on the genomic variation of a species. We made use of the historical fish scale collection of Vonlanthen et al. (2012) to generate population-level whole-genome resequencing data from samples collected before, during and after the period of eutrophication for each of the Lake Constance whitefish species. We used this data to document the changes in genomic diversity through time and over the period of anthropogenic eutrophication.

IV. The genomic basis of adaptation to profundal habitats across the Swiss Alpine whitefish radiation

Populations of many profundal whitefish species were rapidly declining during the period of anthropogenic eutrophication. Today, profundal species can still be found in Lakes Thun, Lucerne and Walen, whilst those that inhabited e.g., Lakes Constance and Zug went extinct. We combined newly generated and existing whole-genome resequencing data of all taxonomically described whitefish species of lakes Constance, Lucerne, Thun and Walen. We used this dataset to produce a phylogenetic tree including all species from all four sampled lakes, and thereby confirmed the reciprocal monophyly of the four lake-specific species flocks. Further, we reveal the genomic landscape of parallel differentiation between the profundal and other species in each lake.

V. Sequencing platform shifts provide opportunities but pose challenges for combining genomic data sets

The combination of sequencing data initially generated for different studies is instrumental to build large genomic datasets, enabling to study fundamental questions in evolutionary biology. However, technological advances and the development of new sequencing platforms can be a challenge in this regard, because minor differences in sequencing chemistry may result in severe bias when data sets from different sequencing platforms are combined. During the generation of sequencing data for this thesis, we experienced such bias that resulted from a change in sequencing chemistry between different sequencing platforms. From our own experience, we discuss the problem of technological advances for the build-up of large sequencing data sets generated on different platforms and develop ideas to correct such bias in the data.

References

Aitken, S. N., & Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. In D. J. Futuyma (Ed.), *Annual Review of Ecology, Evolution, and Systematics,* (Vol. 44, 367-388). Palo Alto: Annual Reviews.

Alexander, T. J., Vonlanthen, P., & Seehausen, O. (2017). Does eutrophication-driven evolution change aquatic ecosystems? *Philosophical Transactions of the Royal Society B-Biological Sciences, 372*(1712), 1-10. doi:10.1098/rstb.2016.0041

Bolnick, D. I., & Near, T. J. (2005). Tempo of hybrid inviability in centrarchid fishes (Teleostei : Centrarchidae). *Evolution, 59*(8), 1754-1767. doi:10.1111/j.0014-3820.2005.tb01824.x

Cahill, S., Chandola, T., & Hager, R. (2022). Genetic variants associated with resilience in human and animal studies. *Frontiers in Psychiatry, 13*(840120), 1-29. doi:10.3389/fpsyt.2022.840120

Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., . . . Heliconius Genome, C. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature, 487*(7405), 94-98. doi:10.1038/nature11041

De-Kayne, R., Selz, O. M., Marques, D. A., Frei, D., Seehausen, O., & Feulner, P. G. D. (2022). Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nature Communications, 13*(4479), 1-13. doi:10.1038/s41467-022-32181-8

Deufel, J., Löffler, H., & Wagner, B. (1986). Auswirkungen der Eutrophierung und anderer anthropogener Einflüsse auf die Laichplätze einiger Bodensee-Fischarten. *Österreichs Fischerei, 39*, 325-336.

Doenz, C. J., Bittner, D., Vonlanthen, P., Wagner, C. E., & Seehausen, O. (2018). Rapid buildup of sympatric species diversity in Alpine whitefish. *Ecology and Evolution, 8*(18), 9398-9412. doi:10.1002/ece3.4375

Eby, L. A., Crowder, L. B., McClellan, C. M., Peterson, C. H., & Powers, M. J. (2005). Habitat degradation from intermittent hypoxia: impacts on demersal fishes. *Marine Ecology Progress Series, 291*, 249-261. doi:10.3354/meps291249

Feller, A. F., Selz, O. M., McGee, M. D., Meier, J. I., Mwaiko, S., & Seehausen, O. (2020). Rapid generation of ecologically relevant behavioral novelty in experimental cichlid hybrids. *Ecology and Evolution, 10*(14), 7445-7462. doi:10.1002/ece3.6471

Feulner, P. G. D., & Seehausen, O. (2019). Genomic insights into the vulnerability of sympatric whitefish species flocks. *Molecular Ecology, 28*(3), 615-629. doi:10.1111/mec.14977

Ghosh, S. M., & Joshi, A. (2012). Evolution of reproductive isolation as a by-product of divergent life-history evolution in laboratory populations of Drosophila melanogaster. *Ecology and Evolution, 2*(12), 3214-3226. doi:10.1002/ece3.413

Gilman, R. T., & Behm, J. E. (2011). Hybridization, species collapse, and species reemergence after disturbance to premating mechanisms of reproductive isolation. *Evolution, 65*(9), 2592-2605. doi:10.1111/j.1558-5646.2011.01320.x

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking barriers: Causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution, 33*(3), 198-212. doi:10.1016/j.tree.2017.12.008

Grant, P. R., & Grant, B. R. (2019). Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America, 116*(46), 23216-23224. doi:10.1073/pnas.1913534116

Grimaldi, E., & Numann, W. (1972). Future of salmonid communities in European subalpine lakes. *Journal of the Fisheries Research Board of Canada, 29*(6), 931-936. doi:10.1139/f72-135

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics, 4*, 1-23. doi: 10.1146/annurev.es.04.110173.000245

Hudson, A. G., Lundsgaard-Hansen, B., Lucek, K., Vonlanthen, P., & Seehausen, O. (2016). Managing cryptic biodiversity: Fine-scale intralacustrine speciation along a benthic gradient in Alpine whitefish (*Coregonus* spp.). *Evolutionary Applications*, *10*, 251-266. doi:10.1111/eva.12446

Hudson, A. G., Vonlanthen, P., Bezault, E., & Seehausen, O. (2013). Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish. *Bmc Evolutionary Biology, 13*(108), 1-17. doi:10.1186/1471-2148-13-108

Hudson, A. G., Vonlanthen, P., & Seehausen, O. (2011). Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society B-Biological Sciences, 278*(1702), 58-66. doi:10.1098/rspb.2010.0925

IPBES (2019): Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (2019). (E. S. Brondizio, J. Settele, S. Díaz, & H. T. Ngo Eds.). Bonn, Germany: IPBES secretariat.

Jablonski, D. (2004). Extinction: past and present. *Nature, 427*(6975), 589-589. doi:10.1038/427589a

Jacobs, A., Carruthers, M., Eckmann, R., Yohannes, E., Adams, C. E., Behrmann-Godel, J., & Elmer, K. R. (2019). Rapid niche expansion by selection on functional genomic variation after ecosystem recovery. *Nature Ecology & Evolution, 3*, 77-86. doi:10.1038/s41559-018-0742-9

Jensen, E. L., & Leigh, D. M. (2022). Using temporal genomics to understand contemporary climate change responses in wildlife. *Ecology and Evolution, 12*(9), 1-8. doi:https://doi.org/10.1002/ece3.9340

Kagawa, K., & Takimoto, G. (2018). Hybridization can promote adaptive radiation by means of transgressive segregation. *Ecology Letters, 21*(2), 264-274. doi:10.1111/ele.12891

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2018). Rapid hybrid speciation in Darwin's finches. *Science, 359*(6372), 224-227. doi:10.1126/science.aao4593

Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science, 341*(6143), 292-295. doi:10.1126/science.1232392

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution, 34*(6), 531-544. doi:10.1016/j.tree.2019.02.008

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications, 8*(14363), 1-11. doi:10.1038/ncomms14363

Meier, J. I., Stelkens, R. B., Joyce, D. A., Mwaiko, S., Phiri, N., Schliewen, U. K., . . . Seehausen, O. (2019). The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nature Communications, 10*(5391), 1-11. doi:10.1038/s41467-019-13278-z

Mendelson, T. C., Imhoff, V. E., & Venditti, J. J. (2007). The accumulation of reproductive barriers during speciation: Postmating barriers in two behaviorally isolated species of darters (Percidae :Etheostoma). *Evolution, 61*(11), 2596-2606. doi:10.1111/j.1558-5646.2007.00220.x

Ostbye, K., Bernatchez, L., Naesje, T. F., Himberg, K. J. M., & Hindar, K. (2005). Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Molecular Ecology, 14*(14), 4371-4387. doi:10.1111/j.1365-294X.2005.02737.x

Powers, S. P., Peterson, C. H., Christian, R. R., Sullivan, E., Powers, M. J., Bishop, M. J., & Buzzelli, C. P. (2005). Effects of eutrophication on bottom habitat and prey resources of demersal fishes. *Marine Ecology Progress Series, 302*, 233-243. doi:10.3354/meps302233

Praebel, K., Knudsen, R., Siwertsson, A., Karhunen, M., Kahilainen, K. K., Ovaskainen, O., . . . Amundsen, P. A. (2013). Ecological speciation in postglacial European whitefish: rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and Evolution, 3*(15), 4970-4986. doi:10.1002/ece3.867

Reilly, P. F., Tjahjadi, A., Miller, S. L., Akey, J. M., & Tucci, S. (2022). The contribution of Neanderthal introgression to modern human traits. *Current Biology, 32*(18), 970-983. doi:doi.org/10.1016/j.cub.2022.08.027

Schluter, D. (1996). Ecological speciation in postglacial fishes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 351*(1341), 807-814. doi:10.1098/rstb.1996.0075

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford: Oxford University Press.

Scholl, J. P., & Wiens, J. J. (2016). Diversification rates and species richness across the tree of life. *Proceedings of the Royal Society B-Biological Sciences, 283*(20161334), 1-9. doi:10.1098/rspb.2016.1334

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution, 19*(4), 198-207. doi:10.1016/j.tree.2004.01.003

Seehausen, O. (2006). Conservation: Losing biodiversity by reverse speciation. *Current Biology, 16*(9), 334-337. doi:10.1016/j.cub.2006.03.080

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., . . . Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics, 15*(3), 176-192. doi:10.1038/nrg3644

Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology, 17*(1), 30-44. doi:10.1111/j.1365-294X.2007.03529.x

Selz, O. M., Doenz, C. J., Vonlanthen, P., & Seehausen, O. (2020). 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

Steinmann, P. (1950). Monographie der schweizerischen Koregonen. Beitrag zum Problem der Entstehung neuer Arten. Spezieller Teil. *Schweizerische Zeitschrift für Hydrologie, 12*(1), 340-491. doi:10.1007/BF02486030

Stelkens, R. B., Brockhurst, M. A., Hurst, G. D. D., & Greig, D. (2014). Hybridization facilitates evolutionary rescue. *Evolutionary Applications, 7*(10), 1209-1217. doi:10.1111/eva.12214

Stelkens, R. B., Schmid, C., & Seehausen, O. (2015). Hybrid breakdown in cichlid fish. *Plos One, 10*(5), 1-11. doi:10.1371/journal.pone.0127207

Stelkens, R. B., Young, K. A., & Seehausen, O. (2010). The accumulation of reproductive incompatibilites in African cichlid fish. *Evolution, 64*(3), 617-632. doi:10.1111/j.1558-5646.2009.00849.x

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology, 15*(2), 343-355. doi:10.1111/j.1365-294X.2005.02794.x

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hubner, S., . . . Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications, 9*(7), 892-908. doi:10.1111/eva.12367

Vamosi, J. C., Magallon, S., Mayrose, I., Otto, S. P., & Sauquet, H. (2018). Macroevolutionary patterns of flowering plant speciation and extinction. *Annual Review of Plant Biology, 69*, 685-706. doi:10.1146/annurev-arplant-042817-040348

Vedder, D., Lens, L., Martin, C. A., Pellikka, P., Adhikari, H., Heiskanen, J., . . . Cabral, J. S. (2022). Hybridization may aid evolutionary rescue of an endangered East African passerine. *Evolutionary Applications, 15*(7), 1177-1188. doi:10.1111/eva.13440

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Muller, R., Lundsgaard-Hansen, B., . . . Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature, 482*(7385), 357-362. doi:10.1038/nature10824

Vonlanthen, P., Roy, D., Hudson, A. G., Largiader, C. R., Bittner, D., & Seehausen, O. (2009). Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *Journal of Evolutionary Biology, 22*(3), 498-514. doi:10.1111/j.1420-9101.2008.01670.x

Wagner, C. E., Harmon, L. J., & Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. *Nature, 487*(7407), 366-369. doi:10.1038/nature11144

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution, 65*(7), 1897-1911. doi:10.1111/j.1558-5646.2011.01269.x

Chapter I

I. Alpine whitefish radiation retains adaptive genomic variation from extinct species 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

Published in:
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, 6*(4), 461-468. doi:10.1038/s41559-022-01665-7

See Appendix for accepted manuscript version.

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 scale samples, we generated whole-genome re-sequencing data of ten individuals of an extinct Lake Constance whitefish species. In comparison with sequencing data of the three sympatric whitefish species, we show that despite the extinction of this taxon, substantial fractions of its genome that have been 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 (Vamosi, Magallon, Mayrose, Otto, & Sauquet, 2018). Extinction can result from demographic decline, from loss of reproductive isolation or from a combination of both (Rhymer & Simberloff, 1996; Vonlanthen et al., 2012). When extinction involves the loss of reproductive isolation (Seehausen, 2006; Seehausen, Takimoto, Roy, & Jokela, 2008), the extinction process can leave a lasting legacy in the genomes of surviving species through introgressive hybridization (Kearns et al., 2018; Rhymer & Simberloff, 1996), potentially even influencing species that will only emerge in the future (Meier et al., 2017). 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 (Barlow et al., 2018; Green et al., 2010; Kuhlwilm, Han, Sousa, Excoffier, & Marques-Bonet, 2019; Palkopoulou et al., 2018). However, apart from these few examples, genomic information for extinct species is still rare (Ottenburghs, 2020). 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 (Rundle & Nosil, 2005; Schluter, 2000), is an important process in the evolution of a substantial proportion of contemporary eukaryotic species diversity (Nosil, 2012; Schluter, 2009). 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 (Ghosh & Joshi, 2012; Rundle & Nosil, 2005; Yeaman & Whitlock, 2011). Ecologically-mediated reproductive isolation, both pre- and postzygotic, results from performance trade-offs between, and adaptation to, alternative fitness optima (Rundle & Nosil, 2005; Schluter, 2000). 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 (Seehausen, van Alphen, & Witte, 1997; Taylor et al., 2006), a process called speciation reversal, potentially resulting in the sudden and rapid extinction of species through introgressive hybridization.

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 (Seehausen et al., 2008). Whilst the potentially widespread impacts of speciation reversal on contemporary biodiversity loss are still underappreciated in conservation (Seehausen, 2006), its genomic consequences are still underappreciated in evolutionary biology. Genetically admixed hybrid populations that emerged from speciation reversal might have enhanced evolvability (Grant & Grant, 2019). In the future, such populations may adapt in new and unexpected ways (Feller et al., 2020; Kagawa & Takimoto, 2018), expand their ranges (Pfennig, Kelly, & Pierce, 2016), and even seed further species diversification (Lamichhaney et al., 2016). 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 outstanding system in which to study ecological speciation and the consequences of its reversal (Hudson, Vonlanthen, Bezault, & Seehausen, 2013; Hudson, Vonlanthen, & Seehausen, 2011; Jacobs et al., 2019). Across the large pre-Alpine lakes of Switzerland more than 30 endemic whitefish species have evolved since the end of the last glacial maximum (Doenz, Bittner, Vonlanthen, Wagner, & Seehausen, 2018; Hudson et al., 2011; Selz, Doenz, Vonlanthen, & Seehausen, 2020; Steinmann, 1950). 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 (Hudson, Lundsgaard-Hansen, Lucek, Vonlanthen, & Seehausen, 2016). Therefore, Alpine whitefish species are highly sensitive to speciation reversal when habitat diversity and suitability along the lacustrine water depth gradient changes (Feulner & Seehausen, 2019; Hudson et al., 2016; Vonlanthen et al., 2009). Anthropogenic eutrophication during the 20th century led to the loss of deep-water spawning habitats, reducing prezygotic isolation between sympatric whitefish species (Vonlanthen et al., 2012). 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 (Vonlanthen et al., 2012). 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 (Feulner & Seehausen, 2019; Hudson et al., 2013; Vonlanthen et al., 2012).

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. 2). 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 (Vonlanthen et al., 2012). Additionally, five private microsatellite alleles of the extinct species were discovered in all extant species after eutrophication (Vonlanthen et al., 2012) and whole-genome resequencing data indicated significanted introgression from the extinct into all three extant species (Frei, 2018), consistent with eutrophication-induced speciation reversal. Using eleven historical samples of the extinct *C. gutturosus*, 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. We here demonstrate that introgression from the extinct into all extant species included genomic variation with signatures of positive selection that shaped the genome of the extinct species before eutrophication, indicating that these regions were potentially adaptive in the extinct species prior to speciation reversal.

Methods

Sample collection and DNA extraction

Historical whitefish scale samples, assembled by David Bittner (see Vonlanthen et al. (2012) for details) and collected before the onset of eutrophication in the upper basin of Lake (1937 and 1948), were used to extract DNA from nine *C. gutturosus* individuals. 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 (Wasko, Martins, Oliveira, & Foresti, 2003) 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 whitefish scale sample, an Illumina paired-end TruSeq DNA Nano library (Illumina GmbH, CH) was produced. 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 were sequenced 2x150 paired-end on a Novaseq 6000 sequencing platform.

Mapping and filtering of sequencing reads

Poly-G strings at the end of the reads were removed using fastp 0.20.0 (Chen, Zhou, Chen, & Gu, 2018). Overlapping paired end reads with total length longer than 25 bp were merged using SeqPrep 1.0 (https://github.com/jstjohn/SeqPrep). Raw reads were aligned to the Alpine whitefish genome assembly (R. De-Kayne, Zoller, & Feulner, 2020) (ENA accession: GCA_902810595.1) with bwa mem version 0.7.12 (Li & Durbin, 2009) 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

To assess the introgression of potentially adaptive genomic variation from the extinct *C. gutturosus* into the three Lake Constance whitefish species during the anthropogenic eutrophication period, we made use of existing re-sequencing data of historical populations from Frei (2022) (*C. arenicolus* (n=3), *C. gutturosus* (n=3; two of these three individuals were

again sequenced to increase sequencing coverage), *C. macrophthalmus* (n=2) and *C. wartmanni* (n=2)). We further combined this data with existing re-sequencing data of the contemporary whitefish population sampled 2015 from Frei (2018) (*C. arenicolus* (n=2), *C. macrophthalmus* (n=2) and *C. wartmanni* (n=2)) and from De-Kayne (2020) (*C. arenicolus* (n=3), *C. macrophthalmus* (n=1) and *C. wartmanni* (n=4)). A complete list of all used samples is in Supplementary Table 1. A *Salmo salar* individual (short read archive accession number: SSR3669756) from Kjaerner-Semb et al. (2016) served as outgroup,

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 (De-Kayne et al., 2020). Instead of hard genotyping, we calculated genotype likelihoods (Li, 2011) and minor allele frequencies (Kim et al., 2011; Skotte, Korneliussen, & Albrechtsen, 2012) at polymorphic sites applying the samtools genotype likelihood model implemented in angsd version 0.925 (Korneliussen, Albrechtsen, & Nielsen, 2014). 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 (Kim et al., 2011) 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 (Korneliussen et al., 2014): -sb_pval 0.05 -qscore_pval 0.05 -edge_pval 0.05 -mapq_pval 0.05, resulting in a total of 477’981 sites.

To visualize general relationships among the four studied species, we produced a maximum likelihood phylogeny using RAxML version 8.2.12 (Stamatakis, 2014). We first

calculated genotype likelihoods of the *S. salar* outgroup at all 477'981 polymorphic sites with angsd 0.925 (Korneliussen et al., 2014), and then inferred genotypes of all individuals including the outgroup and phased these using beagle 4.1 (Browning & Browning, 2007). We then thinned this dataset using VCFtools 0.1.16 (Danecek et al., 2011) 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 version 8.2.12 (Stamatakis, 2014) 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) (Martin & Van Belleghem, 2017). First, we calculated genotype likelihoods in angsd 0.925 (Korneliussen et al., 2014), 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 *S. salar* outgroup individual at the positions identified to be polymorphic in our dataset. We then inferred genotypes from the likelihoods and phased these genotypes with beagle 4.1 (Browning & Browning, 2007), 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 (Bukowicki, Franssen, & Schlotterer, 2016) that are relevant for our TWISST approach, and TWISST has been reported to be robust to within-taxon phasing errors (Marburger et al., 2019). We assessed

coverage of each sample at these polymorphic sites with angsd 0.925 (Korneliussen et al., 2014), 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 version 3.0 (Guindon et al., 2010) and the script phyml_sliding_windows.py (https://github.com/simonhmartin/genomics_general/blob/master/phylo). TWISST (Martin & Van Belleghem, 2017) was performed separately for each post-eutrophication sample, using a four taxon topology in the ordering as shown in (see Fig. 1).

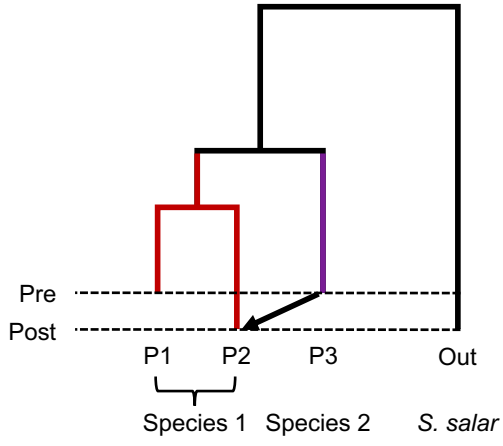


Figure 1: Schematical representation of the population ordering for the TWISST analysis. P1 consisted of all pre-eutrophication samples of an extant species, P2 was the focal text individual (post-eutrophication individual of the same species as P1), P3 were the eleven *C. gutturosus* individuals and the outgroup used was *S. salar*.

All available pre-eutrophication samples of one extant species were in P1, the potential recipient population P2 consisted of one focal individual, all eleven available *C. gutturosus* samples were in (P3) and *S. salar* served 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. 1). Following Meier et al. (2018), 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 performed a two-sided t-test in R (R Core Team, 2018) to evaluate whether the sharing of windows introgressed from *C. gutturosus* was significantly higher between conspecific individuals compared to heterospecifics.

We performed a selection scan using the statistic nSL (Ferrer-Admetlla, Liang, Korneliussen, & Nielsen, 2014). 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 data set of genotype likelihoods obtained from angsd 0.925 (Korneliussen et al., 2014) to only *C. gutturosus* individuals, and then inferred genotypes and phased these using beagle 4.1 (Browning & Browning, 2007). We then calculated the unstandardized nSL statistic with the software selscan 1.3.0 (Szpiech & Hernandez, 2014). Because the sample size consisted of 11 individuals, we included low frequency variants. We then used norm 1.3.0 (Szpiech & Hernandez, 2014) to normalize the unstandardized nSL calculations with default parameters in 50 kb windows along the genome. We considered windows with more than 51.1% of variable sites (top 1 percentile) with a normalized nSL score above 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 (Korneliussen, Moltke, Albrechtsen, & Nielsen, 2013) in angsd 0.925 (Korneliussen et al., 2014) based on 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 0.925 (Korneliussen et al., 2014) and then calculated the maximum likelihood estimate of the folded site allele frequency spectrum using realSFS of angsd 0.925 (Korneliussen et al., 2014). We used the global site allele frequency spectrum to calculate theta per site in realSFS of angsd 0.925 (Korneliussen et al., 2014), and then calculated Tajima's D in 50 kb windows using thetaStat of angsd 0.925 (Korneliussen et al., 2014). 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 (Fig. 3). 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' ($p<0.01$; W= 8352543) (R Core Team, 2018). 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 (De-Kayne et al., 2020); ENA accession: GCA_902810595.1) that overlap in their position with the identified windows under selection were identified using bedtools v.2.28.0 (Quinlan, 2014). 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 2.38.1 (Alexa & Rahnenfuhrer, 2020) separately for each of the three ontology classes cellular component (CC), biological processes (BP), and molecular function (MF). 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 'wilcox.test' (R Core

Team, 2018) and showed that there was no significant difference between the two sets of windows.

Results

Sequencing coverage at polymorphic sites of the 10 sequenced *C. gutturosus* samples (of which two samples were combined with existing sequencing data to increase coverage) ranged between 3.4x and 14.2x (see Supplementary Table 1). When combining the 11 newly sequenced *C. gutturosus* samples with existing sequencing data into a maximum likelihood RAxML tree, all species clustered monophyletically with high bootstrap support (Fig. 2). As expected, branch lengths of historical samples were often longer than those of the contemporary samples of the same species.

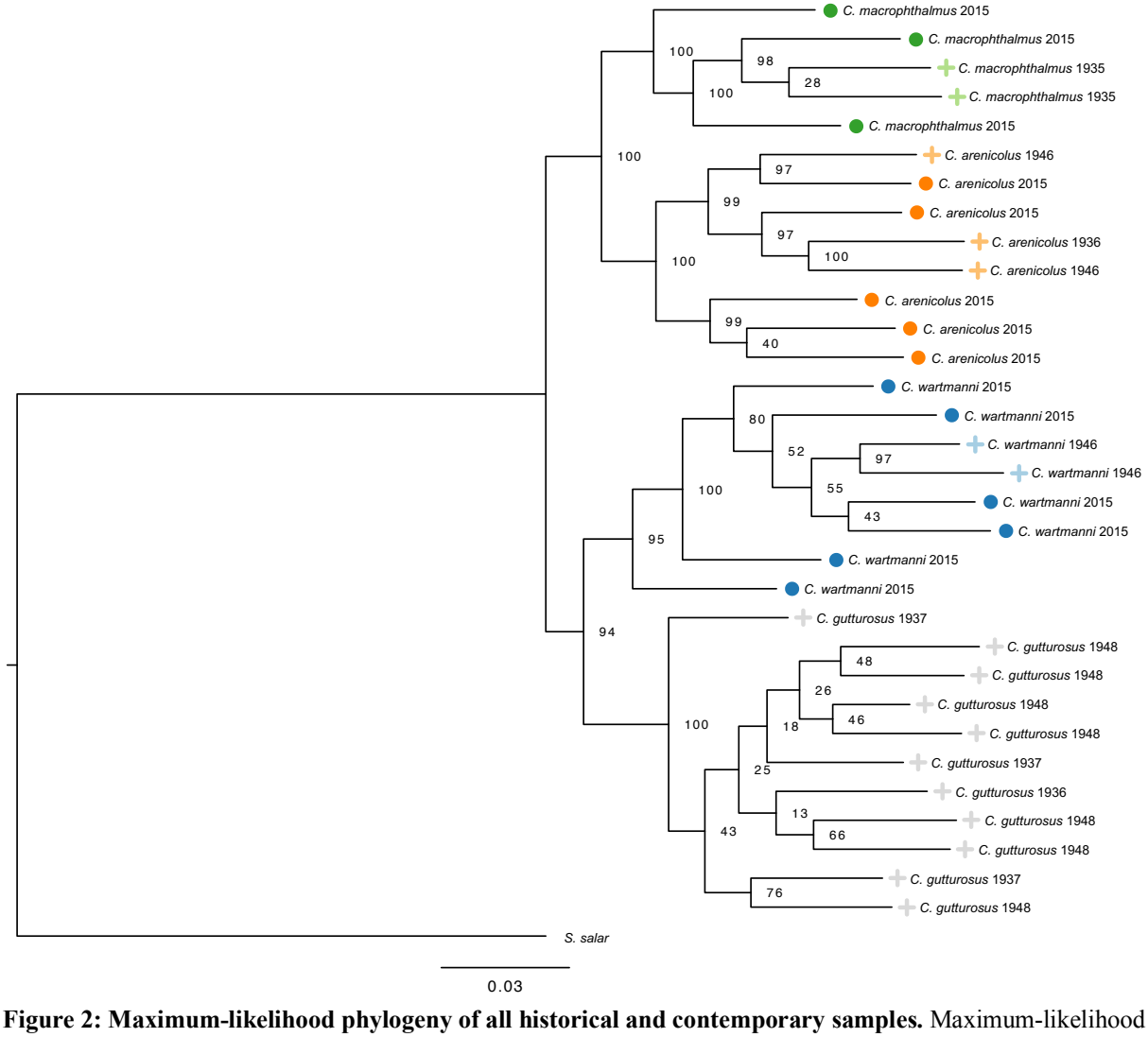


Figure 2: Maximum-likelihood phylogeny of all historical and contemporary samples. Maximum-likelihood phylogeny of all pre- (crosses) and post-eutrophication (points) individuals of the four Lake Constance whitefish species based on 58'831 SNPs. Colours correspond to species (see Fig. 1). Support values 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).

We performed a selection scan using the haplotype based statistic nSL (Ferrer-Admetlla et al., 2014) to determine whether genomic regions that have introgressed from *C. gutturosus* into the three extant Lake Constance whitefish species during speciation reversal have been under positive selection in the now extinct profundal *C. gutturosus* before the anthropogenic eutrophication period started. We considered the highest 1% fraction (315 50 kb windows) of regions showing signals of positive selection in our haplotype-based nSL selection scan (Tajima's D based on genotype likelihoods (Korneliussen et al., 2013) in this top 1% of windows is significantly different from the rest of the genome; Fig. 3) as potentially having conferred adaptation to profundal habitats in *C. gutturosus* (see Supplementary Table 2 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).

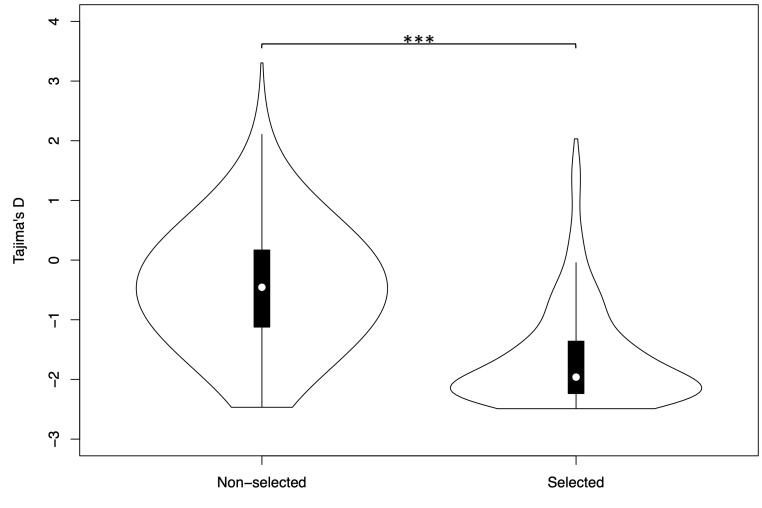


Figure 3: Tajima's D based on genotype likelihoods for windows identified to have been under selection in *C. gutturosus* using nSL. Violin plots of Tajima's D in *C. gutturosus* (n=11) calculated in 50 kb windows comparing the 315 windows identified to be in the top 1 percentile of the nSL analysis 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.

We used topology weighting by iterative sampling of subtrees (Martin & Van Belleghem, 2017) to explore evolutionary relationships in 50 kb windows along the genome to find regions where a topology consistent with introgression from *C. gutturosus* into one of the extant species 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. 4). Windows showing an introgression signature were more frequently shared between individuals of the same species (Fig. 4) than between individuals of different species ($t=57.18$; $p<0.01$; $df=29.34$). Of all windows with evidence for positive selection in *C. gutturosus* in our nSL selection scan, 53.3% have introgressed from *C. gutturosus* into extant whitefish species (Fig. 3). 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). We observed no difference in gene density between introgressed and non-introgressed regions (Fig. 5).

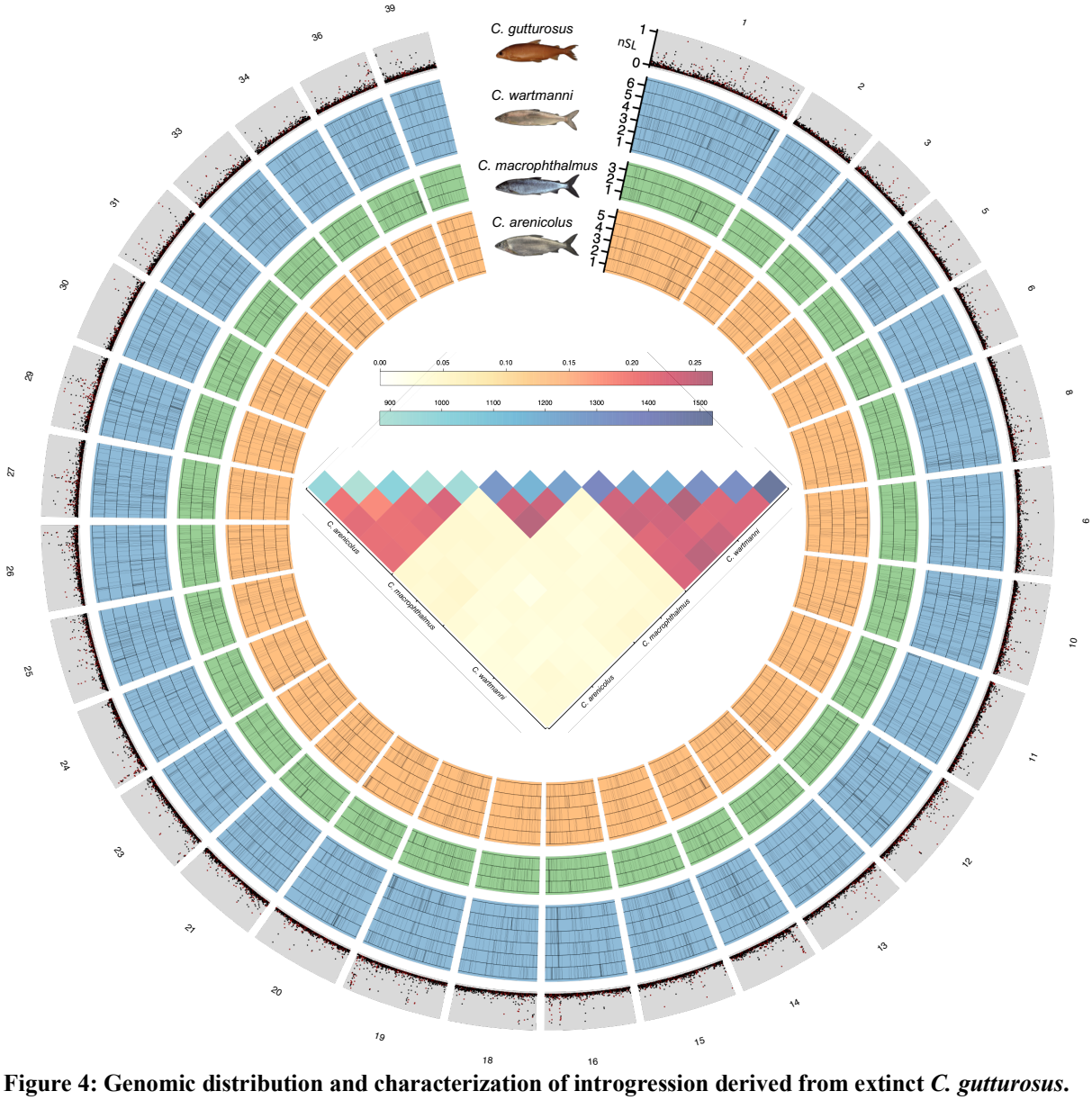


Figure 4: Genomic distribution and characterization of introgression derived from extinct *C. gutturosus*. Each of the three inner tracks corresponds to a species (blue *C. wartmanni*, green *C. macrophthalmus* and orange *C. arenicolus*) and each track is subdivided into individual genomes. Each black 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).

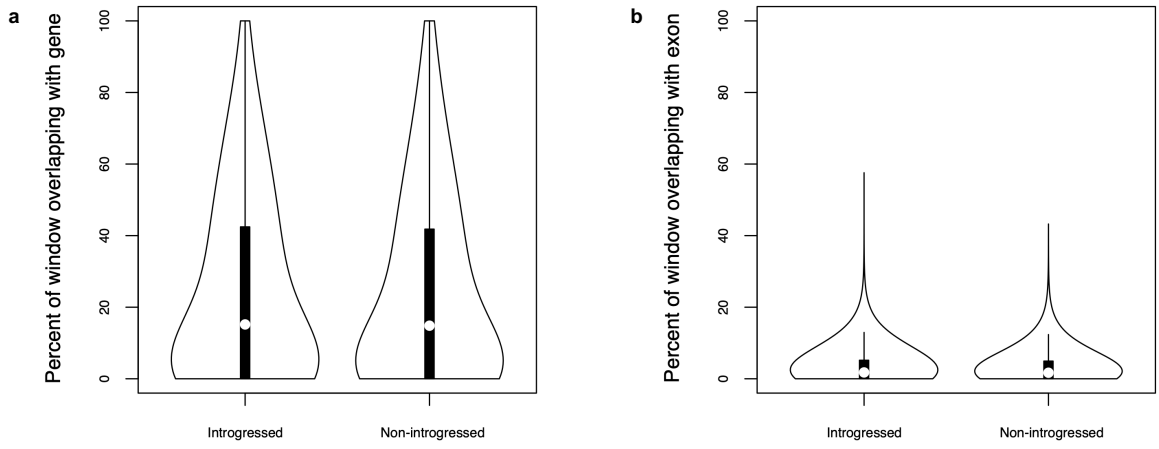


Figure 5: Comparison of gene density in introgressed and non-introgressed windows. a) Comparison of gene density between windows identified to be introgressed and those that did not show evidence for introgression (non-introgressed) from *C. gutturosus* (n=11) across all three extant species (n=14). 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 interquantile range, white dots correspond to medians and spikes are extending to the upper and lower adjacent values.

Discussion

Since species diversity can evolve in response to heterogenous environments, the homogenization of environments can drive species extinction (Seehausen et al., 1997). 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 (Seehausen et al., 1997; Vonlanthen et al., 2012). 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 (Palkopoulou et al., 2018), apes (Kuhlwilm et al., 2019) and bears (Barlow et al., 2018), 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. 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 (Hudson et al., 2013; Seehausen, 2006; Vonlanthen et al., 2012) 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 evolutionary potential to recolonize the lost niche after ecosystem restoration. 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 prior to 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 (Doenz & Seehausen, 2020). Nonetheless, profundal regions remain devoid of whitefish (Alexander & Seehausen, 2021). 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 of speciation (Gilman & Behm, 2011). However, re-emergence appears less likely the more species that 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 years or ~6 whitefish generations (Nussle, Bornand, & Wedekind, 2009)) 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 (Grabenstein & Taylor, 2018), 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 novel adaptive combinations of genotypes can accelerate adaptation and speciation (Seehausen, 2004). Therefore, speciation reversal could potentially facilitate adaptation and diversification in response to changing or even entirely novel 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.

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. 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.

Data availability

The raw sequencing files are accessible on SRA (PRJEB43605). The *S. salar* outgroup sample 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. Scripts used for data

analysis are available on GitHub (https://github.com/freidavid/Genomic-Consequences-of-Speciation-Reversal).

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 for the species assignment. 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.

References

Alexa, A. & Rahnenfuhrer, J. *TopGO: Enrichment Analysis for Gene Ontology*. R package version 2.42.0. doi:10.18129/B9.bioc.topGO (2020).

Alexander, T. & Seehausen, O. *Diversity, distribution and community composition of fish in perialpine lakes - "Projet Lac" synthesis report.* (Bundesamt für Umwelt BAFU, 2021).

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).

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).

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).

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).

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

De-Kayne, R. *The genetic basis of adaptation and speciation in the Swiss Alpine whitefish radiation*. PhD thesis. (University of Bern, 2020).

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).

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).

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

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).

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).

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).

Frei, D. *Whole-genome sequencing of an extinct whitefish species reveals extensive introgression after speciation reversal*. Master thesis. (University of Bern, 2018).

Frei, D., De-Kayne, R., Selz, O. M., Seehausen, O., & Feulner, P. G. D. Genomic variation from an extinct species is retained in the extant radiation following speciation reversal. *Nat. Ecol. Evol.* **6**, 461-468. doi:10.1038/s41559-022-01665-7 (2022).

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).

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).

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).

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).

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

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).

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).

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).

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).

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).

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).

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

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).

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).

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).

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).

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).

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

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).

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

Marburger, S. *et al.* Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* **10**, doi:10.1038/s41467-019-13159-5 (2019).

Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429-438, doi:10.1534/genetics.116.194720 (2017).

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

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).

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

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).

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

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).

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).

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).

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

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).

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

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

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

Seehausen, O. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198-207, doi:10.1016/j.tree.2004.01.003 (2004).

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

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).

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).

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).

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).

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).

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

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).

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).

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).

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

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).

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).

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).

Supplementary Table 1: 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.

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

Supplementary Table 2: Functional enrichment of windows under selection in *C. gutturosus.* GO enrichment analysis for all windows that were under positive selection in *C. gutturosus* (n=11) before its extinction, respectively before the eutrophic phase of the lake started. The first column shows the unique GO identifier for each enriched term, the second column gives the respective terminological description, the third column gives the number of genes annotated with the term within the genome, the fourth column gives the number how often the term was represented within windows with a signature of positive selection, the fifth column gives how often the term was expected in those windows by chance, the next two columns give the p-value using Fisher's exact method based on gene counts (accounting for the GO topology by weighting sixth column, or elimination seventh column) to test for a statistical overrepresentation of the term, and the last column gives one of the three ontologies of interest (CC cellular component, BP biological process, MF molecular function) that have been explored. For each category (CC, BP and MF), the top five entries are shown. Full output table is available as electronic supplementary material[1].

GO.ID	Term	Annotated	Significant	Expected	weight_fisher_P	elim_fisher_P	class
GO:0031094	platelet dense tubular network	11	3	0.07	4.9E-05	4.9E-05	CC
GO:0016528	sarcoplasm	122	5	0.83	0.0015	0.0015	CC
GO:0005747	mitochondrial respiratory chain complex I	43	3	0.29	0.0031	0.0031	CC
GO:0005952	cAMP-dependent protein kinase complex	17	2	0.12	0.0058	0.0058	CC
GO:0098588	bounding membrane of organelle	2294	28	15.55	0.0061	0.0213	CC
GO:0090330	regulation of platelet aggregation	27	3	0.19	0.0009	0.0412	BP
GO:0042311	vasodilation	30	3	0.21	0.0012	0.0012	BP
GO:0048210	Golgi vesicle fusion to target membrane	10	2	0.07	0.0022	0.0022	BP
GO:0060631	regulation of meiosis I	10	2	0.07	0.0022	0.0022	BP
GO:0035845	photoreceptor cell outer segment organization	37	3	0.26	0.0023	0.0023	BP
GO:0015278	calcium-release channel activity	26	4	0.18	3.2E-05	3.2E-05	MF
GO:0031681	G-protein beta-subunit binding	35	3	0.25	0.0019	0.0019	MF
GO:0004692	cGMP-dependent protein kinase activity	10	2	0.07	0.0022	0.0022	MF
GO:0099602	neurotransmitter receptor regulator activity	11	2	0.08	0.0026	0.0026	MF
GO:0031683	G-protein beta/gamma-subunit complex binding	86	4	0.61	0.0033	0.0033	MF

Chapter II

II. Introgression from extinct species facilitates adaptation to its vacated niche

David Frei[1,2], Pascal Reichlin[1], 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

Published as:
Frei, D., Reichlin, P., Seehausen, O., & Feulner, P. G. D. (2022). Introgression from extinct species facilitates adaptation to its vacated niche. *Molecular Ecology*, *32*(4), 841-853. doi:10.1111/mec.16791

Anthropogenic disturbances of ecosystems are causing a loss of biodiversity at an unprecedented rate. Species extinctions often leave ecological niches underutilized, and their colonization by other species may require new adaptation. In Lake Constance, an endemic profundal whitefish species went extinct during a period of anthropogenic eutrophication. In the process of extinction, the deep-water species hybridized with three surviving whitefish species of Lake Constance, resulting in introgression of genetic variation that is potentially adaptive in deep-water habitats. Here, we sampled a water depth gradient across a known spawning ground of one of these surviving species, *Coregonus macrophthalmus*, and caught spawning individuals in greater depths (down to 90m) than historically recorded. We sequenced a total of 96 whole genomes, 11-17 for each of six different spawning depth populations (4m, 12m, 20m, 40m, 60m, and 90m), to document genomic intraspecific differentiation along a water depth gradient. We identified 52 genomic regions that are potentially under divergent selection between the deepest (90m) and all shallower (4-60m) spawning habitats. At 12 (23.1%) of these 52 loci, the allele frequency pattern across historical and contemporary populations suggests that introgression from the extinct species potentially facilitates ongoing adaptation to deep water. Our results are consistent with the syngameon hypothesis, proposing that hybridization between members of an adaptive radiation can promote further niche expansion and diversification. Furthermore, our findings demonstrate that introgression from extinct into extant species can be a source of evolvability enabling rapid adaptation to environmental change and may contribute to the ecological recovery of ecosystem functions after extinctions.

Introduction

The recovery of ecosystems from anthropogenic disturbances is a central factor to predict future consequences of environmental change on biodiversity (Malhi et al., 2020). When species go extinct due to anthropogenic disturbance, previously occupied niche space may become vacant (Prada et al., 2016). Thus, in communities lacking ecological redundancy, extinction can provide surviving species with previously unavailable ecological opportunity (Prada et al., 2016; Wellborn & Langerhans, 2015). Whether and within which time scale such vacant niche space that had previously been occupied by newly extinct species can be filled up again through niche expansion of related or newly emerging species is one critical aspect of determining the functional recovery of an ecosystem on an evolutionary timescale.

Degradation of ecosystems can result in species loss by either demographic decline, or by speciation reversal through the merging of several related species into a single hybrid population (Rhymer & Simberloff, 1996; Taylor, 2006; Tedesco et al., 2016). When reproductive isolation between species is mediated by features of the environment that interact with intrinsic lineage traits, environmental change can induce the loss of reproductive isolation, resulting in speciation reversal through introgressive hybridization (Seehausen, 2006; Seehausen, Takimoto, Roy, & Jokela, 2008; Taylor et al., 2006, Tedesco et al., 2016). Even when ecosystems are restored and thus disturbance was merely transient, speciation reversal can result in the loss of species. However, parts of the genetic variation that had once defined the lost species will then often have been transferred to surviving species through hybridization and introgression (Barlow et al., 2018; Kuhlwilm, Han, Sousa, Excoffier, & Marques-Bonet, 2019; Reilly, Tjahjadi, Miller, Akey, & Tucci, 2022). Thereby, speciation reversal can result in the functional extinction of taxa within a few generations (Rhymer & Simberloff, 1996; Taylor, 2006; Todesco et al., 2016; Vonlanthen et al., 2012), while some of

their genetic variation may persist within extant species (Frei et al., 2022; Gilman & Behm, 2011).

Rapid adaptation and speciation are often associated with re-assembling of old genetic variation originating from hybridization (Marques, Meier, & Seehausen, 2019; Hamid, Korunes, Beleza, & Goldberg, 2021; Moran et al., 2021). Examples include Darwin's finches (Lamichhaney et al., 2015; Lamichhaney et al., 2016), cichlid fish (Irisarri et al., 2018; Meier et al., 2017), *Lycaeides* butterflies (Nice et al., 2013) or *Helianthus* sunflowers (Rieseberg et al., 2003). Hybridization is thought to promote ecological diversification and speciation because it can generate new trait combinations suitable for utilizing resources that could not be utilized before (Marques et al., 2019; Seehausen, 2004). Through these effects, hybridization can fuel entire adaptive radiations, both the onset (Meier et al., 2017) and the continuation of adaptive radiation beyond the first speciation events (Seehausen, 2004). In the context of extinction by speciation reversal, introgressive hybridization during speciation reversal might facilitate the adaptation of a surviving species to an extinct species' vacated niche. Admixture variation generated through hybridization during speciation reversal might be re-assembled into allelic combinations that are adaptive in the now unoccupied habitat previously used by the extinct species. When some of the alleles derived from the extinct species that evolved in response to selective pressures in its former habitat introgress into surviving species, they might facilitate adaptation to the now vacant habitat within surviving species. Such a scenario is in line with the syngameon hypothesis of adaptive radiations, which predicts that hybridization between members of an existing adaptive radiation might induce further adaptation and diversification (Seehausen, 2004).

The Alpine whitefish radiation provides an outstanding opportunity to study the consequences of speciation reversal induced by severe but transient environmental change. Reproductive isolation between sympatric Alpine whitefish species is maintained

predominantly by extrinsic (and possibly intrinsic) prezygotic and extrinsic postzygotic mechanisms (Vonlanthen et al., 2009; Woods et al., 2009; Ingram et al., 2012; Hudson et al., 2016). Many sympatric species differ in the water depth of spawning sites and show differential timing of spawning (Steinmann 1950). Some sympatric species overlap in both, yet retaining significant reproductive isolation, possibly due to behavioural mating preferences (Steinmann 1950; Hudson et al., 2016). Alpine whitefish species are highly sensitive to the alteration of the physiochemical habitat characteristics, because spawning niche differentiation and in turn reproductive isolation, is strongly dependent on the persistence of fine-scale depth-related differences in the specific lacustrine habitat (Vonlanthen et al., 2012; Hudson et al., 2016). Anthropogenic eutrophication during the last century weakened reproductive isolation between Alpine whitefish species, resulting in speciation reversal through introgressive hybridization (Frei et al., 2022; Vonlanthen et al., 2012).

In Lake Constance, a lake between the borders of Germany, Austria and Switzerland, four endemic whitefish species have been taxonomically described. The deep-water species *C. gutturosus* went extinct during eutrophication-induced speciation reversal (Vonlanthen et al., 2012). During the period of anthropogenic eutrophication, its deep-water spawning grounds were lost as a result of decreased oxygen concentrations (Nümann, 1972; Wahl & Löffler, 2009). The anoxic conditions at the water-sediment interface in deep benthic areas of the lake probably prevented successful reproduction of *C. gutturosus* and thereby contributed to its extinction (Deufel, Löffler, & Wagner 1986; Wahl & Löffler, 2009). Recent work demonstrated extensive introgression of this extinct species into all surviving members of the radiation (Frei et al., 2022). Introgression included potentially adaptive alleles that, before eutrophication, had been under positive selection in the extinct species (Frei et al., 2022). Today, oligotrophic conditions of the lake have been largely restored and deep-water habitats

are again accessible for fish. Yet, profundal habitats of Lake Constance are reported to be devoid of any whitefish (Alexander & Seehausen, 2021). However, the genetic variation that had evolved in the extinct deep-water species and introgressed into the extant species during eutrophication-induced speciation reversal may provide the surviving species with alleles that could be adaptive in deep water (respectively adaptive in a now vacant profundal habitat previously occupied by the recently extinct species). Thus, introgression from the extinct profundal *C. gutturosus* could in principle facilitate adaptation to deep water habitats in some of the surviving Lake Constance whitefish species, that had not occupied these greater depths previously.

Here, we sampled a depth transect on known spawning grounds of *C. macrophthalmus*, the deepest spawning of the extant species. Adaptation to the extinct species' former deep-water habitat seems most plausible in this species. We set nets in six depth zones ranging from 4m to 120m of water depth during spawning season. Thus, we sampled the entire historically known Lake Constance whitefish depth range from that of the shallowest spawning *C. macrophthalmus* populations down to the depths (90-120m) where the extinct *C. gutturosus* used to spawn. We then sequenced whole-genomes of 11 to 17 individuals per depth (total n=96) to search for signatures of differentiation and adaptation along the water depth gradient. We demonstrate that the deepest caught *C. macrophthalmus* individuals from 90m depth show morphological and genetic differentiation from the shallower caught individuals, and we identify 52 candidate loci that might be under positive selection in deep water. At twelve of these loci (23.1%), the allele frequency pattern across our six different spawning depth populations of *C. macrophthalmus* together with the allele frequencies in the historical populations of all Lake Constance whitefish species sampled before speciation reversal (from Frei et al. (2022)) suggest that these alleles might have introgressed from the extinct *C. gutturosus* during the period of anthropogenic lake

eutrophication. Thus, our results demonstrate that some alleles that are likely to have introgressed from the extinct species are potentially involved in adaptation of some populations of *C. macrophthalmus* to the deep-water environment historically used as habitat by the now extinct *C. gutturosus*. This suggests that introgressive hybridization during speciation reversal potentially facilitates adaptation within surviving species to the vacated habitat of a recently extinct species.

Materials and Methods

Study system

In Lake Constance, a large pre-alpine lake bordering Germany, Austria and Switzerland, four whitefish species have been taxonomically described (Steinmann, 1950). *C. wartmanni* is most relevant for commercial fisheries and is extensively managed. It is a pelagic species, mostly feeding on planktonic food resources in the open water (Steinmann, 1950). *C. wartmanni* spawns pelagically, close to the surface over deep water (70-250m) late November until early December (Nenning, 1834; Nüsslin, 1907; Schweizer, 1894; Schweizer, 1926; von Rapp, 1858). *C. macrophthalmus* is the species that is of second largest relevance for commercial fisheries. *C. macrophthalmus* is feeding on both pelagic and benthic food resources (Steinmann, 1950). The species has historically been described to spawn on relatively shallow benthic spawning grounds at depths of less than 20m, close to the shore of the lake, starting from mid-November until early January (Nenning, 1834; Nüsslin, 1907; Eckmann & Rösch, 1998; Schweizer, 1894; Schweizer, 1926; Steinmann, 1950). More recent work suggested an extended range of spawning depth of *C. macrophthalmus* between 2 and 50 meter after the lake has returned to an oligotrophic state (Hirsch, Eckmann, Oppelt, & Behrmann-Godel, 2013; Jacobs et al., 2019). *C. arenicolus* is a relatively large bodied species, feeding on large benthic macroinvertebrates (Steinmann, 1950). It has a very short spawning period mid-November and it has been historically described to spawn on very shallow spawning grounds of 1-2 meters depth and mainly on sandy substrate (Nenning, 1834; Schweizer, 1894; Steinmann, 1950; von Rapp, 1858). *C. gutturosus,* went extinct during the period of anthropogenic eutrophication during the 1970's or 1980's (Vonlanthen et al., 2012). *C. gutturosus* was a deep-water specialist, feeding on benthic macroinvertebrates in the profundal regions of the lake (Steinmann 1950). *C. gutturosus* had an extended spawning

period ranging from summer until winter (Steinmann, 1950), during which it spawned benthically in depths around 70-80m or more (Steinmann, 1950; von Rapp, 1858; von Siebold, 1858).

Sampling

Nets were set at seven different water depths (4m, 12m, 20m, 40m, 60m, 90m and 120m) on a known *C. macrophthalmus* spawning ground at the beginning of the spawning season 2019 (November 26th – November 29th). We used benthic gillnets with varying mesh sizes, consisting of panels of 25mm, 35mm and 45mm mesh size to cover the known range of body sizes of spawning whitefish. We caught fish down to 90m, but not anymore in 120m, suggesting that we covered the whole range of depth that is currently used by whitefish for spawning. Individuals were anaesthetized and subsequently euthanized using appropriate concentrations of tricaine methane sulfonate solutions (MS-222) according to the permit issued by the canton of St.Gallen (SG31396). Fin-clips were taken and stored in 100% analytical ethanol until extraction of DNA. Individual specimens were weighed, total length was measured, a standardized picture was taken and a first species assignment was done on site. All the fish caught were fixed in 4% formalin solution for one month, and then transferred through a series of increasing ethanol concentrations (pure water, 30%, 50%) to the final concentration of 70% for long-term storage.

Morphometric analysis

On all fish caught, we measured 23 linear morphometric traits using digital calliper according to Selz et al. (Selz, Doenz, Vonlanthen, & Seehausen, 2020), except for taking the mean of three measurements per trait (instead of the mean of two measurements). The traits

measured were BD (body depth), DHL (dorsal head length), PreD (predorsal length), PostD (postdorsal length), CD (caudal peduncle depth), CL (caudal peduncle length), SL (standard length), HL (head length), HD (head depth), HW (head width), PostO (postorbital length), SN (snout length), ED (eye diameter), EH (eye height), SD (snout depth), SW (snout width), M (length of maxilla), MW (mouth width), UJ (upper jaw length), LJ (lower jaw length), LJW (lower jaw width), IOW (interorbital width), INW (internarial width) (see Table 1 in Selz et al. (2020)). Additionally, we counted the number of gill-rakers (GRC) also according to Selz et al. (Selz et al., 2020). We used individuals that were assigned to *C. macrophthalmus* (n=106) for the following morphological analyses. First, we size corrected our 23 linear morphometric traits by using the residuals of the linear regression of standard length with the specific trait for further analysis. To assess morphological differentiation between fish caught at different depths, we performed a partial least squares regression analysis between all size-corrected traits (excluding standard length and gill-raker count) and depth (4m, 12m, 20m, 40m, 60m, 90m) in R (R Core Team, 2018) using the package "pls" (Mevik & Wehrens, 2007). Finally, we tested whether the first component was significantly correlated with depth using Spearman correlation in R (R Core Team, 2018) to see if morphological differentiation is associated with water depth.

DNA-extraction and sequencing

We sequenced all individuals caught in the 4m, 12m, 40m, 60m, and 90m net. Only for the 20m net, where we caught a total of 31 individuals, we randomly downsampled the number of individuals to 16 to achieve a balanced sampling across all depths. DNA was extracted from fin clips with the Qiagen DNeasy blood and tissue kit (Qiagen AG, CH), using the standard protocol for tissue samples supplied by the manufacturer. DNA concentrations were quantified on a Qubit 2 fluorometer (Thermo Fisher Scientific AG, CH). An Illumina paired-end TruSeq DNA PCR-Free library (Illumina GmbH, CH) was prepared for each fin-

clip sample. Library preparation was performed by the NGS platform of the University of Bern following the manufacturer's instructions. Libraries were then sequenced paired-end 150bp on an Illumina Novaseq 6000 S4 flow cell. Individual sequencing coverage at polymorphic sites (see next section, called across all 91 *C. macrophthalmus* individuals sequenced and with data from at least 85 individuals at each position) was on average ~8.6x and ranged between ~4.3x and ~16.1x in the 91 sequenced *C. macrophthalmus* individuals. Individual coverage did not differ significantly between the different sampling depths according to a one-way ANOVA (p=0.163) performed in R (R Core Team, 2018). Mean coverage was ~8x (range between ~5.6x and ~13.6x) in the 4m spawning depth population, ~7.4x (range between ~4.3x and ~13x) in the 12m spawning depth population, ~8.1x (range between ~4.7x and ~13.9x) in the 20m spawning depth population, ~10.4x (range between ~6.2x and ~16.1x) in the 40m spawning depth population, ~10x (range between ~5x and ~14.8x) in the 60m spawning depth population, and ~8.1x (range between ~5.1x and ~14.1x) in the 90m spawning depth population.

Processing reads and mapping

Raw reads were processed and mapped to the Alpine whitefish reference genome following Frei et al. (2022). In brief, poly-G tails were removed using fastp 0.20.0 (Chen, Zhou, Chen, & Gu, 2018) and overlapping read pairs with overlaps longer than 25bp were subsequently merged using Seqprep 1.0 (https://github.com/jstjohn/SeqPrep). The processed reads were then mapped to Alpine whitefish reference genome (De-Kayne, Zoller, & Feulner, 2020) using BWA 0.7.12 (Li & Durbin, 2009) adjusting the "r" parameter to 1. We marked duplicate reads, fixed mate information and replaced read groups (settings used except for the default parameters were VALIDATION_STRINGENCY=LENIENT and MAX_FILE_HANDLES_FOR_READ_ENDS_MAP=1024) using picard tools 2.20.2 (http://broadinstitute.github.io/picard/).

Population genomic analysis

We then used angsd version 0.925 (Korneliussen, Albrechtsen, & Nielsen, 2014) to calculate genotype likelihoods across all 96 samples caught at the 6 different depths, and additionally included all historical individuals of Frei et al. (2022) (short read archive accession number PRJEB43605) to verify species assignment that was done in the field. Only sites covered with at least two reads in at least 118 individuals (out of a total of 128 individuals) and passing a p-value cut-off of 10E-6 for being variable were included, while all sites with more than two alleles were excluded. 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. The following p-value cut-offs for SNP filters implemented in angsd version 0.925 (Korneliussen et al., 2014) were used: -sb_pval 0.05 -qscore_pval 0.05 -edge_pval 0.05 -mapq_pval 0.05. To verify species assignment based on the resulting 941'976 SNPs with minor allele frequency above 0.05 (default parameter of PCAngsd 0.98), we did a PCA and calculated admixture proportions based on the thirst three eigenvectors (-e 3) using PCAngsd 0.98 (Meisner & Albrechtsen, 2018). In total, we identified one individual to belonging to *C. wartmanni* (or possibly being early generation hybrids) and four individuals belonging to *C. arenicolus* (matching our species assignment done in the field) and thus these five samples have been excluded from subsequent analysis. We used a generalized linear model (glm) in R (R Core Team, 2018) to test whether the *C. gutturosus* admixture proportions were different between the different depths that we sampled.

We then used eleven *C. gutturosus* individuals and the two historical *C. macrophthalmus* individuals from Frei et al. (2022) in combinations with the 91 *C. macrophthalmus* individuals caught at either 4m, 12m, 20m, 40m, 60m, or 90m to test for introgression from *C. gutturosus* into each spawning depth population separately, using the population-based D-statistics (Soraggi, Wiuf, & Albrechtsen, 2018) implemented in angsd

0.925 (Korneliussen et al., 2014). We first calculated genotype likelihoods using only the 104 above mentioned samples and using the same parameters as described above but adjusting the missing data parameter to include sites with data from at least 99 individuals. This resulted in a total of 517'250 SNPs that were then used for the D-statistics. We used a *S. salar* individual from Kjaerner-Semb et al. (Kjaerner-Semb et al., 2016) (short read archive accession number: SSR3669756) as outgroup P4, the eleven *C. gutturosus* individuals as donor population P3, all *C. macrophthalmus* individuals of one of the six sampled depths as P2 and the two historical *C. macrophthalmus* as P1. By that ordering of populations on the four-taxon topology, it is possible to test for excess allele sharing between *C. gutturosus* and post-eutrophication populations of extant species relative to the same species sampled pre-eutrophication, which would be indicative of introgression of *C. gutturosus* into this species during eutrophication. We then repeated the analysis but replaced the donor population P3 with all historical *C. arenicolus* and *C. wartmanni* individuals from Frei et al. (2022) to test for introgression of *C. arenicolus* or *C. wartmanni* respectively into our six *C. macrophthalmus* spawning depth populations that must have happened during eutrophication-induced speciation reversal.

We then calculated genotype likelihoods again, using the same parameters as above but only using the 91 *C. macrophthalmus* samples and adjusting the missing data parameter to include sites where data for at least 85 individuals was available. We used the resulting genotype likelihoods at 1'948'989 polymorphic sites to calculate (weighted) F_{ST} (Bhatia, Patterson, Sankararaman, & Price, 2013) between all possible pairs of spawning depth populations (as well as between 90m and all other spawning depth populations pooled) in angsd 0.925 (Korneliussen et al., 2014) based on one- and two-dimensional site frequency spectra inferred from site allele frequencies (Nielsen, Korneliussen, Albrechtsen, Li, & Wang, 2012). We then again calculated a PCA with PCAngsd (0.98) (Meisner & Albrechtsen, 2018)

at the 1’126’828 SNPs with minor allele frequency above 0.05 (default parameter of PCAngsd 0.98) to visualize population structure across depth within *C. macrophthalmus*. We further performed a selection scan along PC1 using PCAngsd (0.98) (Meisner & Albrechtsen, 2018) according to the method proposed by Galinksy et al. (2016). The method identifies unusual allele frequency shifts along previously inferred PC-axes, making use of the fact that the squared correlation of each SNP to a specific PC-axis, rescaled to account for genetic drift, follows a chi-square distribution (1 d.o.f) under the null hypothesis of the absence of selection (Galinsky et al., 2016). As PC1 separated the fish caught at 90m from all the fish caught shallower (4m, 12m, 20m, 40m, 60m), this selection scan would detect positions that are under selection in deep-water, respectively involved in depth adaptation. Following Pinsky et al. (2021), we FDR-corrected the resulting P-values and assumed SNPs with an FDR-corrected P-value below 0.05 to be under selection. P-values were then log transformed for plotting using R (R Core Team, 2018).

At the 107 SNPs above the *FDR*-corrected significance threshold from the PCA-based selection scan, we used angsd 0.925 (Korneliussen et al., 2014) to calculate allele frequencies from genotype likelihoods of each spawning depth population separately using the method described in Kim et al. (2011), and we fixed the tracked allele to represent the reference allele of the Alpine whitefish reference genome. To remove redundant sites in strong physical linkage, we only considered positions that are more than 5 Mbp apart from each other. We retained 52 SNPs for further analysis. We then additionally calculated the allele frequency in all historical *C. gutturosus* (n=11) individuals, as well as in all historical *C. macrophthalmus* (n=2), *C. arenicolus* (n=3) and *C. wartmanni* (n=2) individuals from Frei et al. (2022). SNPs that have a minor allele frequency above 0.05 in *C. gutturosus*, but are absent from all historical *C. macrophthalmus*, *C. arenicolus and C. wartmanni* have potentially been characteristic for *C. gutturosus* before the eutrophication period. Considering that our data

showed that there was significant *C. gutturosus* introgression, detecting an allele with such a frequency pattern in contemporary populations of the extant species suggests that this allele introgressed from *C. gutturosus* during the anthropogenic eutrophication period. Following this logic, we looked for SNPs with such an allele frequency pattern consistent with *C. gutturosus* introgression among the 52 independent SNPs inferred to be under selection between deep and shallower spawning *C. macrophthalmus* to find SNPs with alleles that potentially introgressed from *C. gutturosus* that may now facilitate deep-water adaptation in deep spawning *C. macrophthalmus*. We tested by permutation if the 52 sites potentially under selection between deep and shallower caught *C. macrophthalmus* are significantly enriched for SNPs with an allele frequency pattern consistent with *C. gutturosus* introgression. We randomly subsampled 52 positions (the same number as inferred to be under selection between deep and shallower spawning *C. macrophthalmus*), and then calculated the proportion of these subsampled SNPs that show an allele frequency pattern consistent with *C. gutturosus* introgression. We repeated this random subsampling 10'000 times to generate a null expectation, and then calculated a P-value by comparing the expected proportion of sites showing an allele frequency pattern of *C. gutturosus* introgression of these 10,000 permutations with the observed proportion calculated within the 52 sites potentially under selection between deep and shallower spawning *C. macrophthalmus*.

Finally, we assessed if the 107 SNPs (in 52 independent genomic regions) inferred to be under selection fall within genes, and if yes, in which genes. Gene annotations (from the Alpine whitefish genome (De-Kayne et al., 2020); ENA accession: GCA_902810595.1) that overlap with the loci potentially under selection were identified using bedtools v.2.28.0 (Quinlan, 2014). We then used the protein sequence of the overlapping gene from the Alpine whitefish genome (De-Kayne et al., 2020; ENA accession: GCA_902810595.1) to perform a

protein-protein BLAST (blastp) search against all genes of the annotation of the *S. salar* genome (taxid 8030). We reported the best hit for each gene (Supplementary Table 4).

Results

Sampling populations of *C. macrophthalmus* along a spawning depth gradient

We sampled the entire known spawning depth range of *C. macrophthalmus* (4m, 12m, 20m, 40m), as *well* as greater depths where the extinct *C. gutturosus* used to spawn (60m, 90m, 120m) during the *C. macrophthalmus* spawning season end of November 2019 (see Figure 1A). Our sampling timepoint was in the middle of the typical spawning season of the targeted *C. macrophthalmus* (early November until early January), but also overlapped the spawning season reported for the now extinct *C. gutturosus*, ranging from July to early January (Steinmann, 1950). In total, we caught 106 *C. macrophthalmus* individuals, of which 93 (~88%) were fully ripe. While most fish were caught at 20m (n=31, Figure 1B), we caught spawning *C. macrophthalmus* individuals down to 90m, but no fish were caught at the greatest depth fished (120m). This suggests that our sampling covered the entire range of depth that is currently used for spawning by whitefish.

To verify our species assignment that was done in the field, we performed a genomic principal component analysis (PCA; Supplementary Figure 1) and a structure analysis (Supplementary Figure 2) *using* genotype likelihoods of 941'976 SNPs. We included all 96 sequenced individuals, as well as historical and contemporary individuals used in Frei et al. (2022) for reference. In both genomic PCA and structure analysis (Supplementary Figures 1 and 2), four individuals of a total of 96 turned out to belong to *C. arenicolus*, and thus were subsequently excluded from further analysis. One individual caught at 90m clustered with *C. wartmanni* in the PCA and looked like an early generation hybrid in the structure analysis. This fish was also excluded from all subsequent analyses to ensure that the results reflect solely the variation within *C. macrophthalmus* (n=91).

Morphological and genomic differentiation along the spawning depth gradient

We performed a partial least squares regression analysis based on linear morphometric measurements of 23 body- and head traits (size corrected by using residuals of linear regression against standard length) against spawning depth on all 106 individuals assigned to *C. macrophthalmus* (91 individuals randomly selected for sequencing, plus the 15 individuals that were caught but not sequenced). We found indications for morphological differentiation along depth (Figure 1D), with component 1 being significantly correlated with depth (rho=-0.44, p=2e-06, Supplementary Figure 3). The traits with the highest loadings on both component 1 (LJW=0.70, SD=0.69 and MW=0.39) and component 2 (SW=-0.55, EH=-0.49 and SNL=-0.35) were related to mouth (and head) shape (Supplementary Figure 4). In contrast, our analysis of the genomic data did not yield any evidence for genomic differentiation along the spawning depth gradient when genome-wide F_{ST} was used as test metric. We performed a genomic PCA based on genotype-likelihoods of 1'126'828 SNPSs in all sequenced individuals genetically assigned to *C. macrophthalmus* (n=91; out of a total of 96 individuals that were randomly selected for sequencing). We found that the principal component explaining most variation (PC1, Figure 1C) separates the deepest caught individuals (90m) from all others, but the genome-wide F_{ST} between the 90m sample and all shallower caught individuals did not differ from zero (weighted F_{ST}= -0.001567; all pairwise genome-wide F_{ST}'s between depth categories were below zero). Similar to the morphological results, genomic PC1 was significantly correlated with depth (rho=-0.37, p=0.0003, Supplementary Figure 3). Taken together, our high-resolution data allows to identify subtle intra-specific differentiation within *C. macrophthalmus* in both genomic and morphological data (Figure 1C and Figure 1D). The correlation of both morphological and genomic variation with depth suggests that the observed differentiation may be related to the onset of adaptation to deep-water. Even though we did not observe genome-wide differentiation based on

genome-wide F_{ST} estimates which reflect neutral demographic processes that affect all SNPs (as the majority of SNPs along the genome are expected to evolve neutrally), our PCA approach demonstrates intraspecific differentiation within *C. macrophthalmus* between fish spawning at 90m and all shallower spawning individuals. However, this pattern might be driven by relatively few loci, which potentially show differentiation in consequence of selective processes.

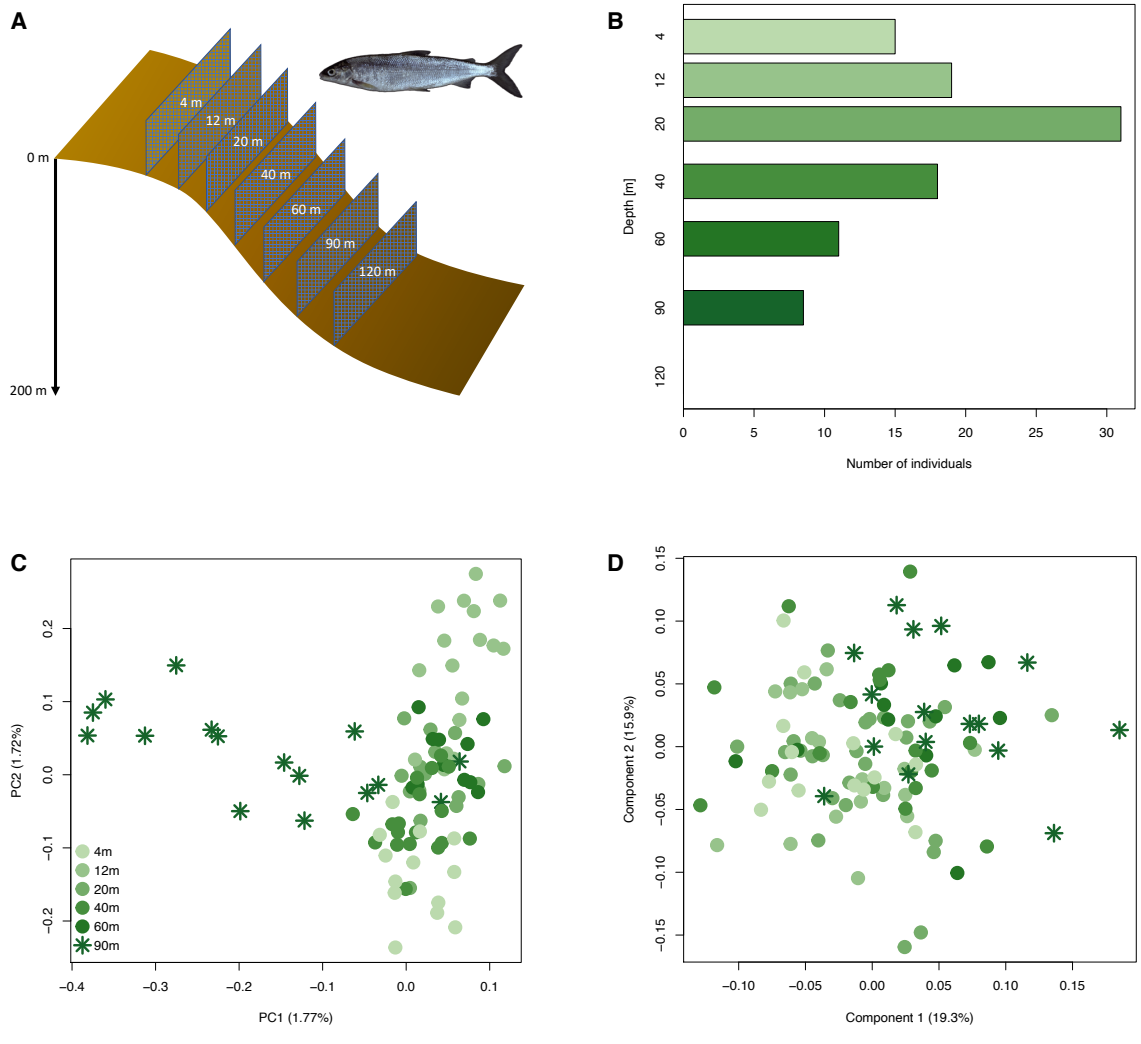


Figure 1: Differentiation along a water depth gradient. A) Schematic overview of the sampling structure with nets set at seven different depths. **B)** Number of individuals caught at each depth within 18h (same sampling effort for each depth). **C**) Genomic variation based on 1'126'828 polymorphic sites illustrated by a principal component analysis (PCA). The depth category of each individual is indicated by different green shadings (the deeper the darker). The 90m spawning depth population is highlighted with asterisks, while all other spawning depth populations are indicated by dots. **D**) Morphological differentiation is displayed as the two major components resulting from the partial least squares regression analysis. Symbols and colours are the same as in Figure 1C.

Introgression from extinct deep-water species

We tested whether the six different *C. macrophthalmus* spawning depth populations (n=11-16) received significant introgression from either *C. gutturosus*, *C. arenicolus* and/or *C wartmanni*, by making use of the historical samples of Frei et al (2022). We detected significant introgression from *C. gutturosus* and *C. wartmanni* into each of our six *C. macrophthalmus* spawning depth populations (n=11-16; Supplementary Table 1 and 2), but we did not detect significant introgression from *C. arenicolus* (Supplementary Table 3). Per-individual *C. gutturosus* admixture proportions were not different between different depths (see Figure 2; p=0.616 in a generalized linear model). However, the variance in admixture proportion was highest in the two deepest nets (60 and 90m) and the individual with the highest admixture proportion (~14%) was caught at 60m.

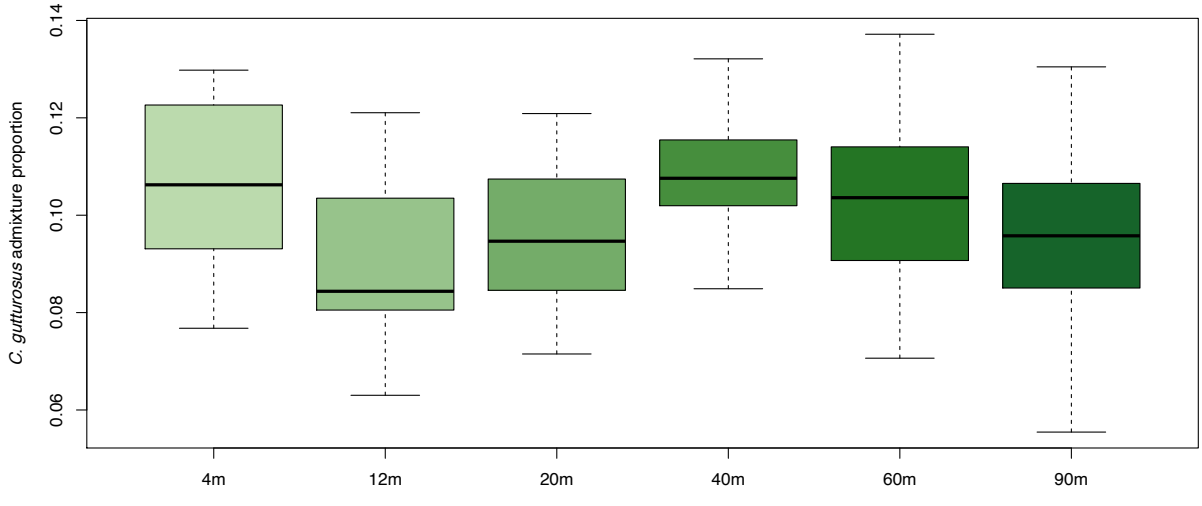


Figure 2: No differences in admixture proportions between spawning depth populations. Boxplots showing the *C. gutturosus* admixture proportions from the PCAngsd admixture analysis (see Supplementary Figure 2) in each spawning depth population. Horizontal bars correspond to medians, and whiskers to 1.5 times the interquantile range. There were no significant differences in admixture proportions between spawning depth populations (p=0.616 in a generalized linear model).

Identifying genomic positions shaped by selection along the water depth gradient

We performed the selection scan proposed by Galinsky et al. (Galinsky et al., 2016) implemented in PCAngsd (Meisner & Albrechtsen, 2018; Meisner, Albrechtsen, & Hanghoj, 2021) to find positions under selection between the *C. macrophthalmus* spawning depth populations. The method works best with data that is continuously distributed in PC-space

(Galinsky et al., 2016; Meisner & Albrechtsen, 2018) (see Figure 1C) and identifies positions that significantly deviate from genetic drift along an axis of differentiation (Meisner et al., 2021). As our PC1 (see Figure 1) is separating most of the 90m *C. macrophthalmus* from all samples from shallower depths, positions that are detected to be under selection along this PC-axis would thus potentially be involved in adaptation to deep water. In total, we found 107 outlier SNPs (FDR-corrected p<0.05) in 52 independent genomic regions (at least 5 Mbp apart from each other) that are potentially under selection between the 90m spawning site and shallower sites (Figure 3A). These 107 outlier SNPs overlapped a total of 30 genes (see Supplementary Table 4).

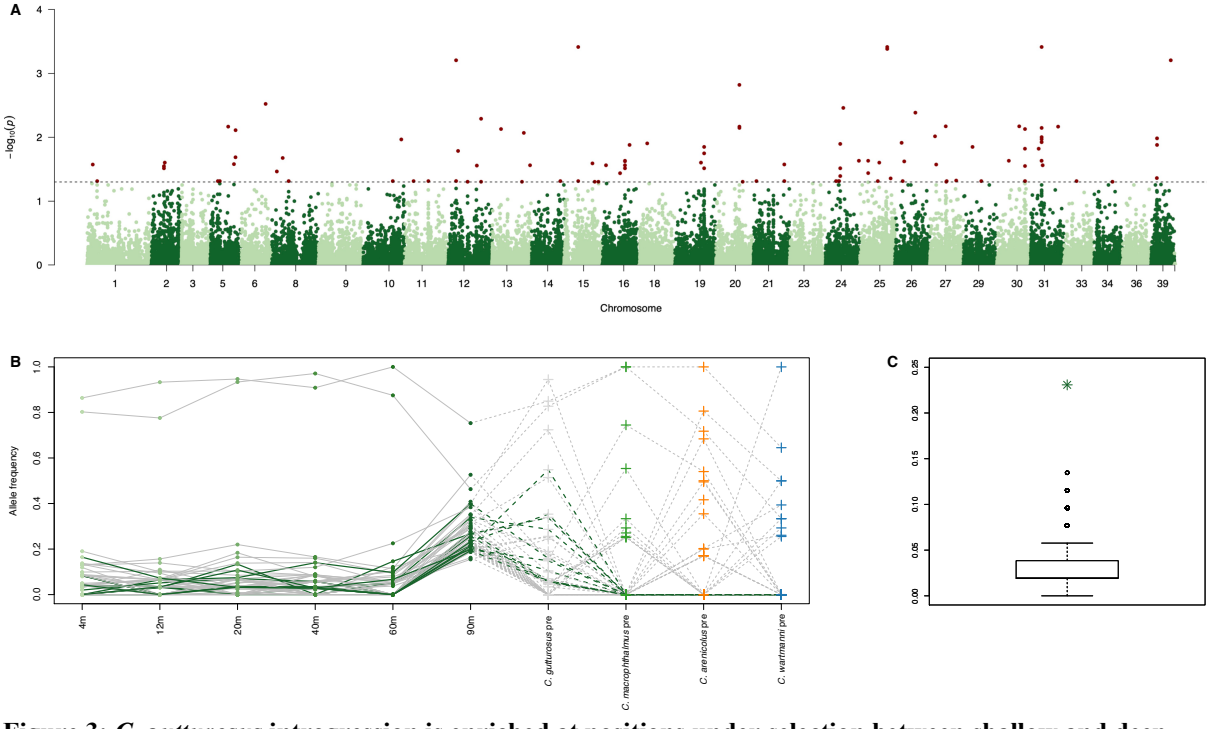


Figure 3: *C. gutturosus* introgression is enriched at positions under selection between shallow and deep water. A) Selection scan along PC1 from our genomic PCA (see Figure 1C) from Galinksy et al. (2016) as implemented in PCAngsd (Meisner & Albrechtsen, 2018; Meisner et al., 2021). As PC1 separates the 90m population from shallower spawning *C. macrophthalmus* populations, this approach identifies positions potentially under selection between deep and shallow spawning individuals. Shown are log-transformed and FDR-corrected p-values. The dashed line indicates the FDR-corrected 0.05 significance threshold, and all positions with p-values below the threshold are coloured in darkred. **B**) Allele frequencies in the six different spawning depth populations (4-90m) and in historical populations from Frei et. al (2022) (indicated with crosses and denoted with "pre"; grey for *C. gutturosus* (n=11), green for *C. macrophthalmus* (n=2), orange for *C. arenicolus* (n=3) and blue for *C. wartmanni* (n=2)). Shown are the 52 positions with an FDR-corrected p-value below 0.05 and at least 5 Mbp apart from each other. SNPs derived from *C. gutturosus* and potentially introgressed into *C. macrophthalmus* during eutrophication (frequency in *C. gutturosus* above 0.05, but allele is absent from all other historical populations) are coloured in green. **C**) The distribution of 10'000 permutations of 52 randomly sampled positions (same number as shown in Figure 2B) along the genome, showing the proportion of alleles derived from *C. gutturosus* and potentially introgressed into *C. macrophthalmus* during eutrophication

(same pattern as dark green trajectories in B). The green asterisk indicates the observed value (the 12 out of 52 in B).

Allele frequency patterns consistent with introgression from extinct deep-water species

To assess whether adaptation to deep water in *C. macrophthalmus* was potentially facilitated by alleles introgressed from *C. gutturosus*, we assessed the population allele frequency in our six spawning depth populations (4m, 12m, 20m, 40m, 60m and 90m) at the 52 independent genomic positions with evidence for selection between the deepest (90m) and all shallower (4-60m) spawning *C. macrophthalmus* populations. Additionally, we also inferred allele frequencies at the same positions in 11 historical *C. gutturosus* individuals, two historical *C. macrophthalmus* individuals, three historical *C. arenicolus* individuals and two historical *C. wartmanni* individuals sampled from before the onset of eutrophication from Frei et al (2022). At 12 out of 52 (23.1%) positions under selection between deep and shallower spawning *C. macrophthalmus*, we found that the alternate allele was present in *C. gutturosus* before the eutrophication period, while the allele was absent in the historical *C. macrophthalmus*, *C. arenicolus* and *C. wartmanni* samples from Frei et al. (2022) (Figure 3B). This pattern of allele frequencies across populations and species suggests that these alleles, potentially involved in adaptation to deep water in *C. macrophthalmus*, likely introgressed from *C. gutturosus* during the anthropogenic eutrophication period. Sites with such an allele frequency pattern consistent with *C. gutturosus* introgression were significantly enriched among the 52 independent sites that are potentially under selection between deep and shallower caught *C. macrophthalmus* ($p<10e-4$ obtained with 10'000 permutations, Figure 3C), suggesting that introgressed alleles from *C. gutturosus* may facilitate adaptation to deep water in *C. macrophthalmus*.

Discussion

Anthropogenic eutrophication of Lake Constance during the last century resulted in the extinction of the endemic profundal whitefish species *Coregonus gutturosus*, caused by a combination of demographic decline and speciation reversal through introgressive hybridization with other species of the same radiation. Introgression during speciation reversal resulted in the persistence of considerable parts of genomic variation from the extinct species within several extant species (Frei et al., 2022). We here show that one of the surviving Lake Constance whitefish species, *C. macrophthalmus*, is currently re-populating the deep-water environment that was left vacated after the extinction of *C. gutturosus.* Our systematic sampling of a spawning depth gradient demonstrated that today, *C. macrophthalmus* is spawning in greater depths (down to 90m) than previously reported for this species (less than ~20m before eutrophication in e.g., Nüsslin (1907) and Schweizer (1926), or ~20m in Eckmann & Rösch (1998) and 2-50m in Jacobs et al. (2019) after eutrophication). Our data suggests that introgression from *C. gutturosus* that occurred during its decline, potentially facilitates ongoing adaptation to the vacated deep-water niche in *C. macrophthalmus*.

Re-population of and adaptation to the vacated deep-water environment

Adaptation to deep water conditions at the lower end of a species' depth range is expected to result in morphological and genomically localized rather than genome-wide differentiation between populations spawning deep and those that spawn shallower, as selection is thought to favour phenotypes or combinations of alleles that increase fitness in deep-water habitats. Founder effects during range expansion might mimic adaptation, and hence could be mistaken for signals of adaptation. However, a founder effect would require some degree of geographical isolation between founder and source populations. As *C.*

macrophthalmus has expanded its ecological niche at a spatial scale that lies within the dispersal distance of a single individual, a founder effect in the 90m spawning depth population is unlikely. Furthermore, a founder effect in the deepest spawning population should be associated with an increase of genetic drift due to reduced effective population size. As a consequence, genetic diversity would decrease, resulting in genome-wide differentiation to all other spawning depth populations. In contrast with this prediction, we did not find evidence for genomic differentiation measured with genome-wide F_{ST}, which would reflect genome-wide differentiation resulting from demographic processes. Our data provides evidence for both subtle morphological and genomic differentiation between the deepest-spawning (90m) and shallower (4-60m) spawning populations of *C. macrophthalmus*. This subtle genomic differentiation might be genomically localized rather than genome-wide, considering the evidence for differentiation between the 90m spawning depth population in the PCA (see Figure 1C), but no evidence for such differentiation when using genome-wide F_{ST}. Additionally, the major axis of genomic differentiation and of morphological differentiation were both correlated with spawning depth. This suggests that the observed intraspecific differentiation between the 90m spawning depth population and all shallower spawning depth populations might indeed be a result of adaptation to the vacant deep-water niche.

Until recently, whitefish have been reported absent from deep-water habitats of Lake Constance (Alexander & Seehausen, 2021). Thus, the colonization of and adaptation to deep water in *C. macrophthalmus* described here has likely started only recently. This is consistent with theoretical work that demonstrated that, when a species goes extinct through hybridization, the re-population of its habitat is likely when disturbance that led to reversal is of only short duration (Gilman & Behm, 2011). Hybridization during the extinction process facilitates the re-emergence of a similar phenotype to that of the extinct species through a combination of alleles derived from surviving and the extinct species, finally enabling the re-

population of the vacated habitat. Adaptation and diversification by re-assembling alleles from two hybridizing species into new adaptive trait combinations is thought to be orders of magnitudes faster than adaptation and speciation based on *de-novo* mutation alone (Marques et al., 2019), and thus might be an important process in rapid adaptation to changing environments.

Introgression facilitates adaption to extinct species' habitat

Introgression from *C. gutturosus* during eutrophication-induced speciation reversal might have provided the contemporary *C. macrophthalmus* population with alleles that are adaptive in deep water. The significant introgression from *C. gutturosus* into *C. macrophthalmus* caught in any depth zone demonstrated by D-statistics in combination with the matching allele frequencies between *C. gutturosus* and our 90m *C. macrophthalmus* sample suggests that parts of the adaptation to deep water in *C. macrophthalmus* could be based on selection on introgressed variation derived from *C. gutturosus*. Ecological selection on introgressed variation is predicted to result in biased ancestry around functionally relevant genomic regions (Moran et al., 2021). However, as both positive and negative selection may act on genomic variation derived from introgression (as for example demonstrated for Neanderthal introgression into some populations of *Homo sapiens*; see e.g., Huerta-Sanchez et al. (2014), Racimo, Sankararaman, Nielsen, & Huerta-Sanchez (2015), Reilly et al. (2022) and Harris & Nielsen (2016)), determining the exact selective forces acting on a specific allele is complex and challenging (Moran et al., 2021). At twelve out of 52 independent positions (23.1%) indicating signatures of positive selection (and thus potentially involved in adaptation to depth), the patterns of allele frequencies suggested that the allele with increased frequency in the 90m spawning depth population might have introgressed from *C. gutturosus*. Even though the sample sizes for the historical populations of the three extant species is limited, our permutation approach demonstrates that alleles that likely introgressed from *C.*

gutturosus are enriched at positions under divergent selection between the deepest (90m) and shallower spawning (4-60m) *C. macrophthalmus* populations. This suggests that introgression from the extinct *C. gutturosus* might facilitate adaptation to its former deep-water habitat in the extant *C. macrophthalmus*. These results are in line with recent work that demonstrated that adaptation based on variation derived from recent admixture events can be very rapid and take only few generations (Hamid et al., 2021). Further, our findings are consistent with the syngameon hypothesis of adaptive radiations, predicting that hybridization between species within an adaptive radiation can promote further diversification and speciation (Seehausen, 2004). Especially when environments change, hybridization within an adaptive radiation might increase the genomic variation of individual species and thereby enhance their adaptive potential, enabling a faster evolutionary response to the novel selective pressures of a changing environment (Grant & Grant, 2019). Consequently, hybridization between members of an adaptive radiation might be especially relevant under environmental change, potentially facilitating the survival of several species or even all species in a radiation through elevated evolvability and faster adaptation to the changing environmental conditions.

Ecological recovery through evolution

Anthropogenic environmental change is affecting ecosystems worldwide whilst a large portion of contemporary species diversity is sensitive to hybridization-driven dynamics (Grabenstein & Taylor, 2018). Thus, the potential for speciation reversal to affect evolutionary trajectories of species and lineages is enormous (Seehausen et al., 2008). Our results suggest that the colonization of a vacated niche could potentially occur on a short evolutionary time-scale when adaptation of an extant species is facilitated by introgression of alleles from the species that occupied this niche in the past and now is extinct. Such introgressed alleles have already been tested by selection in the environment originally inhabited by the extinct species. These alleles likely have the potential to facilitate rapid

adaptation of the recipient species to these environmental conditions, provided that the disturbance resulting in extinction through speciation reversal was transient and of short duration (Gilman & Behm, 2011). This highlights the importance of quick and efficient ecosystem restorations after anthropogenic disturbances to maximize the chance of ecological recovery through evolution.

Hybridization in response to homogenized environments can result in dramatic losses of biodiversity within few generations (Taylor et al., 2006; Grabenstein & Taylor, 2018). However, hybridization can as well facilitate adaptation when environments become more heterogeneous again and thus promote the evolution of new biodiversity (Moran et al., 2021). When alleles that have evolved in a now extinct species introgress into a surviving species, they can outlast the species they evolved in and potentially be re-used to adapt to the extinct species' vacated habitat, other habitats, or changed environmental conditions. The role of hybrid populations has been controversial in conservation biology (Draper, Laguna, & Marques, 2021). However, hybrid populations with high genetic variation and in turn high evolvability, such as those resulting from speciation reversal, can be important for future evolutionary dynamics that could contribute to the ecological recovery of an ecosystem. In turn, efficient and informed conservation measures should consider the implications of the existence of such hybrid populations with high adaptive potential and the evolutionary dynamics that can emerge from them, potentially contributing to the recovery of an ecosystem on a time-scale that is much shorter than the usually assumed evolutionary timescales of millenia.

Acknowledgements

We thank Christoph Birrer, Michael Kugler and Jörg Schweizer from the Amt für Natur, Jagd und Fischerei, St.Gallen and we are very thankful for their valuable information and support during planning and sampling. We want to thank Carmela Dönz, Marta Reyes, Oliver Selz and Lukas Aerne for help during fieldwork and Andres Grolimund for measuring linear morphometric traits and counting gill-rakers. 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. 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.

Data accessibility

The raw sequencing files are accessible on ENA SRA (PRJEB53050). Additional supporting data (including the morphological data, genotype likelihood files and ENA sample accessions of all historical samples used) is deposited on the eawag research data institutional collection (doi:10.25678/0007FH). Scripts used for data analysis are available on GitHub (https://github.com/freidavid/Lake_Constance_Depth_Transect).

Author contributions

DF, OS and PGDF conceived of, designed, and conceptualized the study. PGDF managed and supervised the study. PR and DF together planned and run the field sampling and processed all specimen and samples for further analysis. DF analysed morphological and 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.

References

Alexander, T., & Seehausen, O. (*2021*). *Diversity, distribution and community composition of fish in perialpine lakes - "Projet Lac" synthesis report*. Bern: 2021.

Barlow, A., Cahill, J. A., Hartmann, S., Theunert, C., Xenikoudakis, G., Gonzalez Fortes, G., . . . Hofreiter, M. (2018). Partial genomic survival of cave bears in living brown bears. *Nature Ecology & Evolution, 2*(10), 1563-1570. doi:10.1038/s41559-018-0654-8

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F_{ST}: The impact of rare variants. *Genome Research, 23*(9), 1514-1521. doi:10.1101/gr.154831.113

Chen, S. F., Zhou, Y. Q., Chen, Y. R., & Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), 884-890. doi:10.1093/bioinformatics/bty560

De-Kayne, R., Zoller, S., & Feulner, P. G. D. (2020). A de novo chromosome-level genome assembly of *Coregonus sp.* "Balchen": One representative of the Swiss Alpine whitefish radiation. *Molecular Ecology Resources, 20*(4), 1093-1109. doi:10.1111/1755-0998.13187

Deufel, J., Löffler, H., & Wagner, B. (1986). Auswirkungen der Eutrophierung und anderer anthropogener Einflüsse auf die Laichplätze einiger Bodensee-Fischarten. *Österreichs Fischerei, 39*, 325-336.

Draper, D., Laguna, E., & Marques, I. (2021). Demystifying negative connotations of hybridization for less biased conservation policies. *Frontiers in Ecology and Evolution, 9*(637100). doi:10.3389/fevo.2021.637100

Eckmann, R., & Rösch, R. (1998). Lake Constance fisheries and fish ecology. *Advances in Limnology, 53*, 285-301.

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, 6*(4), 461-468. doi:10.1038/s41559-022-01665-7

Galinsky, K. J., Bhatia, G., Loh, P. R., Georgiev, S., Mukherjee, S., Patterson, N. J., & Price, A. L. (2016). Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *American Journal of Human Genetics, 98*(3), 456-472. doi:10.1016/j.ajhg.2015.12.022

Gilman, R. T., & Behm, J. E. (2011). Hybridization, species collapse, and species reemergence after disturbance to premating mechanisms of reproductive isolation. *Evolution, 65*(9), 2592-2605. doi:10.1111/j.1558-5646.2011.01320.x

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking barriers: causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution, 33*(3), 198-212. doi:10.1016/j.tree.2017.12.008

Grant, P. R., & Grant, B. R. (2019). Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America, 116*(46), 23216-23224. doi:10.1073/pnas.1913534116

Hamid, I., Korunes, K. L., Beleza, S., & Goldberg, A. (2021). Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. *Elife, 10*. doi:10.7554/eLife.63177

Harris, K., & Nielsen, R. (2016). The genetic cost of Neanderthal introgression. *Genetics, 203*(2), 881-891. doi:10.1534/genetics.116.186890

Hirsch, P. E., Eckmann, R., Oppelt, C., & Behrmann-Godel, J. (2013). Phenotypic and genetic divergence within a single whitefish form - detecting the potential for future divergence. *Evolutionary Applications, 6*(8), 1119-1132. doi:10.1111/eva.12087

Hudson, A. G., Lundsgaard-Hansen, B., Lucek, K., Vonlanthen, P., & Seehausen, O. (2016). Managing cryptic biodiversity: Fine-scale intralacustrine speciation along a benthic

gradient in Alpine whitefish (Coregonus spp.). *Evolutionary Applications*, *10*, 251-266. doi:10.1111/eva.12446

Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., . . . Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature, 512*(7513), 194-197. doi:10.1038/nature13408

Ingram, T., Hudson, A. G., Vonlanthen, P., & Seehausen, O. (2012). Does water depth or diet divergence predict progress towards ecological speciation in whitefish radiations?. *Evolutionary Ecology Research, 14*(4), 487-502. doi:10.7892/boris.16333

Irisarri, I., Singh, P., Koblmuller, S., Torres-Dowdall, J., Henning, F., Franchini, P., . . . Meyer, A. (2018). Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nature Communications, 9*(3159). doi:10.1038/s41467-018-05479-9

Jacobs, A., Carruthers, M., Eckmann, R., Yohannes, E., Adams, C. E., Behrmann-Godel, J., & Elmer, K. R. (2019). Rapid niche expansion by selection on functional genomic variation after ecosystem recovery. *Nature Ecology & Evolution, 3*, 77-86. doi:10.1038/s41559-018-0742-9

Kim, S. Y., Lohmueller, K. E., Albrechtsen, A., Li, Y. R., Korneliussen, T., Tian, G., . . . Nielsen, R. (2011). Estimation of allele frequency and association mapping using next-generation sequencing data. *Bmc Bioinformatics, 12*(231). doi:10.1186/1471-2105-12-231

Kjaerner-Semb, E., Ayllon, F., Furmanek, T., Wennevik, V., Dahle, G., Niemela, E., . . . Edvardsen, R. B. (2016). Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection. *Bmc Genomics, 17*(610). doi:10.1186/s12864-016-2867-z

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *Bmc Bioinformatics, 15*(356). doi:10.1186/s12859-014-0356-4

Kuhlwilm, M., Han, S., Sousa, V. C., Excoffier, L., & Marques-Bonet, T. (2019). Ancient admixture from an extinct ape lineage into bonobos. *Nature Ecology & Evolution, 3*(6), 957-965. doi:10.1038/s41559-019-0881-7

Lamichhaney, S., Berglund, J., Almen, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., . . . Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature, 518*(7539), 371-375. doi:10.1038/nature14181

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almen, M. S., Webster, M. T., . . . Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science, 352*(6284), 470-474. doi:10.1126/science.aad8786

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M. G., Field, C. B., & Knowlton, N. (2020). Climate change and ecosystems: threats, opportunities and solutions. *Philosophical Transactions of the Royal Society B - Biological Sciences, 375*(20190104). doi:10.1098/rstb.2019.0104

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution, 34*(6), 531-544. doi:10.1016/j.tree.2019.02.008

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications, 8*(11). doi:10.1038/ncomms14363

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics, 210*(2), 719-731. doi:10.1534/genetics.118.301336

Meisner, J., Albrechtsen, A., & Hanghoj, K. (2021). Detecting selection in low-coverage high-throughput sequencing data using principal component analysis. *Bmc Bioinformatics, 22*(470). doi:10.1186/s12859-021-04375-2

Mevik, B. H., & Wehrens, R. (2007). The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software, 18*(2), 1-23. doi:10.18637/jss.v018.i02

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization. *Elife, 10*. doi:10.7554/eLife.69016

Nenning, S. (1834). Die *Fische des Bodensees nach ihrer äussern Erscheinung*. Konstanz. doi:10.3931/e-rara-72443

Nice, C. C., Gompert, Z., Fordyce, J. A., Forister, M. L., Lucas, L. K., & Buerkle, C. A. (2013). Hybrid speciation and independent evolution of lineages of alpine butterflies. *Evolution, 67*(4), 1055-1068. doi:10.1111/evo.12019

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. R., & Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *Plos One, 7*(7). doi:10.1371/journal.pone.0037558

Nümann, W. (1972). The Bodensee: Effects of exploitation and eutrophication on the Salmonid community. *Journal of the Fisheries Research Board of Canada, 29*(6), 833-884. doi:10.1139/f72-12

Nüsslin, O. (1907). *Coregonus wartmanni* Bloch und *macrophthalmus* Nüssl. - Differentialdiagnose für das Stadium der Dottersackbrut. *Biologisches Zentralblatt, 27*, 440-447.

Pinsky, M. L., Eikeset, A. M., Helmerson, C., Bradbury, I. R., Bentzen, P., Morris, C., . . . Star, B. (2021). Genomic stability through time despite decades of in cod on both sides of the Atlantic. *Proceedings of the National Academy of Sciences of the United States of America, 118*(15). doi:10.1073/pnas.2025453118

Prada, C., Hanna, B., Budd, A. F., Woodley, C. M., Schmutz, J., Grimwood, J., . . . Medina, M. (2016). Empty niches after extinctions increase population sizes of modern corals. *Current Biology, 26*(23), 3190-3194. doi:10.1016/j.cub.2016.09.039

Quinlan, A. R. (2014). BEDTools: The Swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics, 47*(11.12.1-11.12.34). doi: 10.1002/0471250953.bi1112s47

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sanchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics, 16*(6), 359-371. doi:10.1038/nrg3936

Reilly, P. F., Tjahjadi, A., Miller, S. L., Akey, J. M., & Tucci, S. (2022). The contribution of Neanderthal introgression to modern human traits. *Current Biology, 32*(18), PR970-R983. doi:doi.org/10.1016/j.cub.2022.08.027

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., . . . Lexer, C. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science, 301*(5637), 1211-1216. doi:10.1126/science.1086949

Rhymer, J. M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics, 27*, 83-109. doi:10.1146/annurev.ecolsys.27.1.83

Schweizer, W. (1894). Die Felchen des Bodensees und ihre natürliche und künstliche Vermehrung. *Mitteilungen der Thurgauischen Naturforschenden Gesellschaft, 11*, 13-26.

Schweizer, W. (1926). Der Gangfisch im Bodensee (Ober- und Untersee), sein Fang und seine Pflege. *Mitteilungen der Thurgauischen Naturforschenden Gesellschaft, 26*, 3-32.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution, 19*(4), 198-207. doi:10.1016/j.tree.2004.01.003

Seehausen, O. (2006). Conservation: Losing biodiversity by reverse speciation. *Current Biology, 16*(9). doi:10.1016/j.cub.2006.03.080

Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology, 17*(1), 30-44. doi:10.1111/j.1365-294X.2007.03529.x

Selz, O. M., Doenz, C. J., Vonlanthen, P., & Seehausen, O. (2020). 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

Soraggi, S., Wiuf, C., & Albrechtsen, A. (2018). Powerful inference with the D-statistic on low-coverage whole-genome data. *G3-Genes Genomes Genetics, 8*(2), 551-566. doi:10.1534/g3.117.300192

Steinmann, P. (1950). Monographie der schweizerischen Koregonen. Beitrag zum Problem der Entstehung neuer Arten. Spezieller Teil. *Schweiz. Z. Hydrologie, 12*(1), 340-491. doi:10.1007/BF02486030

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: morphological and genetic evidence of the collapse

of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology, 15*(2), 343-355. doi:10.1111/j.1365-294X.2005.02794.x

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hubner, S., . . . Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications, 9*(7), 892-908. doi:10.1111/eva.12367

von Rapp, W. (1858). Ueber den Fang des Kilch (Coregonus acronius). *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg, 14*, 328-331.

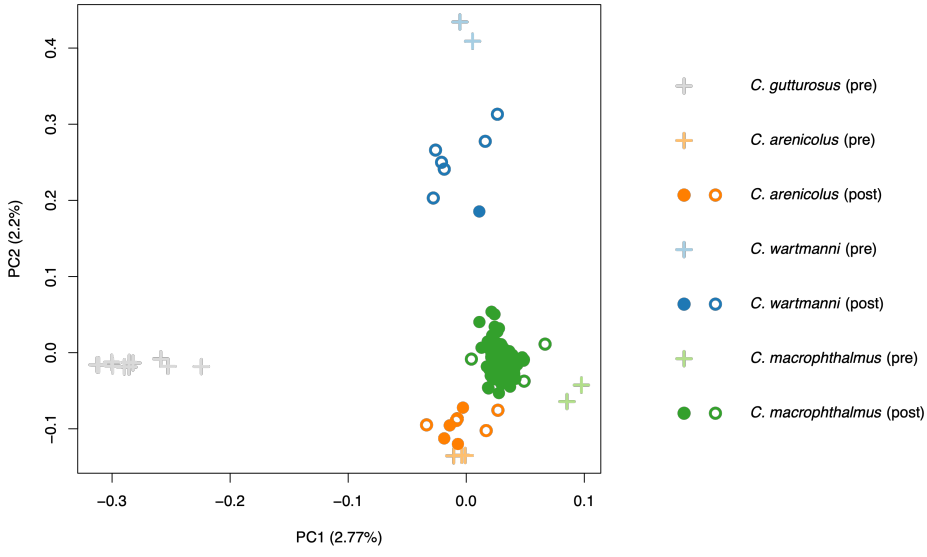
von Siebold, C. T. E. (1858). Kleinere Mittheilungen und Correspondenz - Nachrichten. Ueber den Kilch des Bodensees (Coregonus acronius). *Zeitschrift für wissenschaftliche Zoologie, 9*, 295-299.

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Muller, R., Lundsgaard-Hansen, B., . . . Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature, 482*(7385), 357-362. doi:10.1038/nature10824

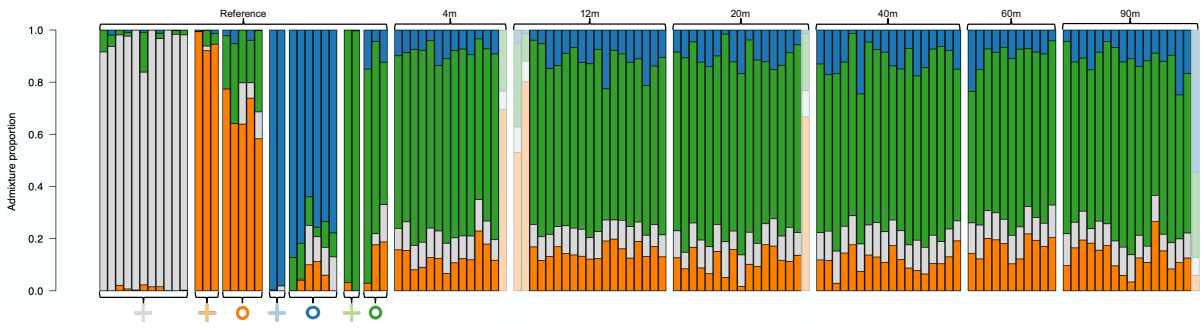
Wahl, B., & Loffler, H. (2009). Influences on the natural reproduction of whitefish (*Coregonus lavaretus*) in Lake Constance. *Canadian Journal of Fisheries and Aquatic Sciences, 66*(4), 547-556. doi:10.1139/f09-019

Wellborn, G. A., & Langerhans, R. B. (2015). Ecological opportunity and the adaptive diversification of lineages. *Ecology and Evolution, 5*(1), 176-195. doi:10.1002/ece3.1347

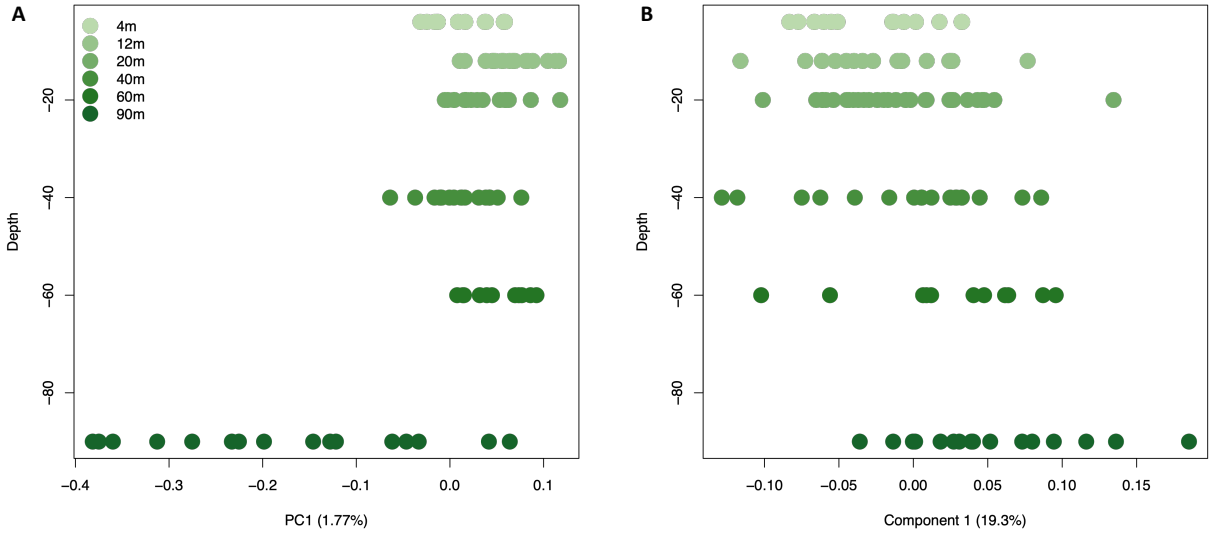
Woods, P. J., Müller, R., & Seehausen, O. (2009). Intergenomic epistasis causes asynchronous hatch times in whitefish hybrids, but only when parental ecotypes differ. *Journal of Evolutionary Biology, 22*(11), 2305-2319. doi:10.1111/j.1420-9101.2009.01846.x



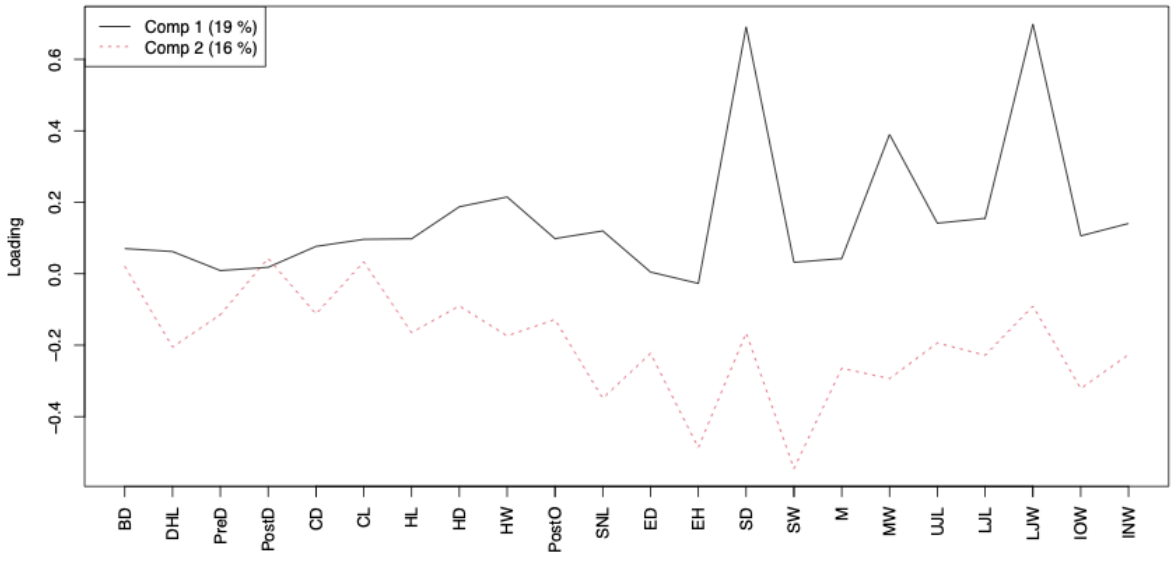
Supplementary Figure 1: PCA including all available Lake Constance samples. Grey is *C. gutturosus*, blue is *C. wartmanni,* green is *C. macrophthalmus* and orange is *C. arenicolus* (see legend to the right). Pre-eutrophication samples from Frei et al. (2022) are indicated with crosses, post-eutrophication samples are indicated with circles. Filled circles are samples caught in the depth transect sampling of this study, empty circles are from Frei et al. (2022).



Supplementary Figure 2: Structure analysis, using all depth transect samples and all samples used in Frei et al. (2022). Grey is *C. gutturosus*, blue is *C. wartmanni*, green is *C. macrophthalmus* and orange is *C. arenicolus*. Reference samples from Frei et al. (2022) are divided into pre-eutrophication (crosses) and post-eutrophication (empty circles) samples. Samples are grouped into the six spawning depth populations, and individuals that were assigned to *C. wartmanni* or *C. arenicolus* are shaded.



Supplementary Figure 3: A) Principal component 1 of the genomic PCA against depth including all caught *C. macrophthalmus* individuals. There was a significant correlation of PC1 with depth (rho=-0.37, p=0.0003). **B)** Component 1 of the partial least squares analysis against depth. There was a significant correlation of component 1 with depth (rho=-0.44, p=2e-6).



Supplementary Figure 4: Loadings of each trait of the partial least squares regression analysis (Figure 1D and Supplementary Figure 3B). All 22 morphometric traits and their loading on component 1 and 2 of the partial least squares regression analysis. The three traits with highest (positive) loadings on component 1 are LJW (lower jaw width), SD (snout depth) and MW (mouth width). The three traits with highest (negative) loadings on component 2 are SW (snout width), EH (eye height) and SNL (snout length).

Supplementary Table 1: D-statistic results for the test of introgression during eutrophication from *C. gutturosus* into each of the spawning depth populations. 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.

D	Z	P-value	P1	P2	P3	Outgroup
0.02	5.81	0.00	*C. macrophthalmus* pre (n=2)	4m	*C. gutturosus* pre (n=11)	*S. salar* (n=1)
0.01	4.78	0.00	*C. macrophthalmus* pre (n=2)	12m	*C. gutturosus* pre (n=11)	*S. salar* (n=1)
0.02	5.97	0.00	*C. macrophthalmus* pre (n=2)	20m	*C. gutturosus* pre (n=11)	*S. salar* (n=1)
0.02	6.31	0.00	*C. macrophthalmus* pre (n=2)	40m	*C. gutturosus* pre (n=11)	*S. salar* (n=1)
0.02	5.46	0.00	*C. macrophthalmus* pre (n=2)	60m	*C. gutturosus* pre (n=11)	*S. salar* (n=1)
0.01	4.19	0.00	*C. macrophthalmus* pre (n=2)	90m	*C. gutturosus* pre (n=11)	*S. salar* (n=1)

Supplementary Table 2: D-statistic results for the test of introgression during eutrophication from *C. wartmanni* into each of the spawning depth populations. 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.

D	Z	P-value	P1	P2	P3	Outgroup
0.01	2.82	0.00	*C. macrophthalmus* pre (n=2)	4m	*C. wartmanni* pre (n=2)	*S. salar* (n=1)
0.01	3.72	0.00	*C. macrophthalmus* pre (n=2)	12m	*C. wartmanni* pre (n=2)	*S. salar* (n=1)
0.01	4.05	0.00	*C. macrophthalmus* pre (n=2)	20m	*C. wartmanni* pre (n=2)	*S. salar* (n=1)
0.01	4.54	0.00	*C. macrophthalmus* pre (n=2)	40m	*C. wartmanni* pre (n=2)	*S. salar* (n=1)
0.01	3.72	0.00	*C. macrophthalmus* pre (n=2)	60m	*C. wartmanni* pre (n=2)	*S. salar* (n=1)
0.01	2.43	0.02	*C. macrophthalmus* pre (n=2)	90m	*C. wartmanni* pre (n=2)	*S. salar* (n=1)

Supplementary Table 3: D-statistic results for the test of introgression during eutrophication from *C. arenicolus* into each of the spawning depth populations. 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.

D	Z	P-value	P1	P2	P3	Outgroup
0.00	0.52	0.60	*C. macrophthalmus* pre (n=2)	4m	*C. arenicolus* pre (n=3)	*S. salar* (n=1)
0.00	-0.56	0.57	*C. macrophthalmus* pre (n=2)	12m	*C. arenicolus* pre (n=3)	*S. salar* (n=1)
0.00	-0.52	0.60	*C. macrophthalmus* pre (n=2)	20m	*C. arenicolus* pre (n=3)	*S. salar* (n=1)
0.00	-0.48	0.63	*C. macrophthalmus* pre (n=2)	40m	*C. arenicolus* pre (n=3)	*S. salar* (n=1)
0.00	-0.19	0.85	*C. macrophthalmus* pre (n=2)	60m	*C. arenicolus* pre (n=3)	*S. salar* (n=1)
0.00	-1.47	0.14	*C. macrophthalmus* pre (n=2)	90m	*C. arenicolus* pre (n=3)	*S. salar* (n=1)

Supplementary Table 4: Genes overlapping with any candidate SNP under selection between shallow and deep water (n=107), and their respective best Blast hits with an annotation of the genome of *Salmo salar*.

Gene name Alpine whitefish genome assembly (De-Kayne et al., 2020)	E value	Perc. identical	Accession	Description
snap_masked-PGA_scaffold11__203_contigs__length_63881516-processed-gene-283.13	2.00E-116	77.38	NP_001136192.1	RNA polymerase II subunit A C-terminal domain phosphatase SSU72
maker-PGA_scaffold11__203_contigs__length_63881516-snap-gene-418.12	0	89.89	XP_014023237.2	phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta-like
maker-PGA_scaffold11__203_contigs__length_63881516-snap-gene-480.10	0	85.07	XP_013987572.1	disks large-associated protein 4 isoform X1
maker-PGA_scaffold12__167_contigs__length_57740044-augustus-gene-463.9	3.00E-14	100	XP_045579496.1	gamma-aminobutyric acid receptor subunit beta-1 isoform X3
maker-PGA_scaffold14__173_contigs__length_55641933-snap-gene-414.19	0	97	XP_013997625.1	nucleoporin NUP188
maker-PGA_scaffold14__173_contigs__length_55641933-snap-gene-446.9	0	92.08	XP_013999495.2	lamin-B1
maker-PGA_scaffold15__168_contigs__length_54025139-augustus-gene-331.8	0	87.41	XP_045563145.1	kinesin-like protein KIF2A isoform X5
maker-PGA_scaffold15__168_contigs__length_54025139-snap-gene-398.18	0	85.17	XP_014026493.1	tenascin isoform X1
maker-PGA_scaffold17__183_contigs__length_51949489-snap-gene-114.25	0	87.47	XP_045562700.1	ARF GTPase-activating protein GIT2a isoform X2
maker-PGA_scaffold18__164_contigs__length_59907985-augustus-gene-424.0	0	87.9	XP_045544677.1	mucin-17
maker-PGA_scaffold19__147_contigs__length_54335267-snap-gene-386.8	0	91.1	XP_014004818.1	phosphatidylinositol 3-kinase regulatory subunit gamma-like isoform X1
maker-PGA_scaffold23__167_contigs__length_50329371-snap-gene-199.0	0	92.39	XP_014001415.2	laminin subunit beta-2 isoform X2
maker-PGA_scaffold23__167_contigs__length_50329371-snap-gene-487.17	1.00E-154	88.26	XP_045548802.1	FYVE, RhoGEF and PH domain-containing protein 1 isoform X1
maker-PGA_scaffold24__152_contigs__length_51033154-snap-gene-280.12	0	90.94	XP_014047632.1	rho-associated protein kinase 2 isoform X3
maker-PGA_scaffold24__152_contigs__length_51033154-snap-gene-396.16	0	85.99	XP_014051926.1	intersectin-2 isoform X1
maker-PGA_scaffold25__179_contigs__length_50922480-snap-gene-111.6	4.00E-156	87.74	XP_014025515.2	harmonin-like isoform X2
maker-PGA_scaffold26__192_contigs__length_48683376-snap-gene-73.17	0	87.65	XP_045545764.1	rap guanine nucleotide exchange factor 1 isoform X1
maker-PGA_scaffold26__192_contigs__length_48683376-snap-gene-229.16	0	91.17	XP_013984216.1	rab9 effector protein with kelch motifs
maker-PGA_scaffold26__192_contigs__length_48683376-snap-gene-381.32	0	72.26	XP_014051430.1	platelet-derived growth factor receptor beta
maker-PGA_scaffold28__172_contigs__length_48977775-snap-gene-133.10	2.00E-142	68.62	XP_014065728.2	consortin-like
maker-PGA_scaffold28__172_contigs__length_48977775-snap-gene-260.13	0	86.12	XP_014066184.1	protein jagged-2-like isoform X1
maker-PGA_scaffold29__157_contigs__length_48675208-snap-gene-407.12	0	84.5	XP_013991973.1	serine--tRNA ligase, cytoplasmic-like
maker-PGA_scaffold30__165_contigs__length_48446552-snap-gene-119.2	3.00E-73	72.61	XP_014034412.1	ras-related protein Rab-26

snap_masked-PGA_scaffold30__165_contigs__length_48446552-processed-gene-161.4	1.00E-36	92.31	XP_014034785.1	cGMP-dependent protein kinase 1 isoform X2
maker-PGA_scaffold30__165_contigs__length_48446552-snap-gene-177.15	2.00E-108	85.02	XP_014034899.2	mitochondrial import inner membrane translocase subunit Tim23
maker-PGA_scaffold33__143_contigs__length_40727438-augustus-gene-264.4	0.00E+00	94.94	XP_014014854.1	unnamed protein product
maker-PGA_scaffold38__206_contigs__length_33962415-snap-gene-84.0	3.00E-175	86.6	XP_014062081.1	EH domain-binding protein 1-like isoform X7
maker-PGA_scaffold38__206_contigs__length_33962415-snap-gene-286.7	0.00E+00	78.07	XP_014012100.1	phosphorylase b kinase regulatory subunit alpha, skeletal muscle isoform isoform X1
maker-PGA_scaffold4__243_contigs__length_45591172-snap-gene-365.20	0.00E+00	84.47	XP_045559993.1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase eta-1 isoform X3
maker-PGA_scaffold7__351_contigs__length_68138733-snap-gene-159.14	0.00E+00	98.25	XP_013998328.2	homeobox protein Dlx5a-like

Chapter III

III. Ecological disturbance reduces genomic diversity across an Alpine whitefish adaptive radiation

David Frei[1,2], Salome Mwaiko[1], 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

Genomic diversity is associated with the adaptive potential of a population and thereby impacts the extinction risk of a species during environmental change. We here used whole-genome resequencing data from all four species of the Lake Constance Alpine whitefish radiation covering a period of strong but transient anthropogenic environmental change and, to track changes in genomic diversity in all species over time. Genomic diversity became strongly reduced during the period of anthropogenic disturbance and has not recovered yet. The decrease in genomic diversity varies between 18-30%, depending on the species. Interspecific allele frequency differences of SNPs located in potentially ecologically relevant genes were homogenized over time. This suggests that in addition to the reduction of genome wide genetic variation, differentiation that evolved in the process of adaptation to alternative ecologies between species might have been lost during ecological disturbance. The erosion of substantial amounts of genomic variation within just a few generations in combination with the loss of potentially adaptive genomic differentiation, both of which had evolved over thousands of years, demonstrates the sensitivity of biodiversity in evolutionary young adaptive radiations towards environmental disturbance. Natural history collections, such as the one used for this study, are instrumental for the assessment of genomic consequences of anthropogenic environmental change. Historical samples enable us to document biodiversity loss against the shifting baseline syndrome and advance our understanding of the needs for efficient biodiversity conservation on a global scale.

Introduction

Genetic diversity represents the most fundamental level of biodiversity. Genomic diversity is central to sustain viable populations and to preserve evolutionary potential, enabling the adaptation to changing environmental conditions (Hoffmann, Sgro, & Kristensen, 2017). As a consequence, genomic diversity is one key component determining the extinction risk of a population during environmental change (Jensen & Leigh, 2022). Disturbance of ecosystems can influence genomic variation through both selective, but also demographic (and selectively neutral) processes, as well as the interaction of both (Banks et al., 2013). As a result, the history of environmental disturbance may be a major driver shaping patterns and dynamics of genomic diversity in many natural systems (Banks et al., 2013). As both frequency and strength of anthropogenic ecological disturbances are increasing (*IPBES*, 2019; Turner, 2010), it is essential to advance our understanding of how such disturbance affects biodiversity at its most basal level, which is genetic and/or genomic diversity (Banks et al., 2013).

Anthropogenic eutrophication during the last century had dramatic consequences on the biodiversity of many perialpine lakes in Switzerland (Feulner & Seehausen, 2019; Frei, De-Kayne, Selz, Seehausen, & Feulner, 2022; Vonlanthen et al., 2012). The effects on many species of the Alpine whitefish were particularly detrimental. In total, about a third of the more than 30 taxonomically described whitefish species went extinct during the period of anthropogenic eutrophication (Selz, Doenz, Vonlanthen, & Seehausen, 2020; Steinmann, 1950; Vonlanthen et al., 2012). High nutrient inputs altered many habitat characteristics of the deep and oligotrophic Swiss lakes, affecting both diet and reproduction of many whitefish species (Vonlanthen et al., 2012). The loss of spawning grounds together with the shift in food resources resulted in the extinction of multiple species through a combination of

demographic decline and speciation reversal through introgressive hybridization (Frei, De-Kayne, et al., 2022; Vonlanthen et al., 2012). The improvement of sewage treatment and phosphorus management towards the end of the last century resulted in many of the Swiss lakes returning close to their natural oligotrophic state (Vonlanthen et al., 2012). Even though the changed environmental conditions were transient and of relatively short duration, the period of cultural eutrophication had severe consequences on the genomic variation of the Alpine whitefish radiation (Frei, De-Kayne, et al., 2022).

In evolutionary young adaptive radiations, such as the Alpine whitefish radiation, sympatric species are still able to hybridize (Schluter, 2000; Seehausen, Takimoto, Roy, & Jokela, 2008). This is because complete reproductive isolation takes orders of magnitudes longer to evolve than the rapid speciation events in such young radiations (Schluter, 2009; Seehausen et al., 2008). The ability to exchange genomic variation might become particularly important during ecological disturbance: When environmental conditions rapidly change into an unfavorable state for a certain species of a young adaptive radiation, habitats can be lost and food resources might become unavailable, resulting in demographic decline. The decreasing population size is strengthening genetic drift, reducing genetic diversity in the declining population. In such a situation, the exchange of genomic variation with other members of the adaptive radiation through hybridization could become beneficial (Frei, Reichlin, Seehausen, & Feulner, 2022). Hybridization might increase genomic variation of the population and enhance its evolvability, increasing the likelihood of adaptation to the changed environmental conditions through evolutionary rescue (Gilman & Behm, 2011).

In order to document the effects of ecological disturbance on genomic variation following natural disturbance, genomic time-series data capturing the disturbance event is essential (Jensen & Leigh, 2022). The Lake Constance whitefish radiation was strongly affected by anthropogenic eutrophication during the last century. Using historical fish scale

samples, previous work demonstrated that all four taxonomically described whitefish species of Lake Constance extensively hybridized during the eutrophication period (Frei, De-Kayne, et al., 2022; Vonlanthen et al., 2012). One species went extinct by a combination of demographic decline and speciation reversal through introgressive hybridization during the period of anthropogenic eutrophication (Frei, De-Kayne, et al., 2022; Vonlanthen et al., 2012), and according to fisheries management, population sizes of all Lake Constance whitefish dramatically decreased over the last decades (Alexander & Seehausen, 2021). The potential to generate temporal whole-genome resequencing data spanning the entire eutrophication event and including four species makes the Lake Constance whitefish radiation an outstanding system to study the effects of ecological disturbance on genomic diversity.

Here, we used natural history collections to sequence population scale data (10-12 whole genomes per population) of each of the four Lake Constance whitefish species before the onset of the anthropogenic eutrophication (before 1950), as well as data from all three extant species during the peak eutrophication period (1970-1980). In combination with existing sequencing data from the three surviving species (n=8-12) collected after the eutrophication period ended and the lake returned to an oligotrophic state, we produced a time-series data set capturing the whole period (pre, during, and post) of anthropogenic eutrophication. During this period of anthropogenic ecological disturbance, we observed a strong decline in genomic diversity over time and found genomic signals of population declines in all species.

Methods

Sample collection

Historical whitefish scale samples previously used in Vonlanthen et al. (2012) and Frei, De-Kayne, et al. (2022) were used to extract DNA from twelve individuals of each population (pre- and during-eutrophication) of each species (see Table S1). These samples were collected from fisheries authorities around the lake during the last century and have been assembled by David Bittner (see Vonlanthen et al. (2012) for details). For the post-eutrophication populations, we used sequencing data (sampled 2015) produced by Frei, De-Kayne, et al. (2022) retrieved from ENA with accession PRJEB43605, as well as data from Frei, Reichlin, et al. (2022) (sampled 2019) retrieved from ENA with accession PRJEB53050 (see Table S1 for sample accessions).

DNA extraction and sequencing

DNA was extracted according to Frei, De-Kayne, et al. (2022). In brief, DNA extraction of historical scale samples was done using the Qiagen DNeasy blood and tissue kit (Qiagen AG, CH). For scale samples, we followed the manufacturer's protocol for crude lysates with minor adjustments (alternative lysis buffer from Wasko et al. (2003) containing 4M UREA and overnight incubation at 37°C).

Libraries were produced using the Accel-NGS 1S Plus DNA library kit (Swift Biosciences) at the NGS platform of the University of Bern. Libraries were then sequenced paired-end 100bp on a Novaseq 6000 S4 flowcell.

Data processing

We removed poly-G strings with fastp (Chen, Zhou, Chen, & Gu, 2018) and then merged overlapping read pairs (with overlaps longer than 25 bp) using SeqPrep 1.0. (https://github.com/jstjohn/SeqPrep). The processed reads were then aligned to the Alpine whitefish reference genome (De-Kayne, Zoller, & Feulner, 2020) with bwa mem version 0.7.12 (Li & Durbin, 2009) and adjusting the "r" parameter to 1. Finally, we used picard-tools (Version 2.20.2; http://broadinstitute.github.io/picard/) to mark duplicate reads (MarkDuplicates), fix mate information (FixMateInformation) and we replaced read groups with (AddOrReplaceReadGroups).

Population genomic analysis

Genotype likelihoods at polymorphic sites were calculated using angsd 0.925 (Korneliussen, Albrechtsen, & Nielsen, 2014), using the samtools genotype likelihood model. For that purpose, we excluded reads with a mapping quality below 30, bases with base qualities below 20 and reads that did not map uniquely to the reference. Only sites passing a p-value cut-off of 10E-6 for being variable, with a sequencing depth above 2x in each individual and with data of at least 80 of all 127 individuals were included. Additionally, we only analysed whitefish chromosomes without any potentially collapsed duplicated regions (De-Kayne et al., 2020) to avoid potential bias. We applied SNP filters to avoid strand bias (-sb_pval 0.05), quality score bias (-qscore_pval 0.05), edge bias (-edge_pval 0.05) and mapping quality bias (-mapq_pval 0.05). This resulted in a total of 355'311 polymorphic sites for further analysis.

To verify the species assignment done in the field when these samples have been collected by fisheries authorities, we performed a PCA using PCAngsd 1.02 (Meisner & Albrechtsen, 2018). We excluded sites with a minor allele frequency below 0.05 (across the

whole dataset), resulting in 128‘164 sites. Default parameters were used, except for using the first three eigenvectors to estimate individual allele frequencies (-e 3). By this PCA approach, we identified 12 individuals suggesting an erroneous species assignment which were excluded from all subsequent analyses (one post-eutrophication *C. wartmanni* was genetically assigned to *C. arenicolus*, and another one to *C. macrophthalmus*; one post-eutrophication and six during-eutrophication *C. macrophthalmus* samples were genetically assigned to *C. wartmanni*, one during-eutrophication *C. macrophthalmus* was assigned to *C. arenicolus* and two pre-eutrophication *C. macrophthalmus* were assigned to *C. gutturosus*; see Figure S1 and Table S1).

Based on the genotype likelihoods inferred in all our 127 samples but excluding individuals identified as potentially misidentified in the field, we calculated Watterson’s theta (θ_ω) and Tajima’s D in 100 kb windows along the genome (Korneliussen, Moltke, Albrechtsen, & Nielsen, 2013). To do this, the folded site allele frequency likelihood for each species and each sampling timepoint separately as well as the folded site allele frequency likelihood for all species pooled together at each sampling timepoint was calculated in angsd (0.925) (Korneliussen et al., 2014; Nielsen, Korneliussen, Albrechtsen, Li, & Wang, 2012). The maximum likelihood estimate of the folded site allele frequency spectrum was inferred using realSFS of angsd (0.925) (Korneliussen et al., 2014). With the global site allele frequency spectrum, we calculated different theta estimators and Tajima’s D in 100 kb windows using thetaStat of angsd (0.925) (Korneliussen et al., 2014; Korneliussen et al., 2013). We used all 100 kb windows to calculate a genome wide average.

We used NgsRelate v2 (Hanghøj, Moltke, Andersen, Manica, & Korneliussen, 2019; Korneliussen & Moltke, 2015) to calculate pairwise relatedness between all individuals of each species at each sampled timepoint. We split the genotype likelihood file generated across all species and timepoints into each single species and timepoints, and used these separate

genotype likelihood fields as input to NgsRelate v2 (Hanghøj et al., 2019; Korneliussen & Moltke, 2015). At each polymorphic site in the genotype likelihood file, we calculated the allele frequency in each species at each timepoint in angsd 0.925 (Korneliussen et al., 2014) using the method from Kim et al. (Kim et al., 2011), and also used this allele frequency information as input to NgsRelate v2 (Hanghøj et al., 2019; Korneliussen & Moltke, 2015), which we then used to calculate pairwise relatedness with default parameters. We finally calculated the mean relatedness in each species and timepoint by averaging across all pairwise relatedness values for each population in R (R Core Team, 2018).

For each of the 355'311 polymorphic site, we calculated the weighted F_{ST} between each species and all other species pooled together (of only the pre-eutrophication populations) in angsd 0.925 (Korneliussen et al., 2014) from one- and two-dimensional site frequency spectra which were inferred from site allele frequencies (Nielsen et al., 2012). The sites with the highest resulting F_{ST} values are most characteristic for the respective species, and thus, might be involved in the adaptation to its habitat. At the ten sites with the highest F_{ST}, we then calculated the allele frequency in each species and at each timepoint in angsd 0.925 (Korneliussen et al., 2014) after the method from Kim et al. (Kim et al., 2011) to track the change in allele frequency differences over time. Additionally, we calculated the allele frequencies in each species and at each timepoint for the SNP (position 30197713 on scaffold 23) within the gene *edar* that has been found to be significantly associated with gill-raker count (De-Kayne et al., 2022), a trait that is relevant for the feeding ecology of each species. We further blasted the protein sequence of the gene *vgll3*, which is known to be relevant for age at maturity in *Salmo salar* (Barson et al., 2015), against the Alpine whitefish genome and found two equivalent best hits. For any SNPs in these two genes (likely paralogous copies of *vgll3*), we as well calculated the allele frequencies in each species and each timepoint to document the change in allele frequencies over time.

Results

We used natural history collections to sequence population genomic time series data, including an entire adaptive radiation and capturing a period of transient but severe ecological disturbance with the aim of documenting the influence of ecological disturbance on the genomic diversity of each single species, but also on the entire adaptive radiation.

Population structure

We performed a PCA based on genotype likelihoods of 128‘164 SNP’s to visualize population structure of the Lake Constance whitefish radiation over time, respectively over the period of anthropogenic eutrophication. Apart from the extinction of *C. gutturosus*, the three extant species cluster closer together post-eutrophication compared to pre-eutrophication, suggesting that the species are today less differentiated than before the onset of eutrophication. This might be the consequence of interspecific hybridization during the eutrophication period as demonstrated in previous work (Frei, De-Kayne, et al., 2022; Frei, Reichlin, et al., 2022; Vonlanthen et al., 2012), also consistent with several potential early generation hybrids in the during-eutrophication sampling time-point.

Nucleotide diversity

Irrespective of species, nucleotide diversity (measured as Watterson’s theta) declined over time (Figure 1B). In each species, nucleotide diversity was highest before the onset of anthropogenic eutrophication, and it was lowest post-eutrophication, while the populations sampled during peak eutrophication indicated values of nucleotide diversity between the pre- and post-eutrophication population of the species. In total, *C. wartmanni* lost ~23%, *C.*

arenicolus lost ~28%, and *C. macrophthalmus* lost ~30% of their original nucleotide diversity from before the onset of the eutrophication period.

Tajima's D

The genome-wide average of Tajima's D of each species was negative before and during the period of anthropogenic eutrophication (Figure 1C), indicative of population expansion after a recent bottleneck. This might reflect the recent colonization and evolution of the radiation within Lake Constance, since the last glacial maximum 10'000-15'000 years ago. However, in each species, Tajima's D was positive (D = 0.45; Figure 1C) after the period of anthropogenic eutrophication ended, potentially indicating a sudden population contraction associated with the altered environmental conditions.

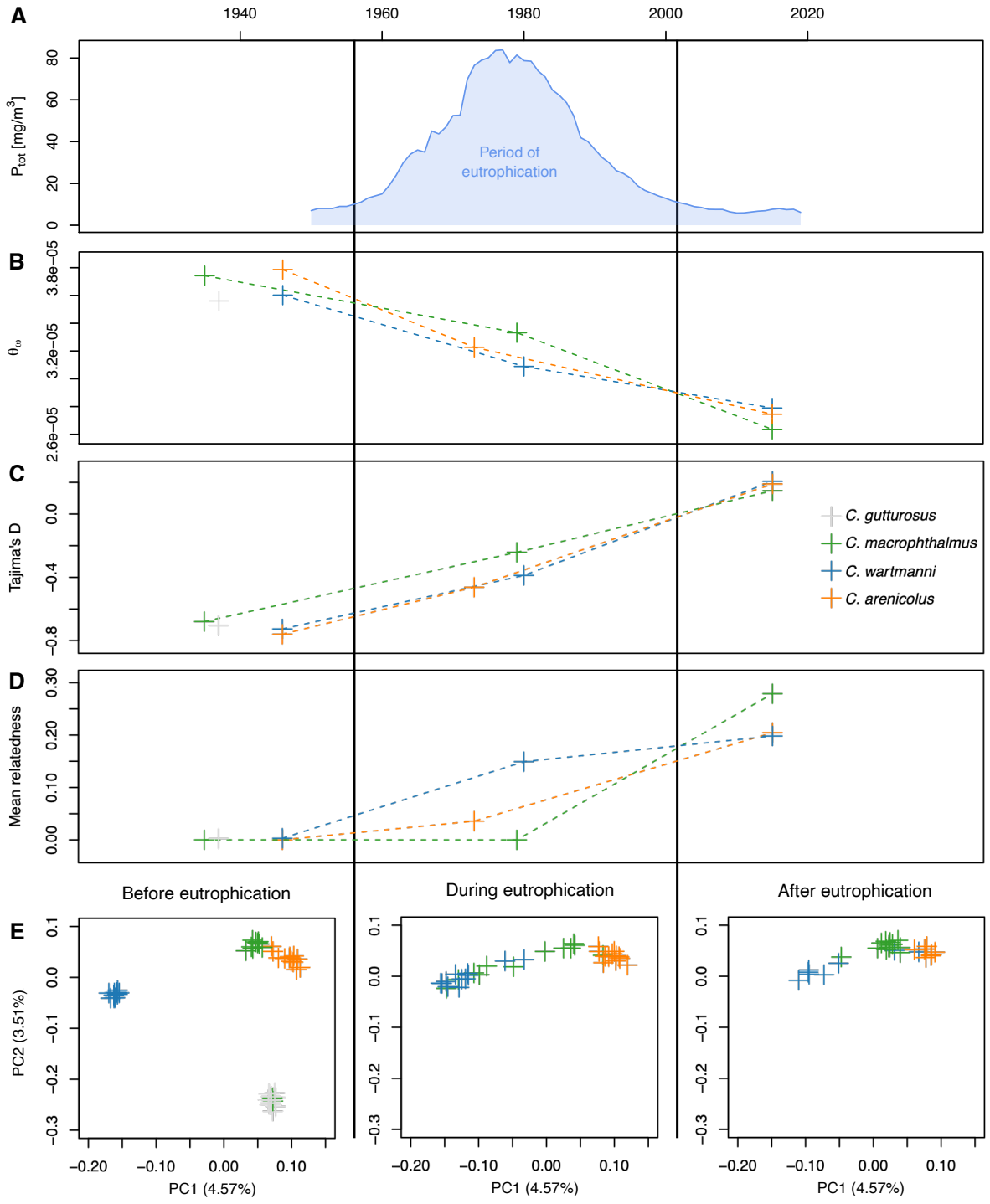


Figure 1: Total phosphate, nucleotide diversity, Tajima's D and relatedness over time. Total phosphate concentration over time (**A**), Watterson's theta (**B**) and Tajima's D (**C**), mean relatedness between individuals of each species and timepoint (**D**), and PCA based on genotype likelihoods separately plotted for each timepoint (**E**). **E)** PCA showing the population structure of the Lake Constance whitefish radiation over time. These plots show the PCA from Supplementary Figure 1, subsetted to the three sampling time-points. Colors correspond to species (grey *C. gutturosus*, green *C. macrophthalmus*, blue *C. wartmanni* and orange *C. arenicolus*), see legend in panel C.

Mean relatedness

For each extant species of Lake Constance whitefish, we calculated the pairwise relatedness between all individual before, during and after the period of anthropogenic eutrophication (Figure S2). Consistent with a decrease in population size, mean relatedness increased in all species over the period of eutrophication. All species showed mean relatedness values below 0.01 before the start of the eutrophication period, while mean relatedness ranged between ~0.2 and ~0.28 in the post eutrophication population of the three extant species (Figure 1D).

Frequency shifts over time

We identified the ten most characteristic alleles of each species by calculating the F_{ST} between each species and all other species pooled into one population at all SNPs along the genome, and for each pairwise comparison used the ten positions with the highest F_{ST} values. In line with previous work that identified few fixed differences between sympatric species of the Alpine whitefish radiation (De-Kayne et al., 2022), we did not detect any fixed differences ($F_{ST} > 0.95$). However, all the sites identified as characteristics for each species had high F_{ST}'s (F_{ST} = 0.82-0.8 for *C. gutturosus*, F_{ST} = 0.73-0.62 for *C. arenicolus*, F_{ST} = 0.92-0.58 for *C. macrophthalmus* and 0.95-0.91 for *C. wartmanni*). In all three extant species, we observed almost an identical pattern at all of the ten most characteristic sites: Allele frequency differences between species were homogenized over the period of eutrophication, because the frequency of the predominant allele in the focal species decreased, while its frequency in all other species increased over time (Figure 2a-c).

We also assessed the allele frequency change in two genes which affect ecologically relevant phenotypes in Salmonids. At a locus in the *edar* gene involved in determining the gill-raker count of whitefish species from De-Kayne et al. (2022), the pre-eutrophication

samples of the species with a low gill-raker count indicated very high allele frequencies (*C. gutturosus* 0.99 and *C. arenicolus* 0.66) while the species with a higher gill-raker count (both *C. wartmanni and C. macrophthalmus* had frequency of 0) showed a very low allele frequency (Figure S3). After the eutrophication period ended, the differences between the surviving species became smaller (*C. macrophthalmus* increased from 0 to 0.18, *C. wartmanni* increased from 0.12, and *C. arenicolus* decreased from 0.66 to 0.43). We detected six polymorphic loci within two *vgll3* paralogs, a gene that is known to be involved in the age at maturity in *S. salar* (Czorlich, Aykanat, Erkinaro, Orell, & Primmer, 2018). For five of this six loci, the allele frequencies varied only little between species and were very low (minor allele frequency below 0.15 across all species and time points). However, one SNP indicated a pattern where frequencies were differentiated between the species before eutrophication, but differentiation was completely lost after eutrophication (Figure S4).

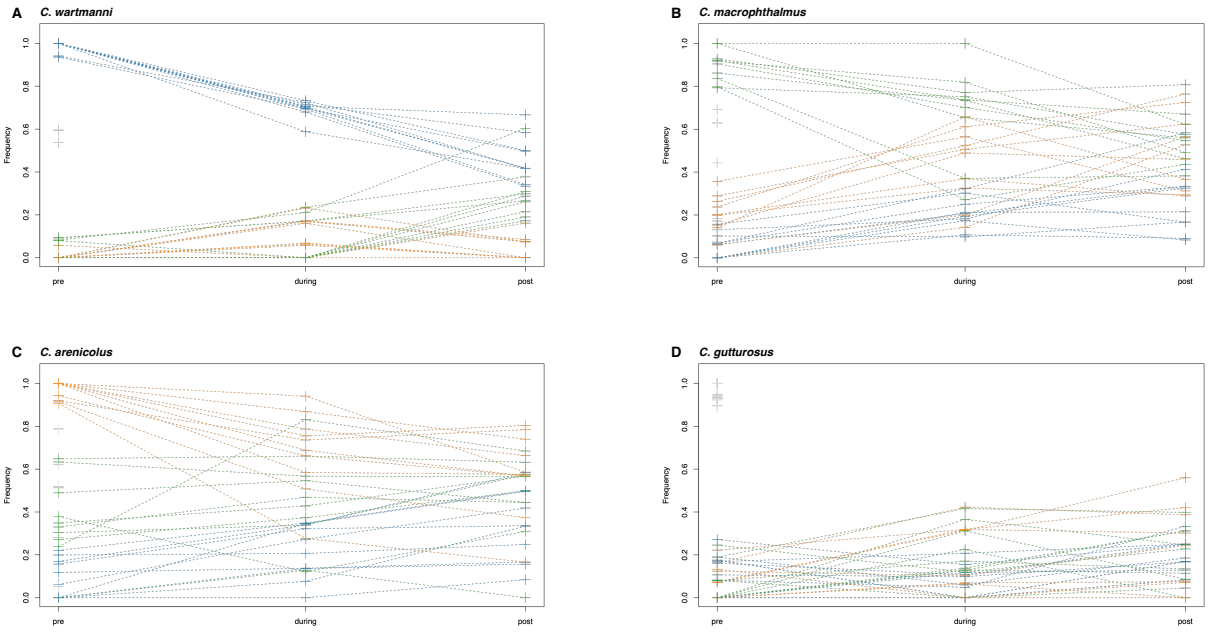


Figure 2: Allele frequency trajectories at most characteristic sites over time. A) The trajectories of the ten most characteristic sites of *C. wartmanni* (top ten highest F_{ST} values when comparing *C. wartmanni* with all other specie of the radiation at the time point before eutrophication). The allele frequencies of each position at the different sampling time-points are connected with a dashed line. The plot shows the allele frequencies of all ten most characteristic sites of C. wartmanni in all three species over time. The color corresponds to species (blue *C. wartmanni*, green *C. macrophthalmus*, orange *C. arenicolus*). **B)** The trajectories of the ten most characteristic sites of *C. macrophthalmus.* **C)** The trajectories of the ten most characteristic sites of *C. arenicolus.* **D)** The trajectories of the ten most characteristic sites of the extinct *C. gutturosus*.

Discussion

Genetic diversity is a core component of the adaptive potential of a population. As a result, the maintenance of genetic diversity is fundamental for fast adaptive responses to rapid environmental change. Hence, genetic diversity is a key component of the extinction risk of species during environmental change (Jensen & Leigh, 2022) and used as a metric to monitor threatened populations (Hoban et al., 2022). Anthropogenic ecological disturbance has the prospect to decrease genetic diversity in natural populations (Themudo et al., 2020). Here, we generated population level whole-genome resequencing data of all extant species of an adaptive radiation before, during and after a severe but transient period of anthropogenic eutrophication. We tracked genomic diversity through time and found that genomic diversity was reduced after the period of eutrophication. Based on our results, we discuss the implications of reduced genetic diversity for the adaptive potential and extinction risk, as well as the relevance of such data for management and conservation.

Population-decline during anthropogenic eutrophication

We observed substantial losses of genetic diversity in each species of the radiation over time, presumably in response to anthropogenic ecological disturbance of the ecosystem. In parallel, we observed a shift in Tajima's D from negative to positive in all species, indicative of a population decline. In line with a decreasing population size, relatedness within each species increased over the period of eutrophication. These results suggest that the period of anthropogenic eutrophication resulted in demographic decline of all three species, probably due to habitat loss and shifts in the available prey community (see Vonlanthen et al. 2012). This reduction in population sizes might potentially strengthened genetic drift, and eventually result in the loss of substantial amounts of genomic variation in a very short period

of time (~30 years or ~6 whitefish generations (Nussle, Bornand, & Wedekind, 2009)). Theory predicts that when genomic diversity is lost in linear fashion, population sizes are still relatively large (Jensen & Leigh, 2022). When population sizes are declining fast, many rare alleles are lost rapidly, leading to strong initial loss of genomic diversity. After this initial loss of rare alleles, all remaining alleles become common and thus, it becomes harder to lose further variation via drift, slowing the loss of genomic diversity down. Hence, when populations sizes become small, the loss of genomic diversity is L-shaped or exponential (Jensen & Leigh, 2022). We here observed rather linear losses of genomic diversity in all three studied species over time. This suggests that despite the population sizes declined during the period of eutrophication, each species might have still retained a relatively large population size.

Evolutionary rescue enabled by hybridization

Previous work showed that the all Lake Constance whitefish species extensively hybridized during the period of anthropogenic eutrophication, while there was little to no selection against genomic variation derived from hybridization (Frei, De-Kayne, et al. 2022). Even though hybridization might not have been able to counteract the loss of genomic diversity caused by the population declines during eutrophication, it could have helped to maintain diversity in each species and to overcome the negative effects of reduced genomic variation. Through such a scenario of evolutionary rescue, interspecific hybridization might become adaptive when populations rapidly decline (Stelkens, Brockhurst, Hurst, & Greig, 2014; Vedder et al., 2022). This suggests that when environments rapidly change, the risk of extinction of species that have evolved complete reproductive isolation against all other sympatric species might be higher than that of species that are still able to hybridize with one

or several other species. Hence, the evolution of complete reproductive isolation of a species can become obstructive for the survival of the species under rapid environmental change.

Frequency differences homogenized in all species

We identified the 10 historically most characteristic SNPs of each species (10 highest F_{ST} values between the pre-eutrophication population of each species and the pre-eutrophication populations of all other species pooled together). These positions have been chosen to reflect some of the adaptation to each species' habitat that have evolved in the course of evolution (da Silva Riberio, Galván , & Pool, 2022). When we compared the allele frequencies at these sites before, during and after eutrophication, we find that the frequency differences become homogenized over the period of eutrophication. Often, the frequency of the predominant allele in the focal species decreased, while its frequency in all other species increased over time , suggesting that the homogenization of allele frequency differences happened in all species. Two ecologically relevant genes, the *edar* locus associated with gill-raker count from De-Kayne et al. (2022) and paralogs of *vgll3* which is involved in the determination of age at maturity in Atlantic salmon (Czorlich et al., 2018), indicated the same pattern of homogenized interspecific allele frequency differences over the period of anthropogenic eutrophication. The loss of frequency differences at ecologically relevant loci is in line with phenotypic data showing reduced ranges of gill-raker numbers after eutrophication (Vonlanthen et al., 2012). Furthermore, the homogenization of allele frequencies differences at ecologically relevant alleles is consistent with extensive hybridization during the period of anthropogenic eutrophication, presumably in combination with reduced divergent selection between species. Hence, parts of the original species differentiation that might have evolved in response to adaptation to the selective pressure in

the habitats of each species have potentially been lost as consequence of anthropogenic eutrophication.

Whole-genome resequencing conflicts with microsatellite results

While Vonlanthen et al. (2012) describe an increase in allelic richness of nine microsatellite markers over the period of eutrophication, using samples from the same historical scale collection as we used for this study, we here report substantial losses of genetic diversity based on whole-genome resequencing data. This difference might be explained by the different data types. Microsatellites are multiallelic markers with a high mutation rate, resulting in fixed differences and private alleles between species or populations in a relatively short time span. However, in evolutionary young adaptive radiations, such as the Alpine whitefish radiation, species differentiation is mainly based on frequency shifts of thousands of single SNPs. As a result of this difference between microsatellite and SNP markers, demographic- (such as population declines) or evolutionary processes (such as hybridization) can have different outcomes on diversity estimates based on the two different marker types. Hybridization between two species whose differentiation is based on moderate frequency shifts at many SNPs might have little impact on their nucleotide diversity, while allelic richness at microsatellites is greatly increased because new alleles were brought into the species that were beforehand private to the other species. If such a hybridization event is taking place during a period of weak population decline, nucleotide diversity at all SNPs might decrease (from demographic decline while hybridization has little), whilst allelic richness at microsatellite loci is increasing (because hybridization is bringing in more new alleles than are lost due to drift during population decline). The contrasting results between microsatellite loci, still often used in the context of conservation and management of natural populations, and whole-genome resequencing data, highlight the importance of marker choice

to draw valid and robust conclusions for the question of interest. Considering the unprecedented contemporary rates of habitat loss and species extinctions, mitigating the consequences of genome-wide losses of genetic variation is central for overcoming the current biodiversity crisis (Kardos et al., 2021) and only possible by making use of genomic data for conservation purposes (Supple & Shapiro, 2018).

The relevance of genomic long-term data for biodiversity conservation

Natural population need genomic diversity to maintain the evolutionary potential enabling a rapid evolutionary response to changing environments (Hoffmann et al., 2017). The erosion of a substantial amounts of genomic variation within few generations in combination with the loss of potentially adaptive genomic differentiation, both evolved over thousands of years, demonstrates the sensitivity of evolutionary young adaptive radiations to environmental disturbance. Genetic erosion might have reduced the potential for resilience to future environmental change through a reduced evolutionary potential, increasing the extinction risk of each species. Therefore, characterizing the genomic change of natural populations across periods of ecological disturbance is fundamental to enhance species conservation and advance our understanding of biodiversity dynamics.

Understanding the genomic consequences of environmental change and its temporal dynamics heavily relies on a suitable baseline from before the onset of environmental disturbance (Jensen & Leigh, 2022). Increasing levels of ecosystem degradation on a global scale can result in lowered subjective threshold for acceptable environmental conditions. Without any historical records about the original condition of a given environment, new generations might consider the situation in which they have been raised as the appropriate baseline level (Soga & Gaston, 2018), a phenomenon termed the “shifting baselines syndrome” (Pauly, 1995). Natural history collections, such as the one used here, can provide

suitable baselines unaffected by anthropogenic influences and are therefore fundamental to counteract the shifting baselines syndrome. However, although it contributes disproportionally to biodiversity conservation and policy, the investment in generating long-term data is declining (Hughes et al., 2017). Thus, the generation of genomic data from historical data representing an appropriate baseline can fundamentally improve our understanding of the evolutionary response of natural populations to anthropogenic disturbance and thereby advance the establishment of targeted and efficient conservation measures.

Acknowledgements

We thank David Bittner for compiling the historical whitefish scale collection, Michael Kugler from the Amt für Natur, Jagd und Fischerei of the Kanton 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. 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.

Data accessibility

The raw sequencing files will become accessible on ENA SRA (PRJEB57733) upon publication. Additional supporting data is deposited on the eawag research data institutional collection and will become accessible upon publication. Scripts used for data analysis will become available on GitHub (https://github.com/freidavid/lake_constance_time_series_data).

Author contributions

DF, OS and PGDF conceived of, designed, and conceptualized the study. PGDF managed and supervised the study. SM performed DNA extractions and molecular lab work.

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.

References

Alexander, T., & Seehausen, O. (2021). *Diversity, distribution and community composition of fish in perialpine lakes - "Projet Lac" synthesis report*. Bern, Switzerland: Bundesamt für Umwelt BAFU.

Banks, S. C., Cary, G. J., Smith, A. L., Davies, I. D., Driscoll, D. A., Gill, A. M., . . . Peakall, R. (2013). How does ecological disturbance influence genetic diversity? *Trends in Ecology & Evolution, 28*(11), 670-679. doi:10.1016/j.tree.2013.08.005

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., . . . Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature, 528*(7582), 405-408. doi:10.1038/nature16062

Chen, S. F., Zhou, Y. Q., Chen, Y. R., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), 884-890. doi:10.1093/bioinformatics/bty560

Czorlich, Y., Aykanat, T., Erkinaro, J., Orell, P., & Primmer, C. R. (2018). Rapid sex-specific evolution of age at maturity is shaped by genetic architecture in Atlantic salmon. *Nature Ecology & Evolution, 2,* 1800-1807. doi:10.1038/s41559-018-0681-5

da Silva Riberio, T., Galván , J. A., & Pool, J. E. (2022). Maximum SNP F_{ST} outperforms full-window statistics for detecting soft sweeps in local adaptation. *Genome Biology and Evolution, 14*(10), 1-20. doi:10.1093/gbe/evac143

De-Kayne, R., Selz, O. M., Marques, D. A., Frei, D., Seehausen, O., & Feulner, P. G. D. (2022). Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nature Communications, 13*(1). doi:10.1038/s41467-022-32181-8

De-Kayne, R., Zoller, S., & Feulner, P. G. D. (2020). A de novo chromosome-level genome assembly of Coregonus sp. "Balchen": One representative of the Swiss Alpine whitefish radiation. *Molecular Ecology Resources, 20*(4), 1093-1109. doi:10.1111/1755-0998.13187

Feulner, P. G. D., & Seehausen, O. (2019). Genomic insights into the vulnerability of sympatric whitefish species flocks. *Molecular Ecology, 28*(3), 615-629. doi:10.1111/mec.14977

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., . . . European Reference Genome Atlas, E. R. (2022). The era of reference genomes in conservation genomics. *Trends in Ecology & Evolution, 37*(3), 197-202. doi:10.1016/j.tree.2021.11.008

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*, *6*(4), 461-468. doi:10.1038/s41559-022-01665-7

Frei, D., Reichlin, P., Seehausen, O., & Feulner, P. G. D. (2022). Introgression from extinct species facilitates adaptation to its vacated niche. *Molecular Ecology*, *32*(4), 841-853. doi:10.1111/mec.16791

Gilman, R. T., & Behm, J. E. (2011). Hybridization, species collapse, and species reemergence after disturbance to premating mechanisms of reproductive isolation. *Evolution, 65*(9), 2592-2605. doi:10.1111/j.1558-5646.2011.01320.x

Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A., & Korneliussen, T. S. (2019). Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *GigaScience, 8*(5). doi:https://doi.org/10.1093/gigascience/giz034

Hoban, S., Archer, F. I., Bertola, L. D., Bragg, J. G., Breed, M. F., Bruford, M. W., . . . Hunter, M. E. (2022). Global genetic diversity status and trends: towards a suite of essential biodiversity variables (EBVs) for genetic composition. *Biological Reviews, 97*(4), 1511-1538. doi:10.1111/brv.12852

Hoffmann, A. A., Sgro, C. M., & Kristensen, T. N. (2017). Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution, 32*(7), 506-517. doi:10.1016/j.tree.2017.03.012

Hughes, B. B., Beas-Luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., . . . Carr, M. H. (2017). Long-term studies contribute disproportionately to ecology and policy. *Bioscience, 67*(3), 271-281. doi:10.1093/biosci/biw185

IPBES (2019): Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (2019). (E. S. Brondizio, J. Settele, S. Díaz, & H. T. Ngo Eds.). Bonn, Germany: IPBES secretariat.

Jensen, E. L., & Leigh, D. M. (2022). Using temporal genomics to understand contemporary climate change responses in wildlife. *Ecology and Evolution, 12*(9). doi:https://doi.org/10.1002/ece3.9340

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., . . . Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences of the United States of America, 118*(48). doi:10.1073/pnas.2104642118

Kim, S. Y., Lohmueller, K. E., Albrechtsen, A., Li, Y. R., Korneliussen, T., Tian, G., . . . Nielsen, R. (2011). Estimation of allele frequency and association mapping using next-generation sequencing data. *Bmc Bioinformatics, 12*, 16. doi:10.1186/1471-2105-12-231

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *Bmc Bioinformatics, 15*, 13. doi:10.1186/s12859-014-0356-4

Korneliussen, T. S., & Moltke, I. (2015). NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics, 31*(24), 4009-4011. doi:https://doi.org/10.1093/bioinformatics/btv509

Korneliussen, T. S., Moltke, I., Albrechtsen, A., & Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *Bmc Bioinformatics, 14*, 14. doi:10.1186/1471-2105-14-289

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics, 210*(2), 719-731. doi:10.1534/genetics.118.301336

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. R., & Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *Plos One, 7*(7), 10. doi:10.1371/journal.pone.037558

Nussle, S., Bornand, C. N., & Wedekind, C. (2009). Fishery-induced selection on an Alpine whitefish: quantifying genetic and environmental effects on individual growth rate. *Evolutionary Applications, 2*(2), 200-208. doi:10.1111/j.1752-4571.2008.00054.x

Pauly, D. (1995). Anecdotes and the shifting base-line syndrome of fisheries. *Trends in Ecology & Evolution, 10*(10), 430-430. doi:10.1016/s0169-5347(00)89171-5

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford: Oxford University Press.

Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science, 323*(5915), 737-741. doi:10.1126/science.1160006

Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology, 17*(1), 30-44. doi:10.1111/j.1365-294X.2007.03529.x

Selz, O. M., Doenz, C. J., Vonlanthen, P., & Seehausen, O. (2020). 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

Soga, M., & Gaston, K. J. (2018). Shifting baseline syndrome: causes, consequences, and implications. *Frontiers in Ecology and the Environment, 16*(4), 222-230. doi:10.1002/fee.1794

Steinmann, P. (1950). Monographie der schweizerischen Koregonen. Beitrag zum Problem der Entstehung neuer Arten. Spezieller Teil. *Schweiz. Z. Hydrologie, 12*(1), 340-491. doi:10.1007/BF02486030

Stelkens, R. B., Brockhurst, M. A., Hurst, G. D. D., & Greig, D. (2014). Hybridization facilitates evolutionary rescue. *Evolutionary Applications, 7*(10), 1209-1217. doi:10.1111/eva.12214

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology, 19*. doi:10.1186/s13059-018-1520-3

Themudo, G. E., Rey-Iglesia, A., Tascon, L. R., Jensen, A. B., da Fonseca, R. R., & Campos, P. F. (2020). Declining genetic diversity of European honeybees along the twentieth century. *Scientific Reports, 10*(1). doi:10.1038/s41598-020-67370-2

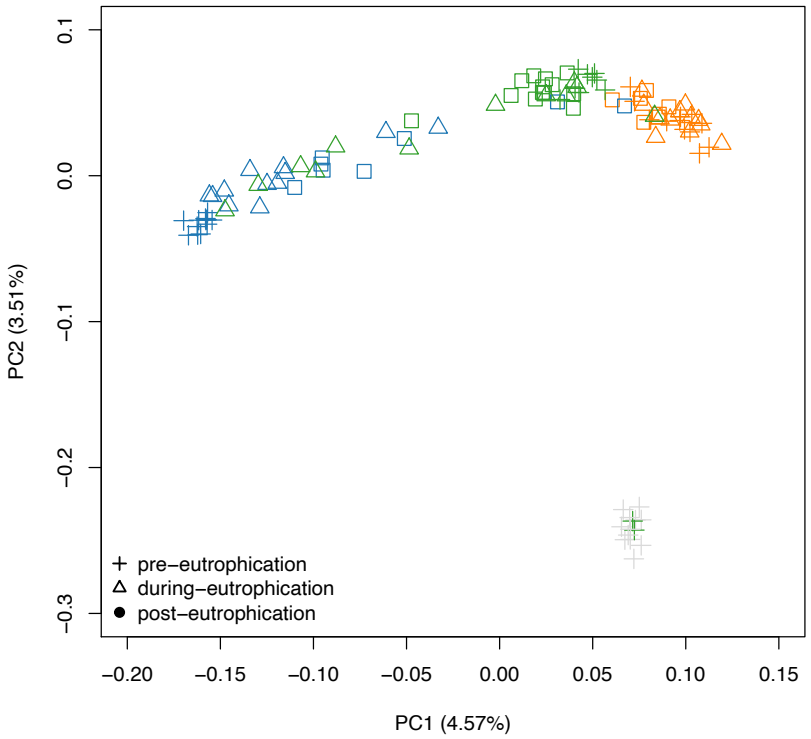
Turner, M. G. (2010). Disturbance and landscape dynamics in a changing world. *Ecology, 91*(10), 2833-2849. doi:10.1890/10-0097.1

Vedder, D., Lens, L., Martin, C. A., Pellikka, P., Adhikari, H., Heiskanen, J., . . . Cabral, J. S. (2022). Hybridization may aid evolutionary rescue of an endangered East African passerine. *Evolutionary Applications, 15*(7), 1177-1188. doi:10.1111/eva.13440

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Muller, R., Lundsgaard-Hansen, B., . . . Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature, 482*(7385), 357-U1500. doi:10.1038/nature10824

Wasko, A. P., Martins, C., Oliveira, C., & Foresti, F. (2003). Non-destructive genetic sampling in fish. An improved method for DNA extraction from fish fins and scales. *Hereditas, 138*(3), 161-165. doi:10.1034/j.1601-5223.2003.01503.x

Supplementary Figures



Supplementary Figure S1: Principal component analysis including all samples and timpepoints. The same PCA is plotted in Figure 1E, but pre-, during and post-eutrophication samples are split into different panels.

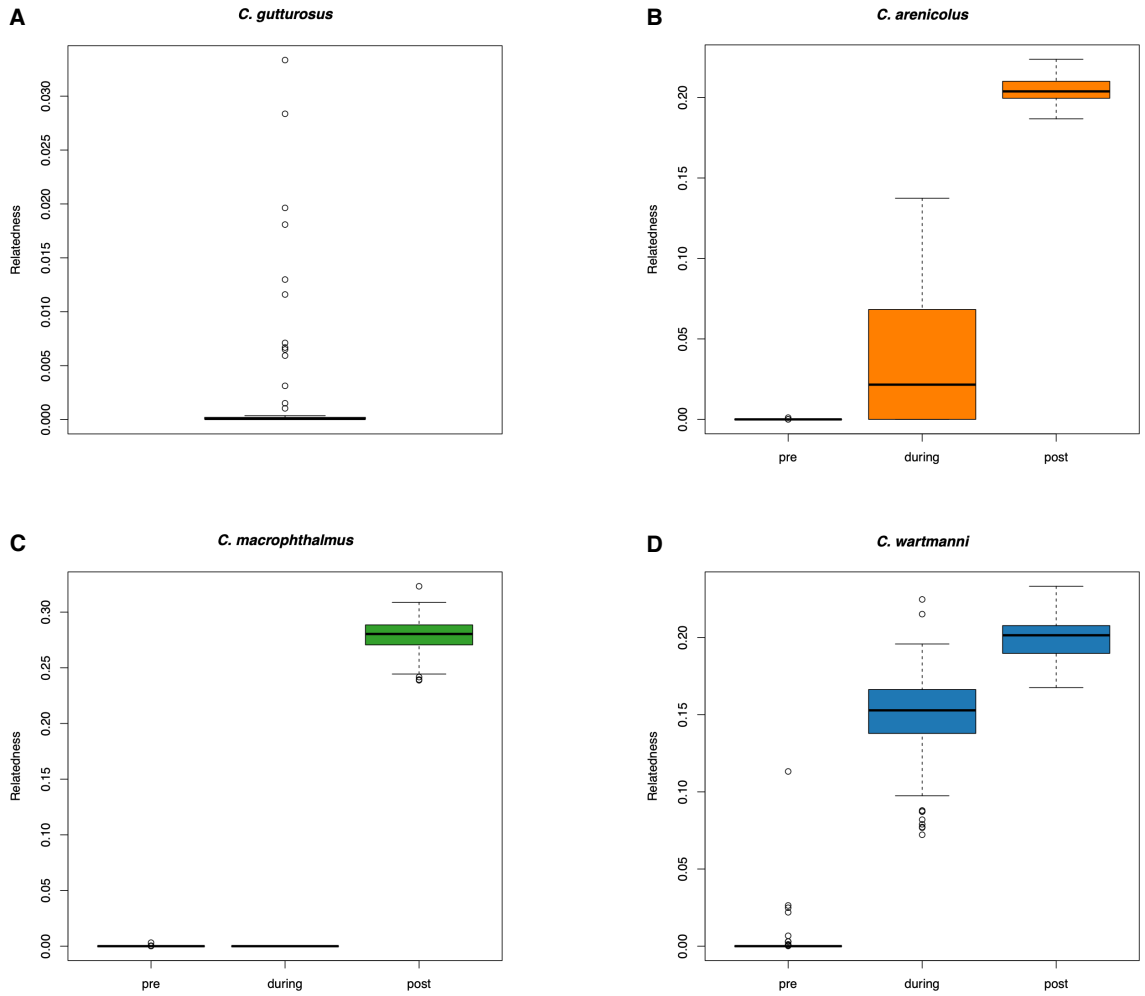


Figure S2: Increased relatedness after the eutrophication period. A) Distribution of pairwise relatedness between the extinct *C. gutturosus* individuals sequenced. **B)** Distribution of pairwise relatedness between *C. arenicolus* individuals from before, during and after the anthropogenic eutrophication period. **C)** Distribution of pairwise relatedness between *C. macrophthalmus* individuals from before, during and after the anthropogenic eutrophication period. **D)** Distribution of pairwise relatedness between *C. wartmanni* individuals from before, during and after the anthropogenic eutrophication period.

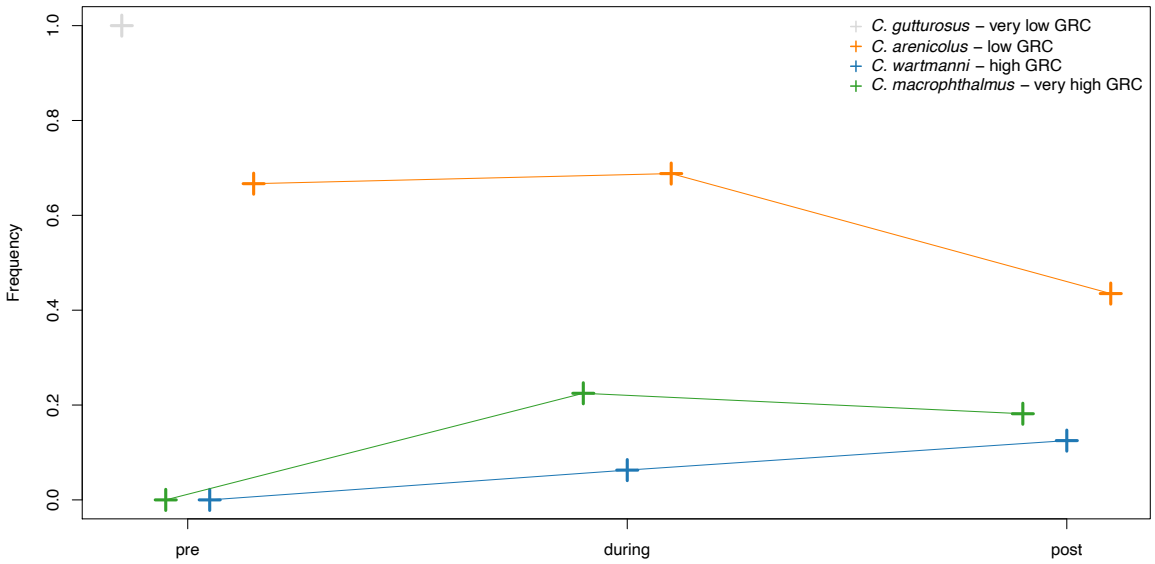


Figure S3: Allele frequency trajectory for the gill-raker count locus from De-Kayne et al. (2022). Plots shows the allele frequencies in each population of the SNP found to be significantly associated with gill-raker count by De-Kayne et al. (2022). The legend shows the symbol and color of each species, as well as the gill-raker count of each species (very low, low, high, very high) according to De-Kayne et al. (2022) and Vonlanthen et al. (2012).

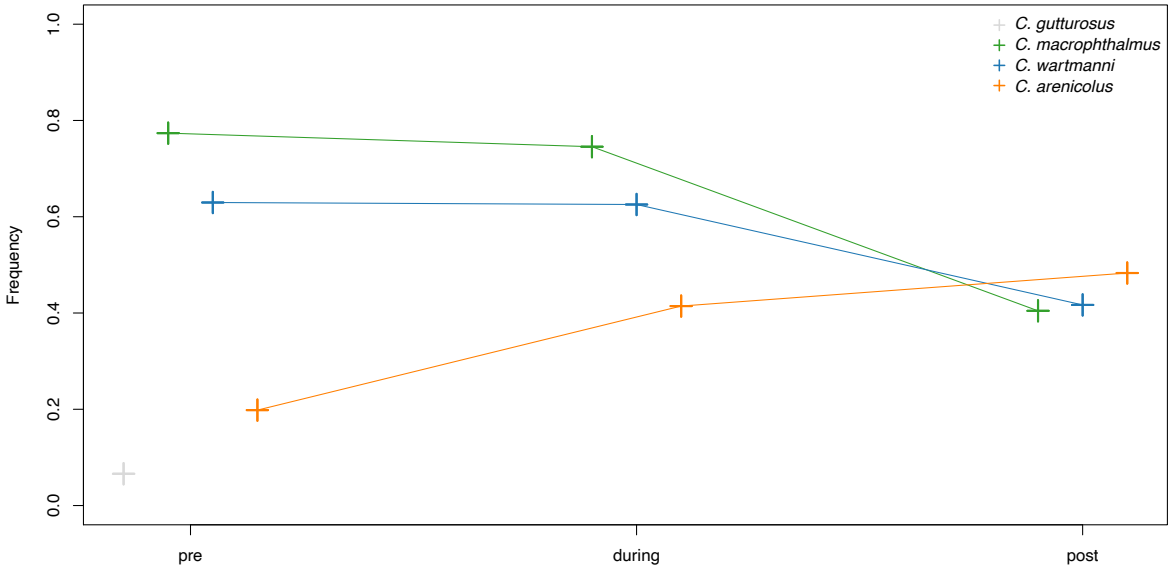


Figure S4: Allele frequency trajectories in *vgll3* paralogs in the whitefish genome assembly (De-Kayne et al. 2020). Plot shows the allele frequencies in each population of the one SNP within the two vgll3 paralogs in the whitefish genome, where at least one of the pre-eutrophcation populations has a minor allele frequency above 0.15 and hence might reflect relevant genetic variation. *vgll3* has been found to controll age at maturity in Atlantic salmon (Barson et al., 2015).

Supplementary Table S1: Overview over all used samples. All used individuals species assignment of the scale sample when it was collected ("Species"), year of collection, sampling timpeoint in relation to the eutrophication period ("Time"), species assignment based on genomic data ("Species Genomic") and ENA sample accessions. Samples for which "Species" and "Species Genomic" columns do not match have been excluded from analyses.

Species	Year	Time	Species Genomic	Lab ID	ENA accession
C. wartmanni	2015	post	*C. wartmanni*	121	ERS6670439
C. wartmanni	2015	post	*C. wartmanni*	122	ERS6670440
C. wartmanni	2015	post	*C. wartmanni*	123	ERS6670441
C. arenicolus	2015	post	*C. arenicolus*	126	ERS6670448
C. arenicolus	2015	post	*C. arenicolus*	127	ERS6670449
C. arenicolus	2015	post	*C. arenicolus*	128	ERS6670450
C. wartmanni	2015	post	*C. wartmanni*	131	ERS6670442
C. macrophthalmus	2015	post	*C. macrophthalmus*	132	ERS6670445
C. macrophthalmus	2019	post	*C. macrophthalmus*	212608	ERS12047271
C. macrophthalmus	2019	post	*C. macrophthalmus*	212623	ERS12047279
C. arenicolus	2019	post	*C. arenicolus*	212631	ERS12047284
C. arenicolus	2019	post	*C. arenicolus*	212633	ERS12047285
C. arenicolus	2019	post	*C. arenicolus*	212634	ERS12047286
C. macrophthalmus	2019	post	*C. macrophthalmus*	212663	ERS12047313
C. macrophthalmus	2019	post	*C. macrophthalmus*	212666	ERS12047316
C. macrophthalmus	2019	post	*C. macrophthalmus*	212671	ERS12047319
C. macrophthalmus	2019	post	*C. macrophthalmus*	212672	ERS12047320
C. macrophthalmus	2019	post	*C. macrophthalmus*	212678	ERS12047326
C. macrophthalmus	2019	post	*C. macrophthalmus*	212694	ERS12047341
C. macrophthalmus	2019	post	*C. macrophthalmus*	212695	ERS12047342
C. macrophthalmus	2019	post	*C. macrophthalmus*	212696	ERS12047343
C. arenicolus	2019	post	*C. arenicolus*	212701	ERS12047348
C. wartmanni	2019	post	*C. macrophthalmus*	212702	this study
C. wartmanni	2019	post	*C. arenicolus*	212703	this study
C. wartmanni	2019	post	*C. wartmanni*	212704	this study
C. wartmanni	2019	post	*C. wartmanni*	212705	this study
C. macrophthalmus	2019	post	*C. macrophthalmus*	212718	ERS12047354
C. macrophthalmus	2019	post	*C. macrophthalmus*	212725	ERS12047361
C. macrophthalmus	2019	post	*C. wartmanni*	212727	ERS12047363
C. gutturosus	1948	pre	*C. gutturosus*	S01_0003_gutturosus_1948	this study
C. gutturosus	1948	pre	*C. gutturosus*	S01_0006_gutturosus_1948	this study
C. gutturosus	1948	pre	*C. gutturosus*	S01_0020_gutturosus_1948	this study
C. gutturosus	1948	pre	*C. gutturosus*	S01_0025_gutturosus_1948	this study
C. gutturosus	1948	pre	*C. gutturosus*	S01_0031_gutturosus_1948	this study
C. gutturosus	1948	pre	*C. gutturosus*	S01_0039_gutturosus_1948	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0001_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0002_arenicolus_1946	this study

C. arenicolus	1946	pre	*C. arenicolus*	S11_0003_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0004_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0005_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0009_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0010_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0014_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0015_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0016_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0017_arenicolus_1946	this study
C. arenicolus	1946	pre	*C. arenicolus*	S11_0019_arenicolus_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0400_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0402_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0403_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0404_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0406_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0409_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0410_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0411_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0414_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0415_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0417_wartmanni_1946	this study
C. wartmanni	1946	pre	*C. wartmanni*	S12_0419_wartmanni_1946	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0001_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0002_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0005_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0011_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0012_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0013_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0015_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0016_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0017_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0018_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0019_arenicolus_1973	this study
C. arenicolus	1973	during	*C. arenicolus*	S13_0020_arenicolus_1973	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0006_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0010_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0011_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0012_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0013_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0014_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0015_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0016_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0017_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0018_wartmanni_1980	this study

C. wartmanni	1980	during	*C. wartmanni*	S16_0020_wartmanni_1980	this study
C. wartmanni	1980	during	*C. wartmanni*	S16_0023_wartmanni_1980	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0026_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0030_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0032_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0035_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0036_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0037_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0038_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0039_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. macrophthalmus*	S19_0040_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. gutturosus*	S19_0046_macrophthalmus_1935	this study
C. macrophthalmus	1935	pre	*C. gutturosus*	S19_0050_macrophthalmus_1935	this study
C. macrophthalmus	1979	during	*C. macrophthalmus*	S19_0098_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. wartmanni*	S19_0102_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. macrophthalmus*	S19_0103_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. macrophthalmus*	S19_0104_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. macrophthalmus*	S19_0107_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. macrophthalmus*	S19_0108_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. arenicolus*	S19_0109_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. wartmanni*	S19_0110_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. wartmanni*	S19_0111_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. wartmanni*	S19_0112_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. wartmanni*	S19_0113_macrophthalmus_1979	this study
C. macrophthalmus	1979	during	*C. wartmanni*	S19_0114_macrophthalmus_1979	this study
C. gutturosus	1937	pre	*C. gutturosus*	S20_0051_gutturosus_1937	this study
C. gutturosus	1937	pre	*C. gutturosus*	S20_0068_gutturosus_1937	this study
C. gutturosus	1937	pre	*C. gutturosus*	S20_0069_gutturosus_1937	this study
C. gutturosus	1937	pre	*C. gutturosus*	S20_0075_gutturosus_1937	this study
C. gutturosus	1937	pre	*C. gutturosus*	S20_0079_gutturosus_1937	this study
C. gutturosus	1937	pre	*C. gutturosus*	S20_0080_gutturosus_1937	this study

Chapter IV

IV. The genomic basis of adaptation to profundal habitats across the Swiss Alpine whitefish radiation

David Frei[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

Parallelism in the form of parallel phenotypic evolution, resulting in similar but independently evolved traits in closely related lineages, has been described for several adaptive radiations. Within the Swiss Alpine whitefish radiation, more than 30 species have been taxonomically described. In nine different Swiss lake systems, several phenotypically similar species have evolved independently. Profundal species, adapted to spawn in depths below 60m, have repeatedly evolved in lakes Thun, Constance, Lucerne, Zug and Walen. We here sequenced and analysed whole-genome resequencing data of all whitefish species of four of these Swiss lakes (Thun, Constance, Lucerne and Walen) to identify genomic signals of parallel profundal adaptation. The results indicate that only a small fraction of the genome (~0.44%) shows signals of parallel differentiation across all six studied profundal species. However, each test for parallel evolution between all possible pairs of profundal species was highly significant. Consistent with previous work investigating parallel evolution of a different Alpine whitefish ecomorph, these results suggest that profundal adaptation shows indications of genomic parallel evolution, but likely also exhibits genotypic redundancy.

Introduction

Many well-described adaptive radiations exhibit a certain degree of parallelism in the form of parallel phenotypic evolution resulting in similar but independently evolved traits in closely related lineages (Futuyma, 1986; Schluter, Clifford, Nemethy, & McKinnon, 2004). Parallel phenotypic evolution in adaptive radiations is often associated with features of the environment (Schluter et al., 2004), and the repeated evolution of similar traits linked to certain environmental characteristics implies a fundamental role of natural selection (Cai et al., 2019; Schluter, 1990). Although genetic parallelism (what genes are re-used in parallel evolution) is relatively well understood, less is known about the evolution of genomic parallelism (what fraction of the genome evolves in parallel and why) and the genomic basis of phenotypes evolved through parallel evolution (Bohutinska et al., 2021).

The number of genes required to produce a given phenotype can vary dramatically. Genetic redundancy is defined as two genes having the same function, such that if one of these two genes is inactivated, it has no effect on the phenotype (Nowak, Boerlijst, Cooke, & Smith, 1997). Following this definition, genotypic redundancy has been defined when more than one genotype can produce the same phenotype (Laruson, Yeaman, & Lotterhos, 2020). In consequence, when a phenotype evolved repeatedly due to adaptation to parallel selective pressures, this is not necessarily predicted to result in genomic parallelism or parallel genomic differentiation. As a result, the fraction of the genome that evolves in parallel differs greatly, even though repeated evolution in response to parallel selective pressures is thought to make evolution predictable (Blount, Lenski, & Losos, 2018; Bohutinska et al., 2021). Hence, the degree of genomic parallelism underlying a repeatedly evolved phenotype can range from full genomic parallelism (parallel evolution at all variable positions of within genes underlying a trait that is evolving in parallel) to the complete absence thereof as a consequence of genotypic redundancy.

Within the nine large perialpine lakes or lake systems of Switzerland, more than 30 Alpine whitefish species have evolved since deglaciation (~ 10'0000 - 15'000 years BP), with up to six species per lake (Hudson, Vonlanthen, & Seehausen, 2011; Selz, Donz, Vonlanthen, & Seehausen, 2020; Steinmann, 1950). In some of the Swiss lakes, phenotypically and ecologically similar species have evolved independently, implying a fundamental role of natural selection in the process of adaptation and speciation of the Alpine whitefish radiation (Hudson et al., 2011; Ostbye, Bernatchez, Naesje, Himberg, & Hindar, 2005; Praebel et al., 2013), a pattern very similar to species-for-species matching (Schluter, 1990) as described in *Anolis* lizards (Mahler, Ingram, Revell, & Losos, 2013).

In several (but not all) of these deep perialpine lakes (e.g., lakes Constance, Lucerne, Thun, and Walen), one or two species adapted to the profundal zones of these lakes have evolved (De-Kayne et al., 2022; Hudson et al., 2011). Either, these profundal species are adapted to feed on benthic macroinvertebrates (e.g., *C. gutturosus* in Lake Constance or *C. profundus* in Lake Thun), or they are adapted to feed on zooplankton (e.g., *C. heglingus* in Lake Walen, *C. nobilis* and *C. muelleri* in Lake Lucerne or *C. albellus* in Lake Thun) (Selz et al., 2020; Selz & Seehausen, 2023; Steinmann, 1950; Vonlanthen et al., 2012). Deep- and shallow water habitats of a lake have contrasting environmental characteristics (e.g., mean and variation in temperature, light conditions, pressure difference, oxygen availability, food resources, parasites and predation), presumably resulting in strong divergent selection between the two habitats (Ingram, Hudson, Vonlanthen, & Seehausen, 2012; Seehausen & Magalhaes, 2010). As a result of adaptation to deep water, profundal species are phenotypically differentiated (see Materials and Method section for a description of the studied profundal species) from other sympatric species (De-Kayne et al., 2022; Doenz, Bittner, Vonlanthen, Wagner, & Seehausen, 2018; Frei, De-Kayne, Selz, Seehausen, & Feulner, 2022; Selz et al., 2020; Selz & Seehausen, 2023; Steinmann, 1950). Often, they are

also genetically strongly differentiated from other sympatric species (De-Kayne et al., 2022; Doenz et al., 2018; Frei, De-Kayne, et al., 2022; Ingram et al., 2012). Because of the high level of genetic differentiation and the clear phenotypic differences, the repeated evolution of profundal whitefish species provides an excellent opportunity to study the genomic landscape of parallel evolution.

Recent work suggested that genomic differentiation underlying the phenotypic divergence between sympatric Alpine whitefish species to be scattered across the genome, largely based on a high number of small effect loci and only very few loci with larger effect (De-Kayne et al., 2022). Altogether, adaptation to the profundal zone includes many multidimensional traits (e.g., body shape, spawning time and depth, pigmentation, eye size) (Doenz et al., 2018; Selz et al., 2020; Selz & Seehausen, 2023; Steinmann, 1950), and thus the genomic basis of profundal adaptation is probably highly polygenic. Hence, there is a lot of potential for genotypic redundancy during the parallel evolution of profundal whitefish species within the Alpine whitefish radiation.

During the last century, 29% of the Alpine whitefish species went extinct as a result of anthropogenic eutrophication of many Swiss lakes (Vonlanthen et al., 2012). Profundal species were heavily affected by the hypoxic conditions during the period of eutrophication that prevented successful reproduction on deep water spawning grounds and resulted in several extinction events (Frei, De-Kayne, et al., 2022; Selz & Seehausen, 2023; Vonlanthen et al., 2012). As a consequence, understanding the genomic basis of profundal adaptation is instrumental to assess the loss of genomic-, functional- and species diversity caused by anthropogenic eutrophication during the last century.

Here, we generated whole-genome resequencing data of all taxonomically described whitefish species for four Swiss lakes with one to two profundal spawning whitefish species.

These are Lake Thun (six species total, two profundal spawning species), Lake Lucerne (six species total, two spawning profundal species), and Lake Walen (two species, one profundal spawning species). We then combined this dataset with previously published data for all four Lake Constance whitefish species (Frei, De-Kayne, et al., 2022; Frei, Reichlin, Seehausen, & Feulner, 2022), including the extinct profundal spawning *C. gutturosus*. Because both profundal species of Lake Zug are extinct (Selz & Seehausen, 2023) and no suitable tissue samples of these two species were available, we could not include the radiation from Lake Zug. Using this data set, we construct a phylogenetic tree to illustrate the relationships of species within and between lakes. We confirmed the reciprocal monophyly of the four lake-specific species flocks providing evidence for the independent and parallel evolution of the studied profundal species. We further reveal the genomic landscape of parallel differentiation between the profundal and other species in each lake.

Materials and Methods

Study system

In all four studied lakes (Thun, Constance, Lucerne, and Walen), one or two profundal whitefish species have been described, although the species of Lake Constance went extinct during the period of anthropogenic eutrophication (Frei, De-Kayne, et al., 2022; Frei, Reichlin, et al., 2022; Vonlanthen et al., 2012). However, due to a historical scale sample collection, sequencing data of this extinct species has been published (Frei, De-Kayne, et al., 2022). The six profundal spawning species used in this study are either benthos feeders (*C. gutturosus* in Lake Constance and *C. profundus* in Lake Thun; Steinmann, 1950; Vonlanthen et al., 2012; Selz et al., 2020), or profundal zooplanktivores (*C. heglingus* in Lake Walen, *C. nobilis* and *C. muelleri* in Lake Lucerne, and *C. albellus* in Lake Thun; Selz et al., 2020; Selz & Seehausen, 2023)

Lake Thun harbours two profundal spawning species: *C. profundus* and *C. albellus*. *C. profundus* and *C. albellus* are both slow growing species (Selz et al., 2020; Steinmann, 1950; Vonlanthen et al., 2012). Both species spawn during summer (and winter) in depths below 60m (Selz et al., 2020; Steinmann, 1950; Vonlanthen et al., 2012). *C. profundus* has a very low gill-raker count as a result of adaptation to feeding on benthic macroinvertebrates, while *C. albellus* exhibits a high gill-raker count as result of adaptation to feeding on zooplankton (Selz et al., 2020; Vonlanthen et al., 2012).

The extinct Lake Constance species *C. gutturosus* is phenotypically similar to *C. profundus* of Lake Thun. *C. gutturosus* was slow growing and spawned during summer and winter in deep regions of the lake (Steinmann, 1950). The species exhibited a very low gill-

raker-count and was feeding exclusively on benthic macroinvertebrates (Steinmann, 1950; Vonlanthen et al., 2012).

C. nobilis and *C. muelleri* in Lake Lucerne also spawn in profundal regions of the lake (Selz & Seehausen, 2023; Vonlanthen et al., 2012). *C. nobilis* is described as summer spawning species, while *C. muelleri* has a long spawning period with peaks during summer and winter (Selz & Seehausen, 2023; Vonlanthen et al., 2012). Both *C. nobilis* and *C. muelleri* have a high gill-raker count, characteristic for species adapted to feed on zooplankton (Selz & Seehausen, 2023; Vonlanthen et al., 2012).

C. heglingus in Lake Walen is ecologically very similar to *C. albellus* in Lake Thun and *C. muelleri* in Lake Lucerne. *C. heglingus* is a slow growing species and spawns in the profundal zone of Lake Walen (Vonlanthen et al., 2012). Both summer and winter spawning populations have been described (Vonlanthen et al., 2012). *C. heglingus* has a high gill-raker count and is adapted to feed on zooplankton (Vonlanthen et al., 2012).

Sample collection

The samples that were sequenced here were collected for previous studies (Hudson et al. (2016) for Lake Lucerne, Doenz et al. (2018) for Lake Thun and Selz (unpublished) for Lake Walen) in accordance with permits issued by the cantons of Zurich and St.Gallen (ZH128/15) for Lake Walen and Lake Constance, Bern (BE68/15) for Lake Thun and Lucerne (LU04/14) for Lake Lucerne. Individual specimens were phenotypically assigned to species. Individuals were anaesthetized and subsequently euthanized using appropriate concentrations of tricaine methane sulfonate solutions (MS-222). Fin-clips were taken and stored in 100% analytical ethanol until extraction of DNA, and muscle tissue was preserved frozen at -80C°. DNA extraction was done using the Qiagen DNeasy blood and tissue kit (Qiagen AG, CH) and following the manufacturers recommendations.

DNA-extraction and sequencing

For each sample where a finclip was available, one Illumina paired-end TruSeq DNA PCR-Free library (Illumina GmbH, CH) was prepared. For muscle tissue samples, an Illumina paired-end TruSeq DNA Nano library (Illumina GmbH, CH) was produced (see Supplementary Table 1 for overview over all sequenced samples). Library preparation was performed by the NGS platform of the University of Bern following the manufacturer's instructions. Libraries were sequenced on a Novaseq 6000 S4 flowcell at the NGS platform of the University of Bern. We additionally used existing sequencing data generated by De-Kayne et al. (2022), downloaded from ENA with accession PRJEB47792 and data generated by Frei et al. (2022), downloaded from ENA with accession PRJEB53050.

Processing reads and mapping

Raw sequencing reads were processed according to Frei et al. (2022). In brief, Poly-G strings were removed using fastp 0.20.0 (Chen, Zhou, Chen, & Gu, 2018) and overlapping paired end reads with 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 (ENA accession: GCA_902810595.1; De-Kayne, Zoller, & Feulner, 2020) with bwa mem version 0.7.12 (Li & Durbin, 2009) and default parameters, except for 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/). We excluded all chromosomes of the Alpine whitefish reference genome with potentially collapsed duplicated regions (De-Kayne et al., 2020).

Population genomic analysis

Genotype likelihoods and minor allele frequencies at polymorphic sites were calculated using the samtools genotype likelihood model in angsd 0.925 (Korneliussen, Albrechtsen, & Nielsen, 2014; Li, 2011). We only included sites covered with at least two reads in at least 159 individuals (of a total of 169 individuals, resulting in a maximum of 6% missing data), passing a p-value cut-off of 10E-6 for being variable and having not more than two different alleles. Not uniquely mapping reads and those with a mapping quality below 30, as well as bases with quality score below 20 were not considered for calculation of genotype likelihoods. The following p-value cut-offs for SNP filters in angsd version 0.925 were used: -sb_pval 0.05 -qscore_pval 0.05 -edge_pval 0.05 -mapq_pval 0.05. This resulted in a total of 8'182'760 polymorphic sites for further analysis.

We then used a minor allele frequency threshold of 5% to perform a PCA on the whole data set using PCAngsd 1.02 (Meisner & Albrechtsen, 2018), resulting in a total of 3'374'468 SNPs. In addition, we split the data set by lake, and again performed a PCA using PCAngsd 1.02 and applying a 5% minor allele frequency threshold.

To find genomic regions where the profundal species of each lake are differentiated, we contrasted the profundal spawning species of each lake with all other non-profundal species of the respective lake pooled together. We calculated the weighted genome-wide F_{ST} and the F_{ST} in 5 kb windows along the genome in angsd 0.925 based on one- and two-dimensional site frequency spectra inferred from site allele frequencies based on genotype likelihoods (Bhatia, Patterson, Sankararaman, & Price, 2013; Korneliussen et al., 2014; Korneliussen, Moltke, Albrechtsen, & Nielsen, 2013; Nielsen, Korneliussen, Albrechtsen, Li, & Wang, 2012). We then assessed the overlap between the 5 kb windows with the highest differentiation (top 5% quantile) in R (R Core Team, 2018). We used the F_{ST} in 5 kb along the

genome between each profundal species and all non-profundal sympatric species to look for signals of parallel differentiation, respectively parallel adaptation to profundal habitats, using the R-package PicMin 0.0.0.9 (Booker, Yeaman, & Whitlock, 2023). PicMin is derived from the theory of order statistics and tests for repeated molecular evolution to estimate significance at the level of individual loci, using the results of genome scans (such as F_{ST} scans) (Booker et al., 2023). The statistical power to detect repeated adaptation increases with the number of lineages that have signals of repeated adaptation of a given locus in multiple lineages (Booker et al., 2023). We followed the vignette of the package on github (https://github.com/TBooker/PicMin/blob/main/vignettes/Arabidopsis-vignette.Rmd) and adjusted the threshold to determine adaptation to 0.05 (which is a suitable threshold for six lineages or species (Booker et al., 2023)). We performed the analysis on all six profundal spawning species combined. Because our aim was to screen the genome for signals of repeated differentiation, we used a relatively inclusive rather than a restrictive approach with a significance threshold of $q<0.1$ (Booker et al., 2023), similar to Bohutinska et al. (2021). In addition, we assessed parallel differentiation separately for each pairwise combination of the six studied profundal species with the PicMin approach (see https://github.com/TBooker/PicMin/blob/main/vignettes/Arabidopsis-vignette.Rmd).

Windows that indicated significant signals of parallel evolution by PicMin were further investigated to identify which genes they overlap. As suggested by Booker et al. (2023), we used a permissive significance threshold ($q<0.5$) for this purpose to maximize the number of genes for the enrichment analysis. Gene annotations (from the Alpine whitefish genome (De-Kayne et al., 2020); ENA accession: GCA_902810595.1) that overlap in their position with the identified windows indicating signals of parallel evolution were identified using bedtools 2.28.0 (Quinlan, 2014). Gene enrichment for specific gene ontology (GO) terms (from https://datadryad.org/stash/dataset/doi:10.5061/dryad.xd2547ddf) within these

windows was tested using the R package topGO 2.38.1 (Alexa & Rahnenfuhrer, 2020) separately for each of the three ontology classes cellular component (CC), biological processes (BP), and molecular function (MF). We used Fisher's exact test applying the "weight" algorithm to each ontology class (with no fdr multiple testing correction in accordance to the topGO manual). The 10 most significant hits for each ontology class were reported.

To visualize phylogenetic relationships between lakes as well as between species within each lake, we produced a maximum likelihood phylogeny. We calculated genotype likelihoods using the whole data set of 169 individuals and including the *C. albula* outgroup from De-Kayne et al. (2022) (ENA sample accession ERS11527697) with angsd 0.925 (Korneliussen et al., 2014). We again used the same parameters as above (data of minimally 159 individuals at each position, at least 2 reads from every individual, p-value cut-off of 10E-6 for being variable, minimum mapping quality of 30 and minimum base quality 20, only sites with two alleles, and we used the identical p-value cutoffs for SNP filters), resulting in a total of 14'546'231 SNPs. We then inferred genotypes and phased these using beagle 4.1 (Browning & Browning, 2007). We thinned this dataset using VCFtools 0.1.16 (Danecek et al., 2011) so that all SNPs were at least 10 kb 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 76'131 SNPs. We then converted the VCF file to a phylip file using the python script vcf2phylip.py (https://github.com/edgardomortiz/vcf2phylip). We used RAxML-NG 1.01 (Kozlov, Darriba, Flouri, Morel, & Stamatakis, 2019) to produce the phylogeny with the ASC_GTRGAMMA substitution model and default parameters. The resulting phylogeny was plotted with Figtree 1.4.4 (https://github.com/rambaut/figtree).

Results

Population structure within and between lakes

To visualize the population structure within and between lakes Constance, Walen, Lucerne, and Thun, we performed a PCA (Fig. 1) based on genotype likelihoods in PCAngsd on our full data set (3'374'468 SNPs) including all lakes. Each lake formed its own cluster in PC-space of principal components one and two, with the exception of *C. acrinasus* from Lake Thun that clustered with the Lake Constance species flock, consistent with earlier work (De-Kayne et al., 2022; Doenz et al., 2018; Selz et al., 2020). Although the PCA suggests that several individuals were misidentified in the field and likely belong to another species (see e.g., *C. zurichensis* of Lake Walen or *C. steinmanni* and *C. acrinasus* in Lake Thun), almost all individuals of the six studied profundal species (*C. profundus*, *C. albellus*, *C. gutturosus*, *C. nobilis*, *C. muelleri* and *C. heglingus*) clustered together and hence were correctly assigned to species (see also Supplementary Figures 1-4 for a separate PCA of each of the four lakes).

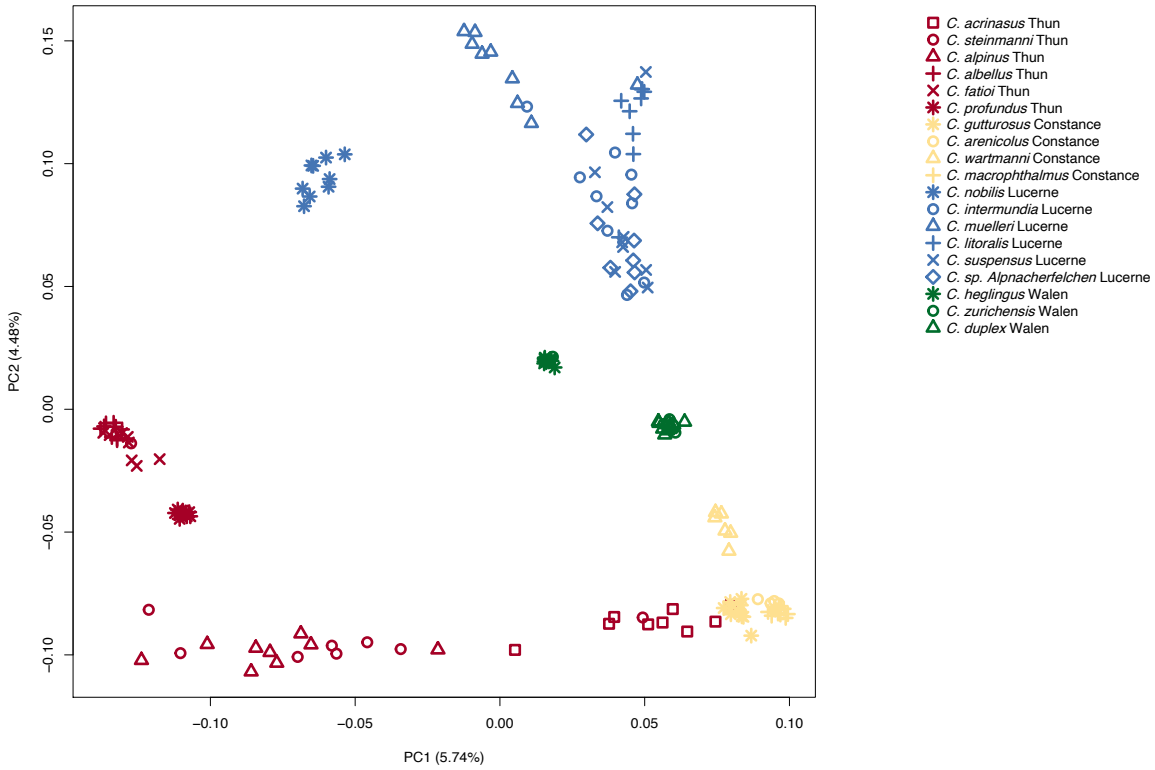


Figure 1: Population structure across all four studied lakes. PCA based on genotype likelihoods of 3'374'468 SNPs polymorphic sites, and including all sampled lakes and individuals. The plot illustrates the differentiation between lakes (different colors, see legend to the left) and amongst species (different symbols, see legend to the left).

Using a thinned dataset comprising of 76'131 sites, we generated a maximum likelihood phylogeny including all sampled lakes and species using RAxML-NG (Figure 2). In line with recent work (De-Kayne et al., 2022), our phylogeny indicated that the species flocks of each lake are monophyletic. One exception is *C. acrinasus* of Lake Thun. *C. acrinasus* clusters within the Lake Constance radiation, consistent with its translocation history from Lake Constance into Lake Thun (De-Kayne et al., 2022; Doenz et al., 2018). Both benthic profundal species (*C. gutturosus* in Lake Constance and *C. profundus* in Lake Thun) showed particularly long branch-lengths relative to the other sympatric species, in line with strong reproductive isolation between benthic profundal and sympatric species.

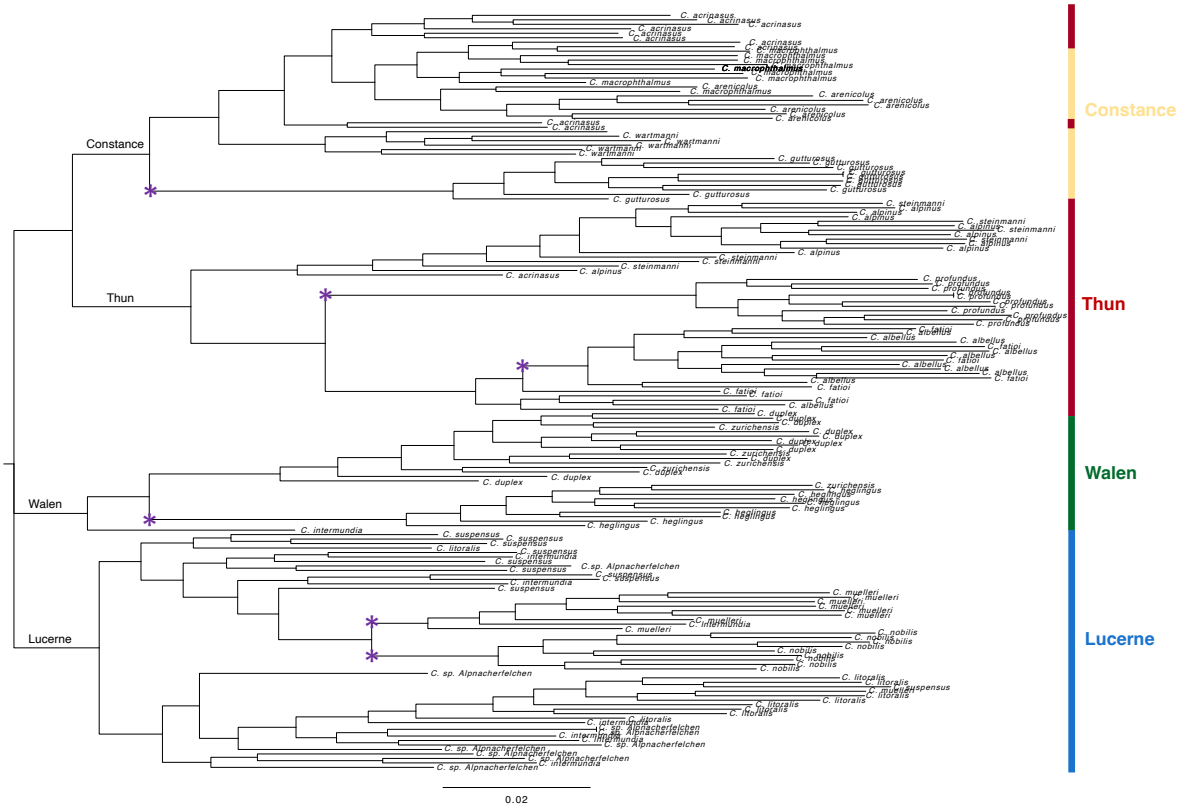


Figure 2: Phylogenetic tree including all individuals and constructed from a thinned data set (76'131 sites). Nodes of the six studied profundal species are highlighted with purple asterisks. The *C. albula* outgroup is not shown. All species from each lake cluster monophyletically (except for *C. acrinasus*, which was translocated from Lake Constance into Lake Thun (De-Kayne et al., 2022; Doenz et al., 2018; Selz et al., 2020)), and lakes are highlighted with colored bars to the right. Tip labels indicate species.

Genomic parallelism of profundal adaptation

To assess the degree of parallel differentiation between the profundal species in each lake, we calculated F_{ST} in 5 kb windows along the genome between the profundal species and all sympatric non-profundal species pooled together. Genome-wide F_{ST} between *C. gutturosus* and all three sympatric Lake Constance species was 0.077, F_{ST} between *C. profundus* and all four sympatric non-profundal Lake Thun species 0.114, F_{ST} between *C. albellus* and all four sympatric non-profundal Lake Thun species 0.059, F_{ST} between *C. nobilis* and all four non-profundal sympatric Lake Lucerne species 0.058, F_{ST} between *C. muelleri* and all four non-profundal sympatric Lake Lucerne species 0.041, and the F_{ST} between *C. heglingus* and the two sympatric Lake Walen species was 0.094. In total, seven of all 299'7515 kb windows (all on chromosome 29 and between position 32'977'500 and

position 33’077’500) were in the top 5 percentile in all six F_{ST} comparisons between profundal and non-profundal species across all four lakes (Supplementary Figure 5).

The two-way PicMin (Booker et al., 2023) analysis based on the F_{ST} between the profundal species and all other non-profundal species of each lake in 5 kb windows along the genome indicated significant parallel genomic differentiation in all pairwise comparisons of profundal species (Figure 3). The resulting p-values were highly significant (see Table 1).

Table 1: P-values of the two-way PicMin analysis. Table shows all p-values of the two-way PicMin analysis between all possible pairs of profundal species. The two comparisons of species within lakes (*C. profundus – C. albellus* in Lake Thun and *C. nobilis – C. muelleri* in Lake Lucerne) are not shown, as they might be biased by sharing of lake specific genomic variation.

		Lucerne		Thun		Constance
		C. nobilis	***C. muelleri***	***C. profundus***	***C. albellus***	***C. gutturosus***
Lucerne	***C. nobilis***	-				
	C. muelleri	-	-			
Thun	***C. profundus***	1.6E-29	6.0E-16	-		
	C. albellus	1.8E-231	5.1E-86	-	-	
Constance	***C. gutturosus***	6.2E-39	3.0E-12	9.7E-32	2.2E-27	-
Walen	***C. heglingus***	2.0E-139	8.5E-145	5.7E-33	2.1E-132	3.6E-58

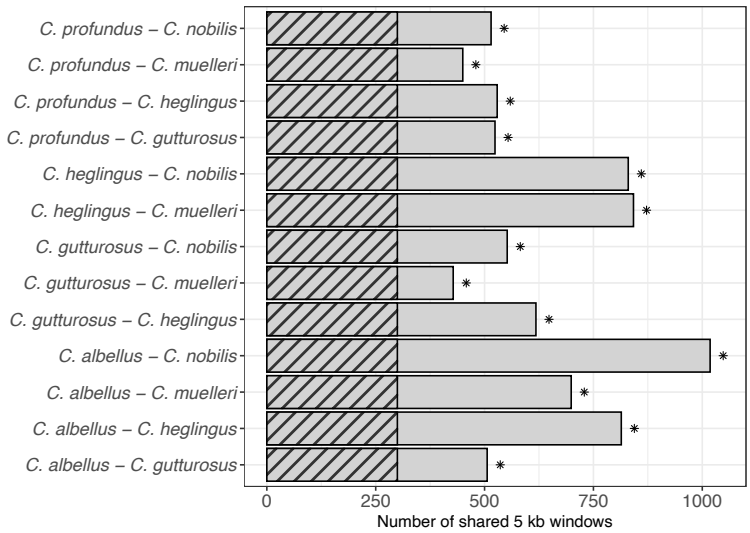


Figure 3: Pairwise signals of parallel genomic differentiation of profundal species. B) Barplots showing the results of the pairwise PicMin analysis. Bars show observed number of windows with signals of parallel differentiation between all possible pairwise comparisons of profundal species. The two comparisons within lakes (*C. nobilis – C. muelleri* in Lake Lucerne, *C. albellus – C. profundus* in Lake Thun) are not shown to avoid bias through lake specific genomic variation. The striped area corresponds to the null expectation for shared differentiated windows without parallel evolution. All comparisons indicated highly significant p-values (see Table 1).

In the PicMin analysis based on the six F_{ST} contrasts between the six profundal and all sympatric non-profundal species, ~0.44% (1304 out of 299'751 windows) of all genomic windows showed signals for parallel evolution (q<0.1; Figure 4 and Supplementary Figure 6). The mean number of species showing parallel evolution at these 1304 windows was ~4.5. We identified 4102 genes that overlapped with windows with p-values below the permissive threshold of q<0.5 used for the functional enrichment analysis. This set of genes was significantly enriched for several gene ontology terms (Supplementary Table 2), including photoperiodism and terms related brain development.

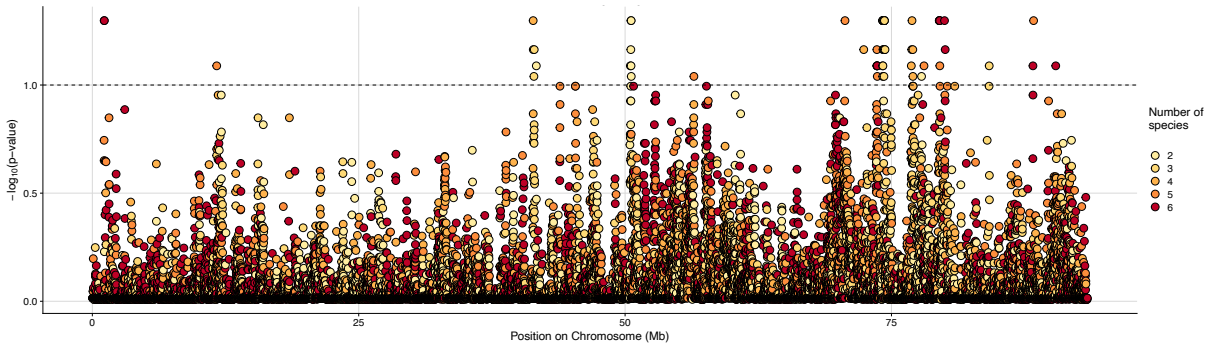


Figure 4: Genomic parallelism in profundal adaptation across the Alpine whitefish radiation. C) PicMin analysis including all six profundal species in all four lakes. The plot shows the results along the first chromosome of the reference genome (WFS1), other chromosomes are shown in Supplementary Figure 6. The plot shows fdr-corrected p-values of 5 kb windows along the genome. Colour corresponds to number the number of species with a signal for parallel evolution. The dashed line indicates the significance threshold (q<0.1).

Discussion

Parallel evolution is a common phenomenon in many well-described adaptive radiations, resulting in similar but evolutionary independent phenotypes (Futuyma, 1986; Schluter et al., 2004). Within the Alpine whitefish radiation, several profundal whitefish species have evolved independently since deglaciation. We generated whole-genome resequencing data of all species of four lakes with one or two profundal spawning species to detect signals of parallel evolution along the genome. Our results show indications for parallel differentiation and adaptation, but also suggest the involvement of genotypic redundancy during the evolution of the four studied profundal species.

Both, our PCA approach and our phylogeny suggested clear genomic differentiation between (most of) the profundal and all other species of each lake. In the phylogeny, almost all individuals from the six studied profundal species cluster monophyletically, a pattern that not all other species show. The Alpine whitefish radiation is evolutionary young and there is evidence for extensive and recent hybridization (Frei, De-Kayne, et al., 2022; Vonlanthen et al., 2012). Hence, the monophyletic clustering of profundal species is remarkable. Branches to the benthic profundal species are long and quite distinct, and in Lake Constance, the profundal species (*C. gutturosus*) is the most basal species of the species flock. Pronounced differences between two environments can result in strong divergent selection (Nosil, 2012), and divergent ecological selection can contribute to the evolution of reproductive isolation (Rundle & Nosil, 2005; Schluter, 2000). The results presented here are consistent with a strong ecological contrast between shallow and deep habitats, and the resulting divergent selection between the profundal and other species in each lake might have contributed to reproductive isolation. The clear genomic differentiation and the long and distinct branches leading to the profundal species in our phylogenetic tree suggest that the profundal species are characterized by a high degree of species-specific genomic variation. Previous work showed

that several profundal species went extinct during the period of anthropogenic eutrophication (Vonlanthen et al., 2012). Hence, the extinction of profundal species resulted in the loss of high amounts of genomic variation relative to the total variation of the entire species flock in each lake.

The degree of parallelism is expected to scale with divergence and closely related lineages are predicted to exhibit a higher level of genomic parallelism than distantly related lineages (Bohutinska et al., 2021). Thus, a high degree of parallelism in the genomic basis of profundal adaptation would be expected, as the Alpine whitefish radiation is evolutionary young and thus, divergence is comparably low. Further, morphological data provided evidence for parallel divergence between other profundal and littoral European whitefish (beyond the Alpine whitefish radiation) in Norway (Praebel et al., 2013). In contrast to the prediction, our inclusive approach identified only a relatively small number of windows along the genome (~0.44%) showing signals of parallel evolution. In other systems with repeated adaptation to environmental contrasts, e.g., *A. saxatilis*, much higher proportions of the genome (2-28%) indicated signals of parallel evolution (Morales et al., 2019; Ravinet et al., 2016; Westram et al., 2014). However, all pairwise tests for genomic parallelism between all possible combinations of profundal species resulted in highly significant p-values. All comparisons including two planktivore species (*C. albellus*, *C. heglingus*, *C. muelleri* and *C. nobilis*) resulted in a higher number of genomic windows with signals of parallel evolution than the comparisons including benthic profundal species. This suggests that the degree of genomic parallelism between the four planktivorous profundal whitefish species is higher than between the two profundal benthic species (*C. gutturosus* and *C. profundus*). Taken together, these findings indicate a low level of genomic parallelism in profundal adaptation across the four studied perialpine lakes, in line with the presence of similar selective pressures in profundal habitats within all four lakes.

The ancestor to the Salmonid lineage experienced a whole-genome duplication 80-100 million years ago (Lien et al., 2016; Macqueen & Johnston, 2014; Near et al., 2012). The genomes of the different Salmonid lineages were uniquely shaped in the course of the independent rediploidization following this salmonid-specific fourth vertebrate whole-genome duplication within each lineage (Robertson et al., 2017). However, this whole-genome duplication event could be a source for genotypic redundancy within Coregonids. Studies in North American whitefish found evidence for low levels of genomic parallelism together with indications of genotypic redundancy (Bernatchez, Laporte, Perrier, Sirois, & Bernatchez, 2016; Renaut, Nolte, Rogers, Derome, & Bernatchez, 2011). Furthermore, ecologically relevant traits have a highly polygenic basis within the Alpine whitefish radiation (De-Kayne et al., 2022), resulting in a high potential for genotypic redundancy. The relatively low level of genomic parallelism across profundal species, but highly significant genomic parallelism in each pairwise comparison implicates a potential role of genotypic redundancy in traits related to profundal adaptation, consistent with results of De-Kayne et al. (2022) for another ecomorph contrast in the Alpine whitefish radiation.

A high degree of genotypic redundancy holds potential to facilitate rapid adaptation, because there are multiple genotypic solutions to the same selective pressure. Hence, genotypic redundancy might decrease the time (in generations) needed for selection to act upon the genomic variation of a population until an adapted genotype evolves, because it potentially increases the amount of genomic variation that can be used by selection (and it thus increases the evolutionary potential) and it reduces potential constraints. Especially the combination of a strong contrast between habitats (such as between shallow water and profundal habitats) and a certain degree of genotypic redundancy might not only facilitate adaptation, but also diversification and speciation. When such an environmental contrast is geographically replicated (e.g., in different lakes or on different islands), divergent adaptation

to such a contrast in the different replicates and the possibility of genotypic redundancy underlying traits might lead to rapid adaptation in each replicate, while some level of genomic parallelism (but not complete genomic parallelism) evolves. Thus, genotypic redundancy is relevant in the context of adaptive radiations due to its capacity to speed up the adaptation and diversification process, but also in the context of environmental change because of its potential to facilitate a rapid evolutionary response to the changing environmental conditions.

References

Alexa, A., & Rahnenfuhrer, J. (2020). TopGO: Enrichment analysis for gene ontology (R package version 2.42.0). doi:10.18129/B9.bioc.topGO

Bernatchez, S., Laporte, M., Perrier, C., Sirois, P., & Bernatchez, L. (2016). Investigating genomic and phenotypic parallelism between piscivorous and planktivorous lake trout (*Salvelinus namaycush*) ecotypes by means of RADseq and morphometrics analyses. *Molecular Ecology, 25*(19), 4773-4792. doi:10.1111/mec.13795

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting FST: The impact of rare variants. *Genome Research, 23*(9), 1514-1521. doi:10.1101/gr.154831.113

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science, 362*(6415), 1-10. doi:10.1126/science.aam5979

Bohutinska, M., Vlcek, J., Yair, S., Laenen, B., Konecna, V., Fracassetti, M., . . . Kolar, F. (2021). Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences of the United States of America, 118*(21), 1-10. doi:10.1073/pnas.2022713118

Booker, T., Yeaman, S., & Whitlock, M. (2023). Using genome scans to identify genes used repeatedly for adaptation. *Evolution*. doi:10.1093/evolut/qpac063

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics, 81*(5), 1084-1097. doi:10.1086/521987

Cai, Z., Zhou, L., Ren, N. N., Xu, X., Liu, R., Huang, L., . . . Ge, S. (2019). Parallel Speciation of wild rice associated with habitat shifts. *Molecular Biology and Evolution, 36*(5), 875-889. doi:10.1093/molbev/msz029

Chen, S. F., Zhou, Y. Q., Chen, Y. R., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), 884-890. doi:10.1093/bioinformatics/bty560

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330

De-Kayne, R., Selz, O. M., Marques, D. A., Frei, D., Seehausen, O., & Feulner, P. G. D. (2022). Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nature Communications, 13*(4479), 1-13. doi:10.1038/s41467-022-32181-8

De-Kayne, R., Zoller, S., & Feulner, P. G. D. (2020). A de novo chromosome-level genome assembly of *Coregonus* sp. *"Balchen"*: One representative of the Swiss Alpine whitefish radiation. *Molecular Ecology Resources, 20*(4), 1093-1109. doi:10.1111/1755-0998.13187

Doenz, C. J., Bittner, D., Vonlanthen, P., Wagner, C. E., & Seehausen, O. (2018). Rapid buildup of sympatric species diversity in Alpine whitefish. *Ecology and Evolution, 8*(18), 9398-9412. doi:10.1002/ece3.4375

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*, *6*(4), 461-468. doi:10.1038/s41559-022-01665-7

Frei, D., Reichlin, P., Seehausen, O., & Feulner, P. G. D. (2022). Introgression from extinct species facilitates adaptation to its vacated niche. *Molecular Ecology*, *32*(4), 841-853. doi:10.1111/mec.16791

Futuyma, D. J. (1986). *Evolutionary Biology* (2 ed.). Sunderland, Mass: Sinauer.

Hudson, A. G., Lundsgaard-Hansen, B., Lucek, K., Vonlanthen, P., & Seehausen, O. (2016). Managing cryptic biodiversity: Fine-scale intralacustrine speciation along a benthic gradient in Alpine whitefish (*Coregonus* spp.). *Evolutionary Applications*, *10*(3), 251-266. doi:10.1111/eva.12446

Hudson, A. G., Vonlanthen, P., & Seehausen, O. (2011). Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society B-Biological Sciences, 278*(1702), 58-66. doi:10.1098/rspb.2010.0925

Ingram, T., Hudson, A. G., Vonlanthen, P., & Seehausen, O. (2012). Does water depth or diet divergence predict progress towards ecological speciation in whitefish radiations? *Evolutionary Ecology Research, 14*(4), 487-502. doi:10.7892/boris.16333

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *Bmc Bioinformatics, 15*(356), 1-13. doi:10.1186/s12859-014-0356-4

Korneliussen, T. S., Moltke, I., Albrechtsen, A., & Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *Bmc Bioinformatics, 14*(289), 1-14. doi:10.1186/1471-2105-14-289

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics, 35*(21), 4453-4455. doi:10.1093/bioinformatics/btz305

Laruson, A. J., Yeaman, S., & Lotterhos, K. E. (2020). The importance of genetic redundancy in evolution. *Trends in Ecology & Evolution, 35*(9), 809-822. doi:10.1016/j.tree.2020.04.009

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21), 2987-2993. doi:10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., . . . Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature, 533*(7602), 200-205. doi:10.1038/nature17164

Macqueen, D. J., & Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B-Biological Sciences, 281*(1778), 1-8. doi:10.1098/rspb.2013.2881

Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science, 341*(6143), 292-295. doi:10.1126/science.1232392

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics, 210*(2), 719-731. doi:10.1534/genetics.118.301336

Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., & Butlin, R. K. (2019). Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. *Science Advances, 5*(12), 1-13. doi:10.1126/sciadv.aav9963

Near, T. J., Eytan, R. I., Dornburg, A., Kuhn, K. L., Moore, J. A., Davis, M. P., . . . Smith, W. L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences of the United States of America, 109*(34), 13698-13703. doi:10.1073/pnas.1206625109

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. R., & Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *Plos One, 7*(7), 1-10. doi:10.1371/journal.pone.0037558

Nosil, P. (2012). *Ecological speciation*. Oxford ; New York: Oxford University Press.

Nowak, M. A., Boerlijst, M. C., Cooke, J., & Smith, J. M. (1997). Evolution of genetic redundancy. *Nature, 388*(6638), 167-171. doi:10.1038/40618

Ostbye, K., Bernatchez, L., Naesje, T. F., Himberg, K. J. M., & Hindar, K. (2005). Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Molecular Ecology, 14*(14), 4371-4387. doi:10.1111/j.1365-294X.2005.02737.x

Praebel, K., Knudsen, R., Siwertsson, A., Karhunen, M., Kahilainen, K. K., Ovaskainen, O., . . . Amundsen, P. A. (2013). Ecological speciation in postglacial European whitefish: Rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and Evolution, 3*(15), 4970-4986. doi:10.1002/ece3.867

Quinlan, A. R. (2014). BEDTools: The swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics, 47*, 11.12.11-11.12.34. doi:https://doi.org/10.1002/0471250953.bi1112s47

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., Andre, C., & Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology, 25*(1), 287-305. doi:10.1111/mec.13332

Renaut, S., Nolte, A. W., Rogers, S. M., Derome, N., & Bernatchez, L. (2011). SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp.). *Molecular Ecology, 20*(3), 545-559. doi:10.1111/j.1365-294X.2010.04952.x

Robertson, F. M., Gundappa, M. K., Grammes, F., Hvidsten, T. R., Redmond, A. K., Lien, S., . . . Macqueen, D. J. (2017). Lineage-specific rediploidization is a mechanism to

explain time-lags between genome duplication and evolutionary diversification. *Genome Biology, 18*(111), 1-14. doi:10.1186/s13059-017-1241-z

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters, 8*(3), 336-352. doi:10.1111/j.1461-0248.2004.00715.x

Schluter, D. (1990). Species-for-species matching. *American Naturalist, 136*(4), 560-568. doi:10.1086/285115

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford: Oxford University Press.

Schluter, D., Clifford, E. A., Nemethy, M., & McKinnon, J. S. (2004). Parallel evolution and inheritance of quantitative traits. *American Naturalist, 163*(6), 809-822. doi:10.1086/383621

Seehausen, O., & Magalhaes, I. S. (2010). Geographical mode and evolutionary mechanism of ecological speciation in cichlid fish. In P. R. Grant & B. R. Grant (Eds.), *In Search of the Causes of Evolution: From Field Observations to Mechanisms.* Princeton: Princeton Universtiy Press.

Selz, O. M., Doenz, C. J., Vonlanthen, P., & Seehausen, O. (2020). 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

Selz, O. M., & Seehausen, O. (2023). A taxonomic revision of ten whitefish species from the lakes Lucerne, Sarnen, Sempach and Zug, Switzerland, with descriptions of seven new species (Teleostei, Coregonidae). *ZooKeys, 1144*, 85-169. doi:10.3897/zookeys.1144.67747

Steinmann, P. (1950). Monographie der schweizerischen Koregonen. Beitrag zum Problem der Entstehung neuer Arten. Spezieller Teil. *Schweizerische Zeitschrift für Hydrologie, 12*(1), 340-491. doi:10.1007/BF02486030

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Muller, R., Lundsgaard-Hansen, B., . . . Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature, 482*(7385), 357-362. doi:10.1038/nature10824

Westram, A. M., Galindo, J., Rosenblad, M. A., Grahame, J. W., Panova, M., & Butlin, R. K. (2014). Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology, 23*(18), 4603-4616. doi:10.1111/mec.12883

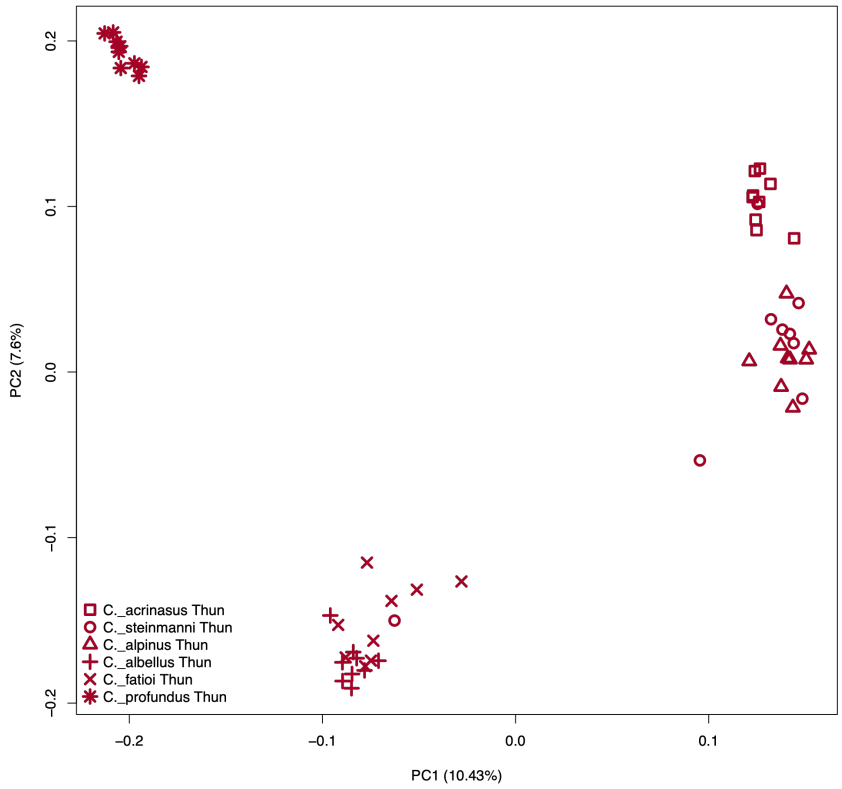


Figure S1: PCA on all individuals from Lake Thun. PCA based on genotype likelihoods of 3'125'823 SNPs polymorphic sites, and including all individuals of Lake Thun. The plot illustrates the differentiation amongst species (plotting symbol according to species, see legend within the plot).

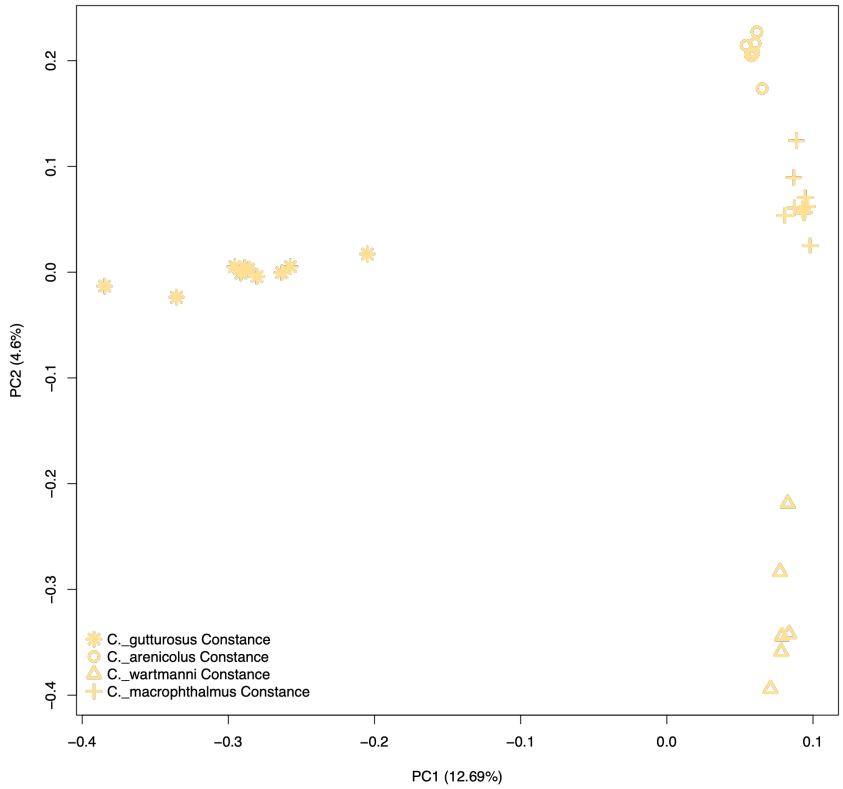


Figure S2: PCA on all individuals from Lake Constance. PCA based on genotype likelihoods of 3'261'099 SNPs polymorphic sites, and including all individuals of Lake Constance. The plot illustrates the differentiation amongst species (plotting symbol according to species, see legend within the plot).

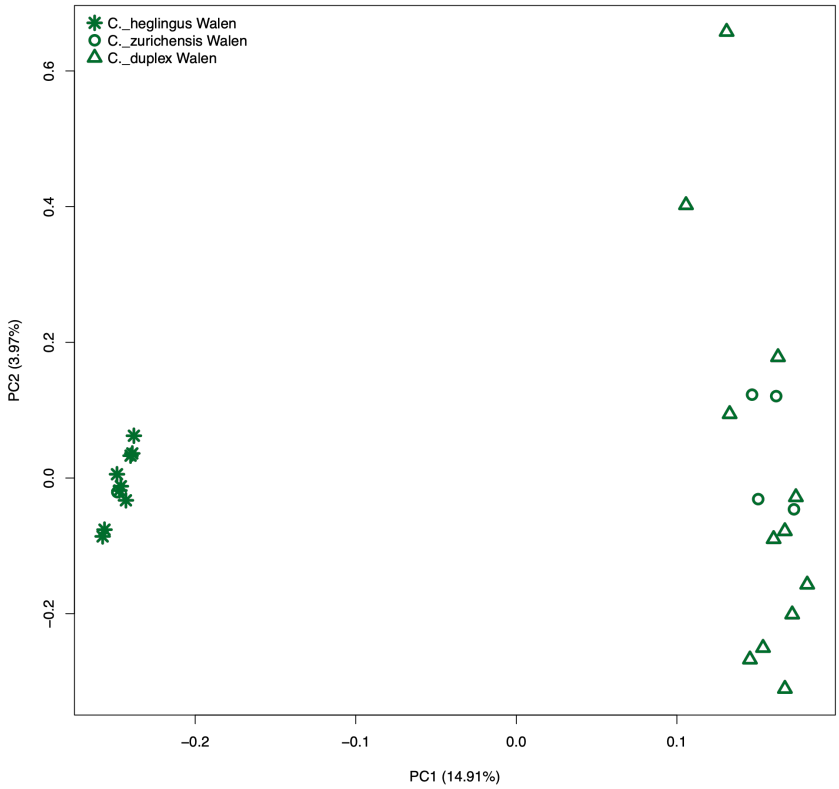


Figure S3: PCA on all individuals from Lake Walen. PCA based on genotype likelihoods of 3'099'149 SNPs polymorphic sites, and including all individuals of Lake Walen. The plot illustrates the differentiation amongst species (plotting symbol according to species, see legend within the plot).

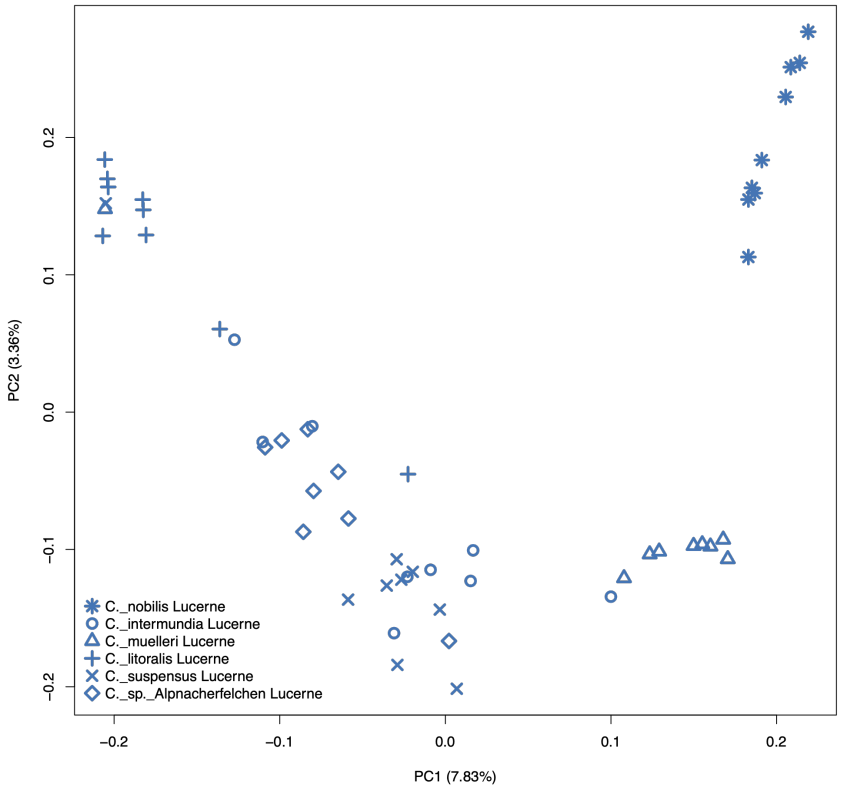
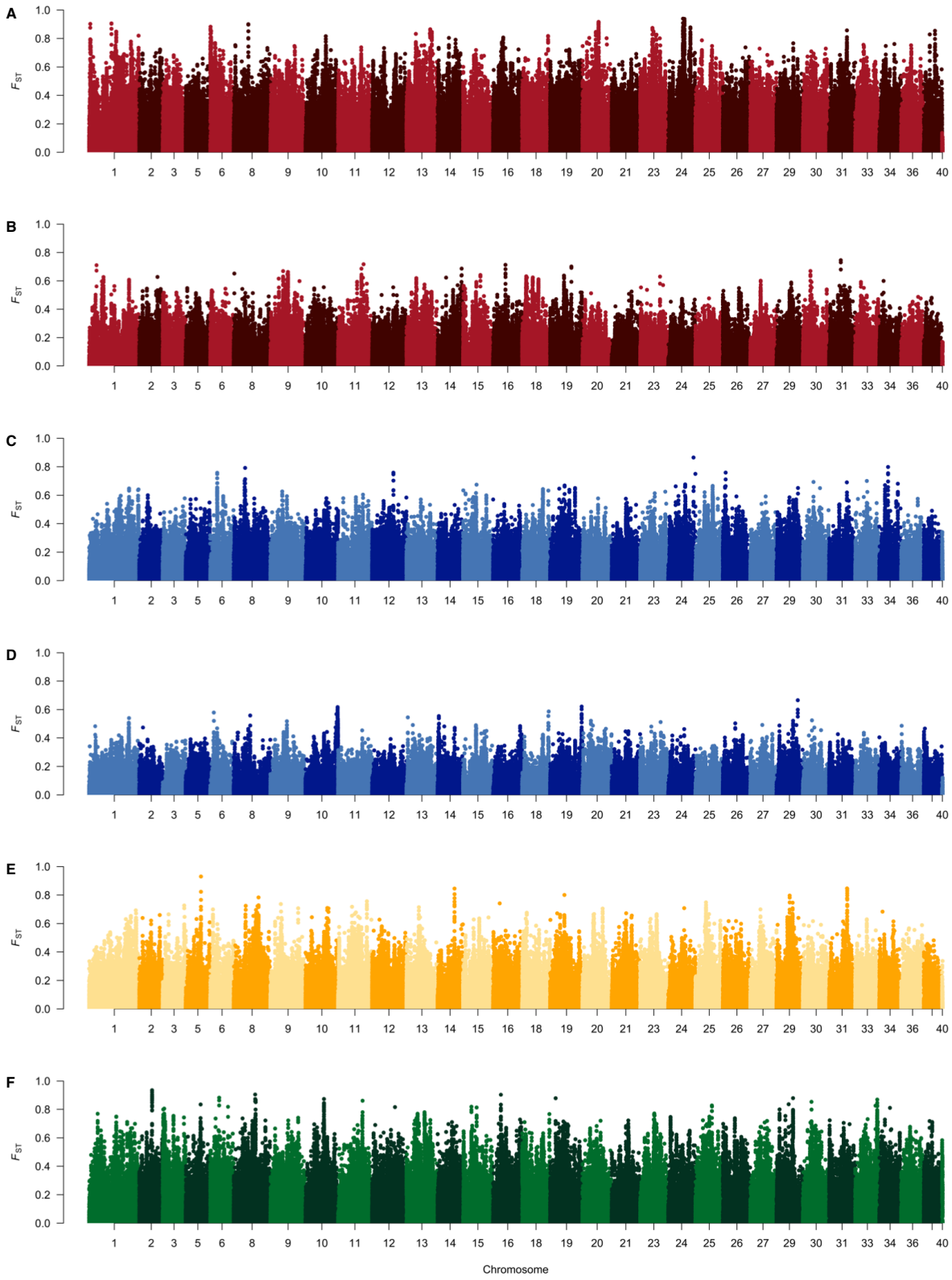
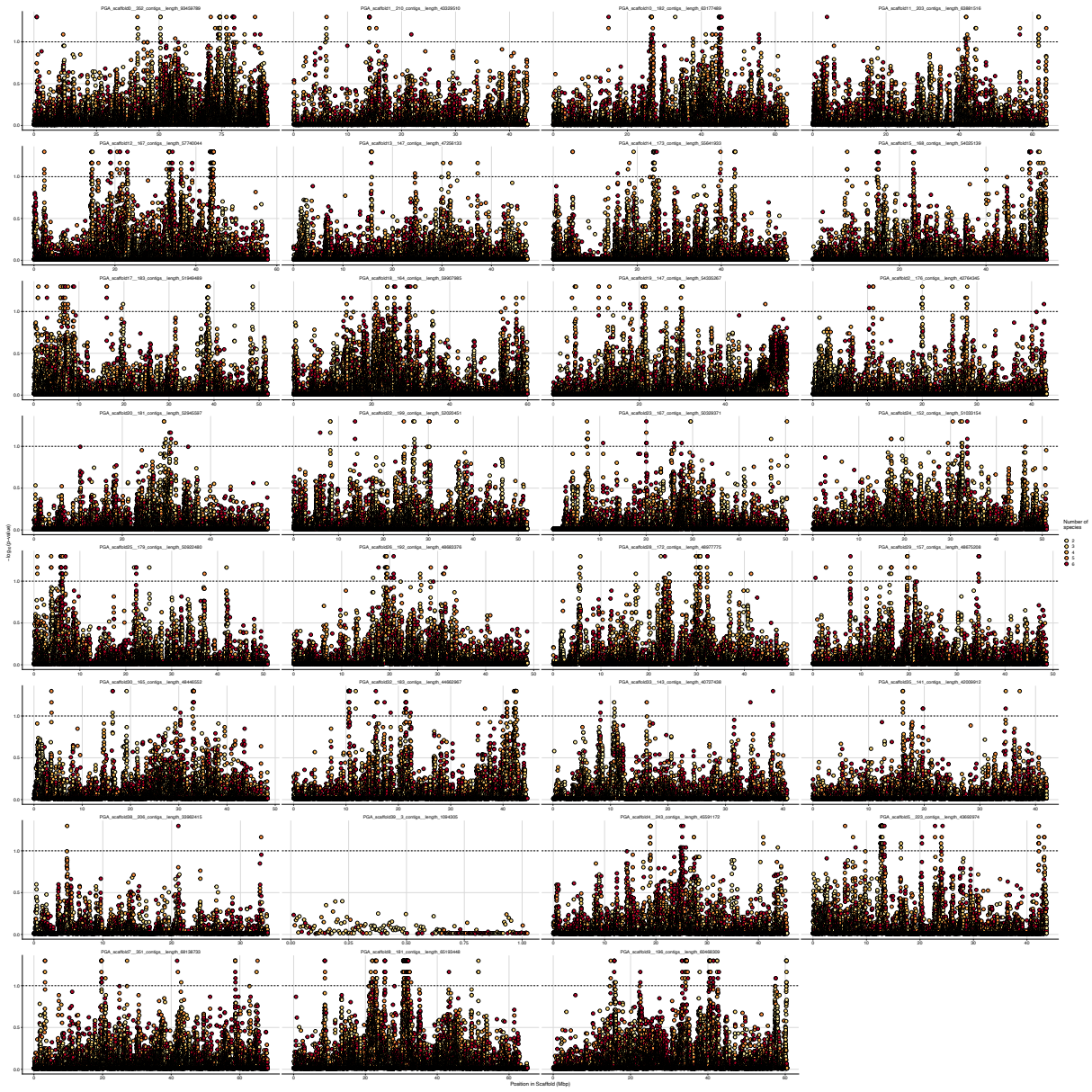


Figure S4: PCA on all individuals from Lake Lucerne. PCA based on genotype likelihoods of 3'237'140 SNPs polymorphic sites, and including all individuals of Lake Lucerne. The plot illustrates the differentiation amongst species (plotting symbol according to species, see legend within the plot).



Supplementary Figure 5: Genomic differentiation of profundal species along the genome. F_{ST} comparisons between profundal and all sympatric non-profundal species of each lake along the genome in 5 kb windows. **A)** *C. profundus* against all non-profundal species of Lake Thun (n=4). **B)** *C. albellus* against all non-profundal species of Lake Thun (n=4). **C)** *C. nobilis* against all non-profundal species of Lake Lucerne (n=4). **D)** *C. muelleri* against all non-profundal species of Lake Lucerne (n=4). **E)** The extinct *C. gutturosus* against all other species of Lake Constance (n=3). **F)** *C. heglingus* against all other species of Lake Walen (n=2). For a list with all species in each Lake see Supplementary Table 1.



Supplementary Figure 6: Parallel evolution along the genome. Manhattan plots showing the PicMin analysis including all six profundal species along all chromosomes. The plot shows fdr-corrected p-values of 5 kb windows along the genome. Colour corresponds to the number of species with a signal for parallel evolution. The dashed line shows the significance threshold (q<0.1).

Supplementary Table 1: Overview over all sequenced samples. Table shows the species assignment, the lab ID number, the lake where the sample has been collected, for which study the sample has been sequenced and tissue type that was used.

Species	Lab ID	Lake	Sequenced by	Tissue
C. acrinasus	123105	Thun	this study	Finclip
C. steinmanni	123117	Thun	this study	Finclip
C. acrinasus	123118	Thun	this study	Finclip
C. alpinus	123131	Thun	this study	Finclip
C. acrinasus	123149	Thun	this study	Finclip
C. steinmanni	123178	Thun	this study	Finclip
C. acrinasus	123201	Thun	this study	Finclip
C. acrinasus	123202	Thun	this study	Finclip
C. alpinus	123231	Thun	this study	Finclip
C. steinmanni	123233	Thun	this study	Finclip
C. albellus	123245	Thun	this study	Finclip
C. acrinasus	123263	Thun	this study	Finclip
C. steinmanni	123264	Thun	this study	Finclip
C. alpinus	123276	Thun	this study	Finclip
C. steinmanni	123288	Thun	this study	Finclip
C. fatioi	123500	Thun	this study	Finclip
C. alpinus	123631	Thun	this study	Finclip
C. profundus	123679	Thun	this study	Finclip
C. profundus	123688	Thun	this study	Finclip
C. profundus	123689	Thun	this study	Finclip
C. albellus	123838	Thun	this study	Finclip
C. fatioi	123840	Thun	this study	Finclip
C. fatioi	123841	Thun	this study	Finclip
C. albellus	123885	Thun	this study	Finclip
C. fatioi	123936	Thun	this study	Finclip
C. albellus	123953	Thun	this study	Finclip
C. albellus	123954	Thun	this study	Finclip
C. albellus	123960	Thun	this study	Finclip
C. profundus	123964	Thun	this study	Finclip
C. profundus	123966	Thun	this study	Finclip
C. profundus	123975	Thun	this study	Finclip
C. fatioi	123990	Thun	this study	Finclip
C. profundus	124003	Thun	this study	Finclip
C. steinmanni	124025	Thun	this study	Finclip
C. alpinus	124029	Thun	this study	Finclip
C. alpinus	124030	Thun	this study	Finclip
C. fatioi	124219	Thun	this study	Finclip
C. profundus	124249	Thun	this study	Finclip
C. heglingus	175718	Walen	this study	Finclip

C. heglingus	175723	Walen	this study	Finclip
C. heglingus	175725	Walen	this study	Finclip
C. zurichensis	175729	Walen	this study	Finclip
C. zurichensis	175731	Walen	this study	Finclip
C. heglingus	175735	Walen	this study	Finclip
C. heglingus	175738	Walen	this study	Finclip
C. heglingus	175742	Walen	this study	Finclip
C. duplex	175801	Walen	this study	Finclip
C. duplex	175802	Walen	this study	Finclip
C. duplex	175806	Walen	this study	Finclip
C. duplex	175807	Walen	this study	Finclip
C. duplex	175809	Walen	this study	Finclip
C. zurichensis	175856	Walen	this study	Finclip
C. zurichensis	175865	Walen	this study	Finclip
C. zurichensis	175868	Walen	this study	Finclip
C. duplex	175870	Walen	this study	Finclip
C. sp. Alpnacherfelchen	26003	Lucerne/Alpnach	this study	Muscle
C. sp. Alpnacherfelchen	26005	Lucerne/Alpnach	this study	Muscle
C. sp. Alpnacherfelchen	26007	Lucerne/Alpnach	this study	Muscle
C. sp. Alpnacherfelchen	26009	Lucerne/Alpnach	this study	Muscle
C. sp. Alpnacherfelchen	26010	Lucerne/Alpnach	this study	Muscle
C. sp. Alpnacherfelchen	26017	Lucerne/Alpnach	this study	Muscle
C. nobilis	26819	Lucerne	this study	Muscle
C. nobilis	26832	Lucerne	this study	Muscle
C. nobilis	26836	Lucerne	this study	Muscle
C. nobilis	26841	Lucerne	this study	Muscle
C. nobilis	26850	Lucerne	this study	Muscle
C. nobilis	26852	Lucerne	this study	Muscle
C. intermundia	28433	Lucerne	this study	Muscle
C. muelleri	28438	Lucerne	this study	Muscle
C. muelleri	28441	Lucerne	this study	Muscle
C. muelleri	28461	Lucerne	this study	Muscle
C. muelleri	28462	Lucerne	this study	Muscle
C. intermundia	28464	Lucerne	this study	Muscle
C. intermundia	28466	Lucerne	this study	Muscle
C. muelleri	28485	Lucerne	this study	Muscle
C. muelleri	28488	Lucerne	this study	Muscle
C. litoralis	28506	Lucerne	this study	Muscle
C. litoralis	28592	Lucerne	this study	Muscle
C. litoralis	28593	Lucerne	this study	Muscle
C. litoralis	28595	Lucerne	this study	Muscle
C. litoralis	28597	Lucerne	this study	Muscle
C. suspensus	28611	Lucerne	this study	Muscle
C. suspensus	28615	Lucerne	this study	Muscle

C. suspensus	28632	Lucerne	this study	Muscle
C. intermundia	28646	Lucerne	this study	Muscle
C. intermundia	28647	Lucerne	this study	Muscle
C. intermundia	28681	Lucerne	this study	Muscle
C. suspensus	28714	Lucerne	this study	Muscle
C. suspensus	28715	Lucerne	this study	Muscle
C. suspensus	28720	Lucerne	this study	Muscle
C. litoralis	28733	Lucerne	this study	Muscle
C. gutturosus	DF_1	Constance	Frei et al. (2022)	Finclip
C. gutturosus	DF_4	Constance	Frei et al. (2022)	Finclip
C. gutturosus	DF_6	Constance	Frei et al. (2022)	Finclip
C. gutturosus	DF_8	Constance	Frei et al. (2022)	Finclip
C. gutturosus	DF_12	Constance	Frei et al. (2022)	Finclip
C. gutturosus	S01_0006	Constance	Frei et al. (2022)	Finclip
C. gutturosus	S01_0020	Constance	Frei et al. (2022)	Finclip
C. gutturosus	S01_0039	Constance	Frei et al. (2022)	Finclip
C. gutturosus	S20_0051	Constance	Frei et al. (2022)	Finclip
C. gutturosus	S20_0079	Constance	Frei et al. (2022)	Finclip
C. arenicolus	126	Constance	Frei et al. (2022)	Finclip
C. arenicolus	127	Constance	Frei et al. (2022)	Finclip
C. arenicolus	128	Constance	Frei et al. (2022)	Finclip
C. arenicolus	212631	Constance	Frei et al. (2022)	Finclip
C. arenicolus	212633	Constance	Frei et al. (2022)	Finclip
C. arenicolus	212634	Constance	Frei et al. (2022)	Finclip
C. arenicolus	212701	Constance	Frei et al. (2022)	Finclip
C. wartmanni	121	Constance	Frei et al. (2022)	Finclip
C. wartmanni	122	Constance	Frei et al. (2022)	Finclip
C. wartmanni	123	Constance	Frei et al. (2022)	Finclip
C. wartmanni	131	Constance	Frei et al. (2022)	Finclip
C. wartmanni	212704	Constance	Frei et al. (2022)	Finclip
C. wartmanni	212705	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	132	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212621	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212639	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212660	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212661	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212673	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212682	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212685	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212686	Constance	Frei et al. (2022)	Finclip
C. macrophthalmus	212689	Constance	Frei et al. (2022)	Finclip
C. acrinasus	12	Thun	De-Kayne et al. (2022)	Finclip
C. acrinasus	13	Thun	De-Kayne et al. (2022)	Finclip
C. acrinasus	14	Thun	De-Kayne et al. (2022)	Finclip

C. alpinus	16	Thun	De-Kayne et al. (2022)	Finclip
C. alpinus	17	Thun	De-Kayne et al. (2022)	Finclip
C. alpinus	19	Thun	De-Kayne et al. (2022)	Finclip
C. fatioi	21	Thun	De-Kayne et al. (2022)	Finclip
C. fatioi	22	Thun	De-Kayne et al. (2022)	Finclip
C. fatioi	23	Thun	De-Kayne et al. (2022)	Finclip
C. profundus	26	Thun	De-Kayne et al. (2022)	Finclip
C. profundus	27	Thun	De-Kayne et al. (2022)	Finclip
C. profundus	29	Thun	De-Kayne et al. (2022)	Finclip
C. intermundia	47	Lucerne	De-Kayne et al. (2022)	Muscle
C. intermundia	48	Lucerne	De-Kayne et al. (2022)	Muscle
C. intermundia	50	Lucerne	De-Kayne et al. (2022)	Muscle
C. steinmanni	56	Thun	De-Kayne et al. (2022)	Finclip
C. steinmanni	58	Thun	De-Kayne et al. (2022)	Finclip
C. steinmanni	59	Thun	De-Kayne et al. (2022)	Finclip
C. albellus	61	Thun	De-Kayne et al. (2022)	Finclip
C. albellus	63	Thun	De-Kayne et al. (2022)	Finclip
C. albellus	64	Thun	De-Kayne et al. (2022)	Finclip
C. sp. Alpnacherfelchen	66	Lucerne/Alpnach	De-Kayne et al. (2022)	Finclip
C. sp. Alpnacherfelchen	68	Lucerne/Alpnach	De-Kayne et al. (2022)	Finclip
C. sp. Alpnacherfelchen	70	Lucerne/Alpnach	De-Kayne et al. (2022)	Finclip
C. duplex	76	Walen	De-Kayne et al. (2022)	Finclip
C. duplex	78	Walen	De-Kayne et al. (2022)	Finclip
C. duplex	79	Walen	De-Kayne et al. (2022)	Finclip
C. heglingus	81	Walen	De-Kayne et al. (2022)	Finclip
C. heglingus	82	Walen	De-Kayne et al. (2022)	Finclip
C. heglingus	83	Walen	De-Kayne et al. (2022)	Finclip
C. duplex	86	Walen	De-Kayne et al. (2022)	Finclip
C. duplex	87	Walen	De-Kayne et al. (2022)	Finclip
C. duplex	90	Walen	De-Kayne et al. (2022)	Finclip
C. nobilis	103	Lucerne	De-Kayne et al. (2022)	Muscle
C. nobilis	104	Lucerne	De-Kayne et al. (2022)	Muscle
C. nobilis	106	Lucerne	De-Kayne et al. (2022)	Muscle
C. suspensus	107	Lucerne	De-Kayne et al. (2022)	Muscle
C. suspensus	108	Lucerne	De-Kayne et al. (2022)	Muscle
C. suspensus	109	Lucerne	De-Kayne et al. (2022)	Muscle
C. muelleri	110	Lucerne	De-Kayne et al. (2022)	Muscle
C. muelleri	111	Lucerne	De-Kayne et al. (2022)	Muscle
C. muelleri	113	Lucerne	De-Kayne et al. (2022)	Muscle
C. litoralis	114	Lucerne	De-Kayne et al. (2022)	Muscle
C. litoralis	118	Lucerne	De-Kayne et al. (2022)	Muscle
C. litoralis	119	Lucerne	De-Kayne et al. (2022)	Muscle

Supplementary Table 2: Functional enrichment of windows with signals of parallel evolution. GO enrichment analysis for all windows that showed signals of parallel evolution in our PicMin analysis including all four lakes. The first column shows the unique GO identifier for each enriched term, the second column gives the respective terminological description, the third column gives the number of genes annotated with the term within the genome, the fourth column gives the number how often the term was represented within windows with a signature of positive selection, the fifth column gives how often the term was expected in those windows by chance, the next column give the p-value using Fisher's exact method based on gene counts, and the last column gives one of the three ontology classes of interest (CC cellular component, BP biological process, MF molecular function) that have been explored.

GO.ID	Term	Annotated	Significant	Expected	P-value	Class
GO:0021678	third ventricle development	7	5	0.76	0.00025	BP
GO:0009648	photoperiodism	5	4	0.54	0.00062	BP
GO:2001056	positive regulation of cysteine-type end...	9	5	0.97	0.00126	BP
GO:0007186	G protein-coupled receptor signaling pat...	78	17	8.41	0.0031	BP
GO:1990504	dense core granule exocytosis	8	4	0.86	0.00659	BP
GO:0070593	dendrite self-avoidance	5	3	0.54	0.01059	BP
GO:0001964	startle response	9	4	0.97	0.01086	BP
GO:0098038	non-replicative transposition, DNA-media...	2	2	0.22	0.01163	BP
GO:0032222	regulation of synaptic transmission, cho...	2	2	0.22	0.01163	BP
GO:0048149	behavioral response to ethanol	2	2	0.22	0.01163	BP
GO:0005220	inositol 1,4,5-trisphosphate-sensitive c...	7	5	0.68	0.00016	MF
GO:0005217	intracellular ligand-gated ion channel a...	26	9	2.54	0.00052	MF
GO:0004991	parathyroid hormone receptor activity	3	3	0.29	0.00093	MF
GO:0005451	monovalent cation:proton antiporter acti...	12	5	1.17	0.00387	MF
GO:0015385	sodium:proton antiporter activity	12	5	1.17	0.00387	MF
GO:0051139	metal ion:proton antiporter activity	12	5	1.17	0.00387	MF
GO:0015491	cation:cation antiporter activity	40	10	3.9	0.00419	MF
GO:0015298	solute:cation antiporter activity	41	10	4	0.00506	MF
GO:0008289	lipid binding	206	32	20.1	0.0055	MF
GO:0046875	ephrin receptor binding	18	6	1.76	0.00565	MF
GO:0030017	sarcomere	35	2	3.22	0.85	CC
GO:0005737	cytoplasm	970	86	89.23	0.67	CC
GO:0045261	proton-transporting ATP synthase complex...	6	0	0.55	1	CC
GO:0061689	tricellular tight junction	5	0	0.46	1	CC
GO:0005773	vacuole	51	2	4.69	0.96	CC
GO:0033176	proton-transporting V-type ATPase comple...	28	2	2.58	0.74	CC
GO:0015934	large ribosomal subunit	34	2	3.13	0.83	CC
GO:0030667	secretory granule membrane	9	0	0.83	1	CC
GO:1990716	axonemal central apparatus	3	0	0.28	1	CC
GO:0042588	zymogen granule	8	0	0.74	1	CC

Chapter V

V. Sequencing platform shifts provide opportunities but pose challenges for combining genomic datasets

Rishi De-Kayne[1,2†], David Frei[1,2,†], Ryan Greenway[1], Sofia L. Mendes[3], Cas Retel[1], Philine G.D. Feulner[1,2]

[†] These authors contributed equally

[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

[3]Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

Published as:
De-Kanye, R., Frei, D., Greenway, R., Mendes, S.L., Retel, C., Feulner, P.G.D. (2020). Sequencing platform shifts provide opportunities but pose challenges for combining genomic datasets. *Molecular Ecology Resources*, *21*(3), 653-660. doi:10.1111/1755-0998.13309

Technological advances in DNA sequencing over the last decade now permit the production and curation of large genomic datasets in an increasing number of non-model species. Additionally, this new data provides the opportunity for combining datasets, resulting in larger studies with a broader taxonomic range. Whilst the development of new sequencing platforms has been beneficial, resulting in a higher throughput of data at a lower per-base cost, shifts in sequencing technology can also pose challenges for those wishing to combine new sequencing data with data sequenced on older platforms. Here, we outline the types of studies where the use of curated data might be beneficial, and highlight potential biases that might be introduced by combining data from different sequencing platforms. As an example of the challenges associated with combining data across sequencing platforms, we focus on the impact of the shift in Illumina's base calling technology from a four-channel to a two-channel system. We caution that when data is combined from these two systems, erroneous guanine base calls that result from the two-channel chemistry can make their way through a bioinformatic pipeline, eventually leading to inaccurate and potentially misleading conclusions. We also suggest solutions for dealing with such potential artifacts, which make samples sequenced on different sequencing platforms appear

more differentiated from one another than they really are. Finally, we stress the importance of archiving tissue samples and the associated sequences for the continued reproducibility and reusability of sequencing data in the face of ever-changing sequencing platform technology.

Opportunities: Combining and extending datasets across time and space

DNA sequencing data reflecting the diversity of life is accumulating, as technological developments continue to increase the basepair yield of sequencing runs, whilst lowering the per-basepair prices. This data continues to facilitate comparative studies of genome structure for more and more organisms, spanning the tree of life (Baker et al., 2020; Cheng et al., 2018; Leebens-Mack et al., 2019; Morris et al., 2018; Peter et al., 2018; Shen et al., 2018; Shi et al., 2018; Zhang et al., 2014). Further, the field of molecular ecology is flourishing, with more and more studies investigating the genetic variation within and among closely related groups of organisms (Brawand et al., 2014; Lamichhaney et al., 2015; Tollis et al., 2018). However, for molecular ecologists working on non-model species, budgets still limit the amount of sequence data that can be produced. As a result, exhaustive experimental designs, which include the sampling of many individuals from many different populations, are rare (but are emerging; (Feulner et al., 2015; Greenway et al., 2020; Martin et al., 2016; Soria-Carrasco et al., 2014; Stankowski et al., 2019; Vijay et al., 2016)). The effort to publicly archive sequence data that has already contributed to publications helps to maintain the reproducibility of sequencing studies, whilst prolonging the value of such sequence data in perpetuity. Additionally, this practice of sequence data storage provides the opportunity to expand datasets beyond those that one laboratory is capable of producing (in terms of time, labour, and finances) to increase the impact of studies despite a potentially limited budget. Repositories like the Short Read Archive (SRA) -- part of the International Nucleotide Sequence Database Collaboration (INSDC) that includes the NCBI Sequence Read Archive (SRA), the European Bioinformatics Institute (EBI), and the DNA Database of Japan (DDBJ) -- are essential for both the reproducibility of genetic and genomic studies, and the reusability of sequencing data. Although combining datasets is challenging for many sequencing approaches, particularly those that sequenced anonymous reduced representations of the

genome (i.e. microsatellites, amplified fragment length polymorphisms, and maybe even restriction site associated DNA sequencing and genotyping by sequencing; but see Leigh, Lischer, Grossen, & Keller (2018) for an example), the increasingly common approach of re-sequencing whole-genomes (even for a broader range of non-model organisms) makes the possibility of combining datasets more inviting.

Between the continued growth of sequencing data repositories and the continued ability to sequence more DNA quicker and cheaper the following types of studies are increasingly carried out:

(1) Broad macroevolutionary studies. Typically, such macroevolutionary studies benefit from a wide taxon sampling and few individuals suffice, making the combination of samples from different published datasets particularly useful. Often these analyses are restricted to more conserved regions of the genome. For example, Zhang et al. (2020) compiled a comprehensive dataset of 365 species of asterids representing all 17 orders containing published and newly sequenced whole genomes and transcriptomes to resolve the deep asterid phylogeny. In another example, Greenway et al. (2020) focus on the Poeciliidae family of fish, to demonstrate that adaptation to extreme, here sulfide-rich, environments has evolved convergently in ten independent lineages, by combining already published and newly sequenced transcriptome sequences.

(2) Microevolutionary studies investigating spatial variation across populations or closely related taxa. Such studies typically focus on one study system but rely on a larger sampling to reflect the variation within species or populations. These studies may benefit from combining newly sequenced material with archived sequence data from previous projects to produce larger within-system datasets. By taking advantage of existing sequence data, these combined datasets facilitate analyses of genomic differentiation across a much

broader geographic sampling or among more individuals than would be otherwise possible. Here, the curated data is used to evaluate patterns in comparable populations to widen the perspective, i.e. to show whether a pattern is general or specific to the population under investigation. For example, Ravinet, Kume, Ishikawa, & Kitano (2020) evaluated if patterns of divergence and introgression between Japan Sea and Pacific Ocean stickleback resemble patterns at other locations where these species co-occur. In a comprehensive study conducted by Samuk et al. (2017), the authors compiled multiple genotyping by sequencing and whole genome sequencing datasets to a global evaluation of 1300 stickleback individuals across 51 populations, to show that putative adaptive alleles tend to occur more often in regions of low recombination. Bergland, Behrman, O'Brien, Schmidt, & Petrov (2014) used curated data to check haplotypes under seasonal selection in *Drosophila melanogaster* for between-species divergence with a sister species (*D. simulans*). Most recently, Jones, Mills, Jensen, & Good (2020) combined new and published whole-genome and exome sequences with targeted genotyping of *Agouti*, a pigmentation gene introgressed from black-tailed jackrabbits, to investigate the evolutionary history of local seasonal camouflage adaptation in Snowshoe hares from the Pacific Northwest.

(3) Studies investigating temporal variation within and between population and species. Such studies involve combining datasets across time scales and often contain sequencing data that originated from a variety of sample types including museum collections, long-term preserved fossils or hard tissues, and contemporary fresh samples. For example, the use of museum specimens facilitated the investigation of independent temporal genomic contrasts spanning a century of climate change for two co-distributed chipmunk species (Bi et al., 2019) and a paleogenomics approach investigated the temporal component of adaptation to freshwater in sticklebacks by sequencing the genomes of 11-13,000-year-old bones and comparing them with 30 modern stickleback genomes (Kirch, Romundset, Gilbert, Jones, &

Foote, 2020). Experimental approaches combining previous sequencing efforts with new samples are also commonly used to increase our understanding of temporal variation. Tenaillon et al. (2016) compiled sequence data from several other publications in addition to new sequences to strengthen their conclusions on the tempo and mode of *E. coli* genome evolution. Bottery, Wood, & Brockhurst (2019), after having shown that tetracycline resistance requires multiple mutations, used curated data to investigate if the mutation establishment order was repeatable. This by no means exhaustive selection of examples highlights that the growing amount of sequence data provides the opportunity for endless combinations of datasets to be analysed to address a multitude of questions.

Challenges: Biases change with technological developments

One technological advance which sped up the Illumina workflow and made it more cost-effective was a change from four-channel chemistry, where each of the four DNA bases is detected by a different fluorescent dye, to a two-channel chemistry, that uses only two different fluorescent dyes (Illumina). In these two-channel workflows, as implemented in the NextSeq and NovaSeq platforms, a guanine base (G) is called in the absence of fluorescence (Figure 1). Hence, it is difficult to differentiate between no signal and a G, resulting in an overrepresentation of poly-G strings in sequence data from both NextSeq and NovaSeq (Chen, Zhou, Chen, & Gu, 2018).

To most accurately capture biological variation in a given sample or population, it is important to differentiate between potentially erroneous and correct base calls, which is often done using base quality scores. However, erroneous poly-G base calls produced on the NextSeq and NovaSeq platforms can be difficult to detect, because, as a result of the two-colour chemistry, they are not always associated with reduced base qualities. Unfortunately, read trimming software packages that were written for the older four-colour systems do not

flag or trim poly-G tails. Although one might think that mapping should remove the effect of these overrepresented Gs without the need for read trimming, it has been shown that some may still trickle through a bioinformatics pipeline and influence variant calling steps. A comprehensive empirical study making use of cancer cell lines to benchmark systematic differences between technologies revealed that NovaSeq instruments produced more stretches of Gs than HiSeqX in both paired-end reads (Arora et al., 2019). Arora et al. (2019) further confirmed that the bias remained detectable in the mapped reads and resulted in a relatively large number of T > G mutations among the variants unique to the NovaSeq instrument. To reduce the potential down-stream impact of these poly-G strings, newer trimming software packages such as fastp (Chen et al., 2018) check the source of the data and implement poly-G trimming by default for the two-colour systems. This not only improves the computational efficiency of sequence alignment, but should also reduce the impact of erroneous variant calling on these bases.

The impact of these changes in base calling and the subsequent erroneous G calls on the biological interpretation may vary with the chosen experimental design and other sources of variation such as for example DNA quality. Although the biases resulting from not trimming off or filtering out poly-G strings might be mild or irrelevant when analysing data produced from high quality input DNA from a single system, this may not be true when data from different technologies are combined across various biological units (e.g. across populations, species, treatments, or time points). On top of variation in the quality of input DNA, a range of variation in sequencing approaches exists, along with differences in library preparation, including variation in read length or whether reads are single-end or paired-end. Where different individuals within a single dataset have been sequenced with variation in these methodological factors biases may also be exacerbated, potentially producing misleading results. Variation in length of sequences reads across a dataset for example has

been shown to lead to pronounced allele frequency differences between populations and subsequently suggested false biological trends (Leight et al. 2018). Metagenomic work suggested that both library preparation and sequencing platform had systematic effects on the microbial community description (Poulsen, Pamp, Ekstrøm, & Aarestrup, 2019; Sato et al., 2019). In summary, attention should be paid to DNA quality, library preparation protocols, and the sequencing platform used when analysing and interpreting publicly available genomic data.

Although the prospect of combining datasets to improve our power to detect patterns is alluring, it is important to consider the ways in which these data may result in misleading conclusions. Combining datasets often means combining data from different sequencing platforms, as DNA sequencing technology continues to develop through time. Unfortunately, some of the developments (e.g. the change from four-channel to two-channel chemistry in Illumina sequencing machines) have changed the way in which uncertainties in base calling are presented in the sequencer's output files. If managed incorrectly, these changes hamper our ability to combine datasets obtained with different sequencing technologies, and the subsequent genotyping and analysis of these combined datasets may be biased (in the worst cases leading to erroneous conclusions). The most straightforward way to prevent this is a well-thought out experimental design, a step which can often be overlooked in a time where sequencing data is being produced so rapidly (see Mason (2017) for sound advice on experimental design). As has been shown for sequencing reduced-representation libraries, it is crucial for any type of sequencing experiment to carefully consider types of errors that may be introduced during laboratory work and data processing, and how to minimize, detect and remove these errors (O'Leary, Puritz, Willis, Hollenbeck, & Portnoy 2018). However, it may be difficult to achieve the ideal or optimal study design when an investigation integrates new information with already existing data (e.g. with individuals and treatments randomised

across sequencing batches). Despite this limitation there are a number of approaches that can help to rectify some of these imbalances and allow the combination of multiple genomic datasets whilst minimising the impact of cross-platform biases.

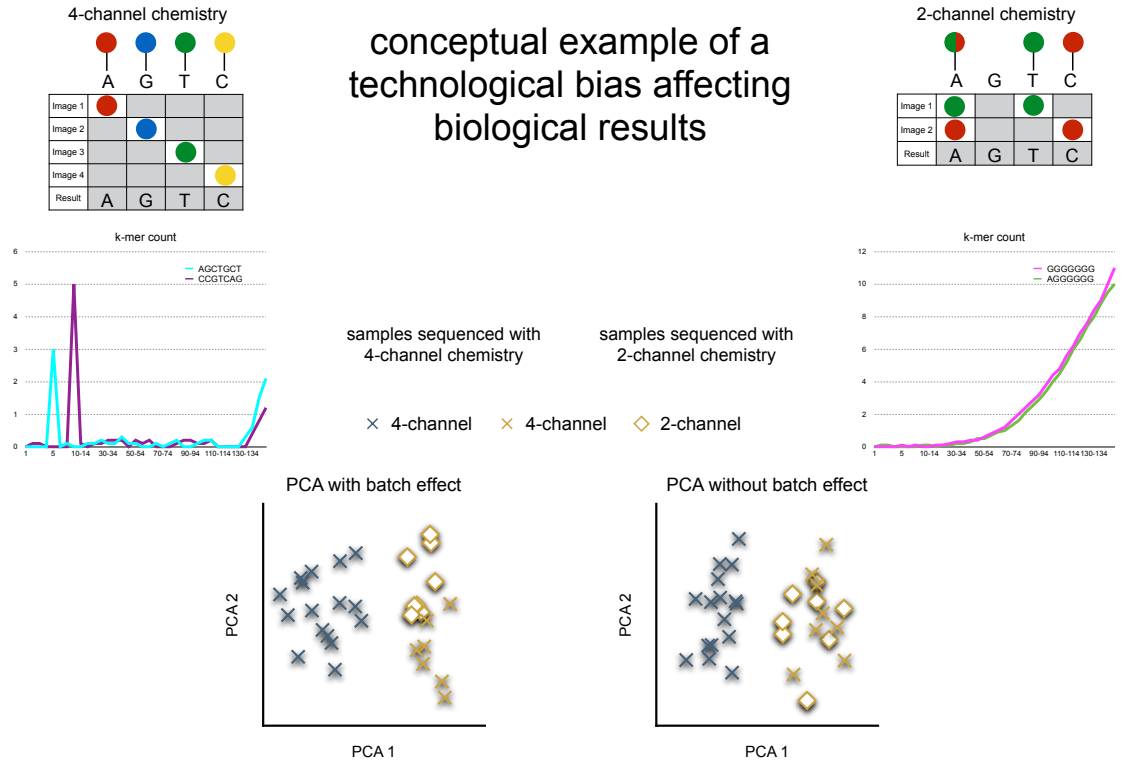


Figure 1: Example of a technological difference between sequencing chemistries, which introduces a bias (overrepresentation of G k-mers) in the sequenced reads and result in a batch effect visible when genotypes are evaluated in a principal component analysis (PCA).
Top: Schematic redrawn from Illumina representing the differences between 4-channel chemistry evaluating each of the four bases by a distinct fluorescence label, and 2-channel chemistry representing the four bases with two dyes only.
Middle: Redrawn examples of the one aspect of a typical FastQC (Andrews, 2010) report, which evaluates the count of each short nucleotide of length k (default = 7) starting at each position along the read. Any given k-mer should be evenly represented across the length of the read. The y axis reports the relative enrichment (log2 observed over expected counts) of the 7-mers over the read length (x axis). The graph presents those k-mers which appear at specific positions with greater than expected frequency. In the left panel reads sequenced with 4-channel chemistry are represented which show a slight overrepresentation of two random 7-mers represented by different colours (typically the report would plot the first six hits). The overrepresentation is small and most pronounced at the beginning of the read (to the left of the x axis), a pattern often found in high quality sequencing libraries due to slight, sequence dependent efficiency of DNA shearing or a result of random priming. In the right panel, an overrepresentation of poly-G-mers toward the end of the reads is exemplified as typical for raw reads sequenced with 2-channel chemistry. Note the difference in the logarithmic scale between left and right panel.
Bottom: Conceptual representation of a batch effect resulting from technological differences. Each sample's genotype, compiled of a large number of loci distributed across the whole genome, is represented as a coloured symbol in multivariate space, where PC axis one and two reflect two primary axes of variation in the dataset. The left panel would reflect a dataset with a batch effect. The fact that samples are separated by sequencing technology on PC axis 2 indicates the presence of a technological bias. In the right panel, batch effects have been reduced, e.g. by trimming off poly-G tails. Symbols in the PCA differentiate samples sequenced with either 2-channel (diamond) or 4-channel (cross) chemistry, colours differentiate different populations or species (biological differences). The left panel is imagined to be based on a data set of untrimmed reads, PC axis 2 separates samples due to technological differences. That effect is gone in the right panel, after read trimming was applied.

Ways forward: Suggestions on how to minimise technological bias when integrating datasets

Despite the ease with which new datasets can be produced it is critical that researchers do not forgo project planning and experimental design steps and aim to understand and reduce the potential impact of intrinsic data biases. These planning steps should be similar to those carried out for the sequencing of new samples and could include an assessment of the dataset (1) and the pipeline for analysis (2):

(1) When compiling a combined dataset, it is important to consider the key question that is being addressed and to evaluate how many samples of each population, species, treatment, or time unit are needed to have the power to draw meaningful conclusions. It is also worth evaluating the trade-offs between sequencing new samples or using existing data (e.g. if only a handful of samples are missing could it be worthwhile to sequence more samples so that all individuals are sequenced the same way, reducing the likelihood that biases or batch effects will cause problems downstream in the analysis). If datasets will be combined to address a specific question then it is important to asses which specific sequenced samples are available and how many different datasets these samples come from. It is important to be conscious of, and carefully document, the different technologies used for library preparation and sequencing across samples and datasets, and if possible, to glean an understanding of the origin and quality of the input DNA. Ideally, the dataset would be compiled in a way that minimizes the number of differences between samples from different sources. Further, it is critical to strive to randomise samples from different biological units across different sequencing batches (Meirmans 2015). It can be particularly beneficial to repeat sequencing of one or a few representatives from a curated dataset to evaluate and correct potential biases. If feasible, repeated sequencing of the same individual allows to identify problematic loci that are not genotyped identically or consistently across technologies

despite originating from the same individual. We therefore urge researchers wherever possible to archive tissue and/or DNA. These collections can be of tremendous value, as they facilitate the repeated sequencing of past samples into newly compiled datasets to determine whether any variants or alleles may have been erroneously missed because of technological biases. Using archived tissue or DNA in this way is one of the only possibilities to verify new sequence variants found using future technologies.

(2) Once it is decided that integrating dataset from various sources provides the best power to answer a particular question, it is important to determine which checks should be implemented in the analysis pipeline to avoid misleading biological interpretation of the data. The ways in which biological and technological differences are distributed across the compiled dataset should be reported and critical steps that would identify potentially problematic sequence artifacts and biases should be implemented in the bioinformatic pipeline. It is also crucial to determine how potential artifacts and biases amongst datasets will be handled. Figure 2 provides a suggestion for a pipeline evaluating known differences between sequencing data produced with four-channel chemistry (e.g. HiSeqX) and two-channel chemistry (e.g. NovaSeq). We suggest comparing the FastQC report (Andrews, 2010) between samples sequenced with the two technologies to each other. Any systematic difference across FastQC reports might be relevant, however, when samples sequenced with different sequence chemistry that affects the base calling are combined reports on per base sequence and k-mers content are particularly worth paying attention to (see Figure 1 for an example, illustrating differences in k-mer counts). To see whether mapping reduces sequencing artefacts, FastQC can be re-run on only the reads that mapped well and will be used for genotyping. If biases persist, read trimming should be considered. Here fastp (Chen et al., 2018) could be used to trim poly-G tails efficiently. Once reads have been mapped, variants have been called, and genotypes have been determined, genotypes should be

evaluated for potential batch effects. Here, we recommend identifying individuals sampled using different datasets and/or technologies with specific symbols or colours allowing the possible differences between these artificial groups to be highlighted (see section above). For example, in a Principal Component Analysis (PCA) which represents the various technological and sample differences by different symbols and biological differences (i.e. populations or species) by colour, any PC axis separating symbols instead of colours suggests there might be some technological bias causing batch effects (Figure 1). However, biases might not always show up as batch effects and are especially problematic when one population or other biological unit is the only one sequenced with a different technology. In this scenario, artifacts and biological differences would be confounded and as a result artifacts and biases would be hard to detect (not visible as a batch effect in a PCA) and correct for. For this reason, we suggest that researchers aim to sequence biological units (species, populations, treatments, or time points) across each batch to avoid confounding biological differences with library or other technical effects. Alternatively, a bias might (although not necessarily) show up as a mutational bias relative to the reference, which can be evaluated and compared to published biases resulting from sequencing platform shifts (see Arora et al. (2019)). To reduce biases and undesired batch effects, the filtering parameters for variant calls and genotypes will need to be adjusted. One way to find the optimal filtering settings could be to determine which filtering thresholds allow you to minimize the differences between the detected batches. Specifically, it may be useful to compare distributions of quality scores between reference and alternate allele, which should look very similar in the absence of batch effects. However, we do not recommend solely relying on this to remove biases in the reads

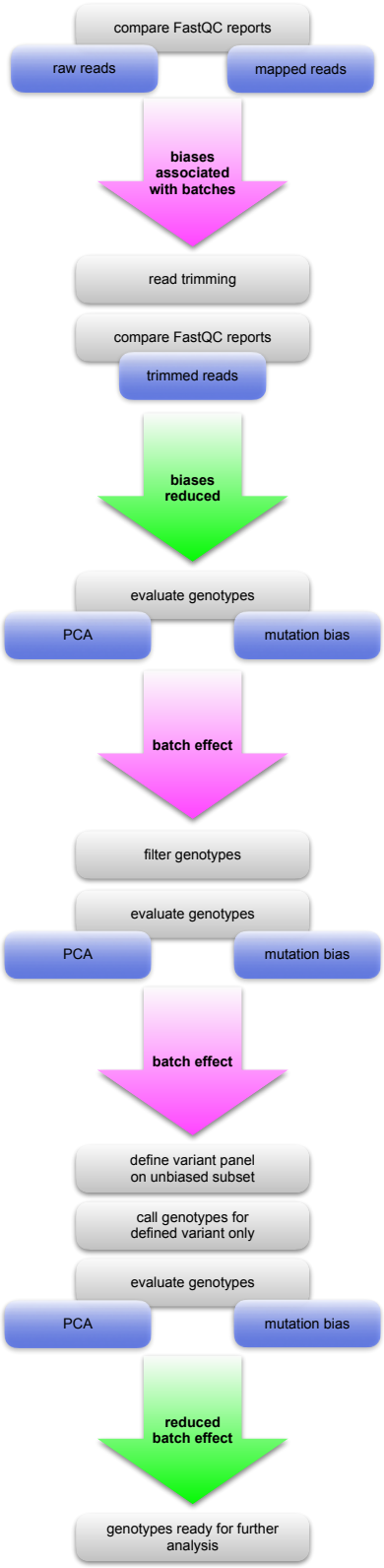


Figure 2: Flow diagram of an exemplified pipeline evaluating and accounting for biases caused by different sequencing technologies in a compiled data set. For more details see text.

(such as poly-Gs in NovaSeq data) but mention this as one option that might help to reduce other sources of undesired batch effects. If none of these approaches suffice to identify and remove biases, one potential solution could be to define variable sites in a subset of the data, which only represents one technology, and then call genotypes on the whole dataset for only those regions. This comes with a potential ascertainment bias depending on how broadly biological units are represented in such a subset, but should reduce spurious variation caused by technological differences. Such an approach is similar to defining a SNP panel and then using SNPchips or other technologies to genotype a larger sampling (Kim et al., 2018). As all datasets are different, different approaches might be needed to reduce any effects of technological differences in compiled datasets. Critically, in each of these scenarios the identification and removal of biases associated with technological shifts serves to reduce the possibility of incorrectly or erroneously inferring biological patterns or processes.

Finally, we want to emphasise the huge value of community efforts to archive sequencing data that makes science reproducible and reusable. We hope that we have demonstrated not only how technological shifts may pose challenges for the meaningful reusability of data, but also that the removal of biases associated with such shifts allows us to address new and exciting biological questions. We highlight the importance and value of accurate documentation, archiving of tissue and DNA samples, and sequence data, and urge researchers to assess the experimental design of their research projects to ensure scientifically sound and robust results.

Acknowledgements

We thank David Marques and the Fish Ecology and Evolution group at Eawag for their helpful comments and fruitful discussions on the topic. We are grateful to three reviewers for their constructive comments on the manuscript.

RD is supported by an SNSF grant (31003A_163446) awarded to PF. DF is 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. D.F. received additional financial support from the Swiss Federal Office for the Environment and Eawag (Eawag Discretionary Funds 2018-2022). SLM is supported by an FCT scholarship (SFRH/BD/145153/2019) granted by the Portuguese National Science Foundation (Fundação para a Ciência e a Tecnologia – FCT). The funders had no role in study design, decision to publish, or preparation of the manuscript.

Author's contributions

RD, DF, and PF conceived of the presented ideas based on the experience and insights of DF. RD and PF drafted the manuscript. PF drafted the figures. All authors contributed to the discussion and critical revision of the final manuscript.

Competing interest declaration

The authors declare no competing interests.

References

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arora, K., Shah, M., Johnson, M., Sanghvi, R., Shelton, J., Nagulapalli, K., . . . Robine, N. (2019). Deep whole-genome sequencing of 3 cancer cell lines on 2 sequencing platforms. *Scientific Reports, 9*, 19123. doi:10.1038/s41598-019-55636-3

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., & Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nature Microbiology, 5*(7), 887-900. doi:10.1038/s41564-020-0715-z

Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *Plos Genetics*, 10(11), e1004775. doi:10.1371/journal.pgen.1004775

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., . . . Good, J. M. (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *Plos Genetics*, 15(5), e1008119. doi:10.1371/journal.pgen.1008119

Bottery, M. J., Wood, A. J., & Brockhurst, M. A. (2019). Temporal dynamics of bacteria-plasmid coevolution under antibiotic selection. *Isme Journal, 13*(2), 559-562. doi:10.1038/s41396-018-0276-9

Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., . . . Di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature, 513*(7518), 375-381. doi:10.1038/nature13726

Chen, S. F., Zhou, Y. Q., Chen, Y. R., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), 884-890. doi:10.1093/bioinformatics/bty560

Cheng, S. F., Melkonian, M., Smith, S. A., Brockington, S., Archibald, J. M., Delaux, P. M., . . . Wong, G. K. S. (2018). 10KP: A phylodiverse genome sequencing plan. *Gigascience, 7*(3). doi:10.1093/gigascience/giy013

Feulner, P. G. D., Chain, F. J. J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., . . . Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *Plos Genetics*, 11(2), e1005414. doi:10.1371/journal.pgen.1004966

Greenway, R., Barts, N., Henpita, C., Brown, A. P., Rodriguez, L. A., Pena, C. M. R., . . . Shaw, J. H. (2020). Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. *Proceedings of the National Academy of Sciences of the United States of America, 117*(28), 16424-16430. doi:10.1073/pnas.2004223117

Jones, M. R., Mills, L. S., Jensen, J. D., & Good, J. M. (2020). The origin and spread of locally adaptive seasonal camouflage in snowshoe hares. *American Naturalist, 196*(3), 316-332. doi:10.1086/710022

Kim, J. M., Santure, A. W., Barton, H. J., Quinn, J. L., Cole, E. F., Visser, M. E., . . . Great Tit HapMap, C. (2018). A high-density SNP chip for genotyping great tit (*Parus major*) populations and its application to studying the genetic architecture of exploration behaviour. *Molecular Ecology Resources, 18*(4), 877-891. doi:10.1111/1755-0998.12778

Kirch, M., Romundset, A., Gilbert, M. T. P., Jones, F. C., & Foote, A. D. (2020). Pleistocene stickleback genomes reveal the constraints on parallel evolution. *bioRxiv*, 2020.2008.2012.248427. doi:10.1101/2020.08.12.248427

Lamichhaney, S., Berglund, J., Almen, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., . . . Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature, 518*(7539), 371-375. doi:10.1038/nature14181

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., . . . One Thousand Plant, T. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature, 574*(7780), 679-685. doi:10.1038/s41586-019-1693-2

Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Molecular Ecology Resources* **18**: 778-788. doi: 10.1111/1755-0998.12779

Marques, D. A., Lucek, K., Sousa, V. C., Excoffier, L., & Seehausen, O. (2019). Admixture between old lineages facilitated contemporary ecological speciation in Lake Constance stickleback. *Nature Communications, 10*, 4240. doi:10.1038/s41467-019-12182-w

Martin, S. H., Most, M., Palmer, W. J., Salazar, C., McMillan, W. O., Jiggins, F. M., & Jiggins, C. D. (2016). Natural selection and genetic diversity in the butterfly *Heliconius melpomene*. *Genetics, 203*(1), 525-541. doi:10.1534/genetics.115.183285

Mason, C. C. (2017). Four study design principles for genetic investigations using next generation sequencing. *Bmj-British Medical Journal*, 359, j4069. doi:10.1136/bmj.j4069

Meirmans, P. G. (2015). Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology* 24: 3223-3231. doi: 10.1111/mec.13243

Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., . . . Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America, 115*(10), E2274-E2283. doi:10.1073/pnas.1719588115

O'Leary Shannon, J., Puritz Jonathan, B., Willis Stuart, C., Hollenbeck Christopher, M., & Portnoy David, S. (2018). These aren't the loci you're looking for: Principles of

effective SNP filtering for molecular ecologists. *Molecular Ecology* 27:3193–3206. doi: 10.1111/mec.14792

Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergstrom, A., . . . Schacherer, J. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature, 556*(7701), 339-344. doi:10.1038/s41586-018-0030-5

Poulsen, C. S., Pamp, S. J., Ekstrøm, C. T., & Aarestrup, F. M. (2019). Library preparation and sequencing platform introduce bias in metagenomics characterisation of microbial communities. *bioRxiv*, 592154. doi:10.1101/592154

Ravinet, M., Kume, M., Ishikawa, A., & Kitano, J. Patterns of genomic divergence and introgression between Japanese stickleback species with overlapping breeding habitats. *Journal of Evolutionary Biology, 00,* 1-14. doi:10.1111/jeb.13664

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology, 26*(17), 4378-4390. doi:10.1111/mec.14226

Sato, M. P., Ogura, Y., Nakamura, K., Nishida, R., Gotoh, Y., Hayashi, M., . . . Hayashi, T. (2019). Comparison of *the* sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. *DNA Research, 26*(5), 391-398. doi:10.1093/dnares/dsz017

Shen, X. X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., . . . Rokas, A. (2018). Tempo and mode of genome evolution in the budding yeast subphylum. *Cell, 175*(6), 1533-1545. doi:10.1016/j.cell.2018.10.023

Shi, M., Lin, X. D., Chen, X., Tian, J. H., Chen, L. J., Li, K., . . . Zhang, Y. Z. (2018). The evolutionary history of vertebrate RNA viruses. *Nature, 561*(7722), E6. doi:10.1038/s41586-018-0310-0

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., . . . Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science, 344*(6185), 738-742. doi:10.1126/science.1252136

Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *Plos Biology*, 17(7), e3000391. doi:10.1371/journal.pbio.3000391

Tenaillon, O., Barrick, J. E., Ribeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., . . . Lenski, R. E. (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature, 536*(7615), 165-170. doi:10.1038/nature18959

Tollis, M., Hutchins, E. D., Stapley, J., Rupp, S. M., Eckalbar, W. L., Maayan, I., . . . Kusumi, K. (2018). Comparative genomics reveals accelerated evolution in conserved pathways during the diversification of anole lizards. *Genome Biology and Evolution, 10*(2), 489-506. doi:10.1093/gbe/evy013

Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications, 7*, 10. doi:10.1038/ncomms13195

Zhang, G. J., Li, C., Li, Q. Y., Li, B., Larkin, D. M., Lee, C., . . . Avian Genome, C. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science, 346*(6215), 1311-1320. doi:10.1126/science.1251385

Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., . . . Ma, H. (2020). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Molecular Biology and Evolution*. doi:10.1093/molbev/msaa160

Synthesis

Synthesis

Genomic consequences of eutrophication induced speciation reversal

When biodiversity evolves as a consequence of adaptation to divergent selection between different habitats in a heterogenous environments, the homogenization of the environment through environmental change can result in the loss of biodiversity (Seehausen, van Alphen, & Witte, 1997). Throughout the last decades, the focus of many conservation efforts was on mitigating demographic consequences of environmental change, such as maximizing available habitats through restoration efforts (Seehausen, Takimoto, Roy, & Jokela, 2008). However, when environmental change affects features of the environment that are coupled to ecologically-mediated reproductive isolation between sympatric species, reproductive isolation between species can be weakened, resulting in hybridization that erodes species differentiation (Seehausen, 2006; Seehausen et al., 1997; Vonlanthen et al., 2012). Such a situation often coincides with a declining population size as a consequence of habitat degradation, and hence, dramatically accelerates the process of biodiversity loss and extinction (Seehausen et al., 2008). In the context of evolutionary young adaptive radiations, environmental change can reduce reproductive isolation between numerous species at the same time (Seehausen et al., 1997; Vonlanthen et al., 2012). The resulting hybridization and introgression can lead to the collapse of the radiation into a hybrid swarm within only few generations (Seehausen et al., 1997; Vonlanthen et al., 2012). The contemporary high rates of deforestation, desertification, urbanization, water extraction and eutrophication increase habitat fragmentation and homogenize natural habitats (Fahrig, 2003; Grabenstein & Taylor, 2018; Haddad et al., 2015). Hence, understanding the consequences of anthropogenic environmental disturbance and their effect on evolutionary forces maintaining reproductive

isolation between coexisting species is fundamental for the conservation of contemporary biodiversity (Grabenstein & Taylor, 2018).

Within Switzerland, anthropogenic eutrophication of several large perialpine lakes during the last century resulted in the loss of ~29% of all described Alpine whitefish species through a combination of speciation reversal and demographic decline (Vonlanthen et al., 2012). Anthropogenic eutrophication had dual consequences on the Swiss lakes: First, hypoxic conditions (especially at the water-sediment interface) resulted in the loss of deep-water spawning grounds, reducing prezygotic reproductive isolation (Deufel, Löffler, & Wagner, 1986). And second, a shift in the available food resources reduced divergent selection between feeding niches, weakening postzygotic reproductive isolation and decreasing the strength of selection against hybrids (Nümann, 1972; Vonlanthen et al., 2012). The combination of both resulted in speciation reversal through introgressive hybridization. Together with the habitat loss due to the eutrophic conditions, speciation reversal led to the extinction of numerous endemic Swiss Alpine whitefish species (Vonlanthen et al., 2012).

The aim of this thesis was to investigate the genomic consequences of eutrophication-induced speciation reversal on the Alpine whitefish radiation by comparing genomes of individuals sampled before, during and after the period of eutrophication. By making use of the historical fish scale sample collection of Vonlanthen et al. (2012), we generated population level whole-genome resequencing data of all species of the Lake Constance whitefish radiation, including data of the now extinct deep-water species *C. gutturosus*. We showed that hybridization and introgression that contributed to the extinction of the profundal *C. gutturosus* during the period of eutrophication resulted in the retention of considerable proportions of its genomic variation within the surviving species of the radiation (Chapter I). However, by tracking the change in genomic diversity in each of the three surviving Lake Constance whitefish species through time, we demonstrated that all three surviving species

lost substantial amounts of genomic diversity throughout the last century, presumably due to demographic decline in response to eutrophic conditions (Chapter III). We sequenced six spawning depth populations (4m, 12m, 20m, 40m, 60m, 90m) of the deepest spawning extant Lake Constance whitefish species *C. macrophthalmus*, sampled along a spawning depth gradient. The results indicated that introgressed genomic variation from the extinct deep-water species *C. gutturosus* might facilitate ongoing adaptation to deep water in *C. macrophthalmus* (Chapter II), even though the total genomic variation of the species got reduced during eutrophication (see Chapter III). Furthermore, we generated a whole-genome resequencing data set containing all taxonomically described whitefish species of lakes Thun, Lucerne, Walen and Constance to compare differentiation between the profundal and all other sympatric species across these four lakes. The data indicated a low level of genomic parallelism and the potential for genotypic redundancy regarding profundal adaptation within the Alpine whitefish radiation (Chapter IV).

Genomic erosion due to environmental disturbance

A sufficient level of genomic diversity is a prerequisite to sustain a fit population and to preserve its evolutionary potential (Hoffmann, Sgro, & Kristensen, 2017). High genomic diversity enables a population to rapidly adapt to changing environmental conditions, and hence, also affects the risk of extinction during environmental change (Grant & Grant, 2019; Jensen & Leigh, 2022). Anthropogenic eutrophication resulted in extensive interspecific hybridization between Lake Constance whitefish species. Hybridization alone is expected to increase the genomic variation in a population or species (Grant & Grant, 2019). However, during the period of anthropogenic eutrophication, several evolutionary and demographic processes, such as habitat loss and thus presumably demographic decline, came together. Although hybridization might have increased the genomic diversity of each Lake Constance

whitefish species to a certain extent, the net outcome was a loss of diversity in each of the three surviving species (and in the Lake Constance radiation as a whole; see Chapter III). The consequences of this loss of genomic diversity for each species, such as the effects on the fitness and evolvability of each species are difficult to predict. However, the re-establishment of the same level of diversity will take much longer than it took to lose it.

During the past few years, fishing yields of the three surviving Lake Constance whitefish species were on a historically low level. Even though this reduction in whitefish catches of professional fishermen is a complex and multilayered problem, the decreased genomic diversity could play a role, because genetic diversity is correlated with fitness (Reed, 2005; Reed & Frankham, 2003). The high fishing pressures and the current practice of spawning fisheries and stocking, intended to increase the population size and fishing yields, might be detrimental in this regard, because it probably reduces the number of individuals that contribute to the next generation. Hence, management practices should be critically re-considered to maximize the chances for the populations to fully recover from the disturbances during the last century and to re-establish the original levels of genetic diversity in each species.

The potential benefit of hybridization during environmental change

The concept of disruptive and divergent selection, central to ecological speciation, predicts that intermediate phenotypes have decreased fitness compared to more extreme phenotypes (Rundle & Nosil, 2005; Schluter, 2000). Hence, in the absence of ecological disturbance, parental species, which are often characterized by a more extreme phenotype, have a higher fitness than hybrids that often exhibit intermediate phenotypes. However, during speciation reversal, environmental disturbance might change the fitness landscape from multimodal towards unimodal or flat (Vonlanthen et al., 2012), and hence, the fitness

disadvantage of hybrids could disappear (Grabenstein & Taylor, 2018; Seehausen et al., 2008). In some situations of environmental change, hybrids might even have increased fitness compared to the original parental species (Seehausen, 2006; Taylor et al., 2006).

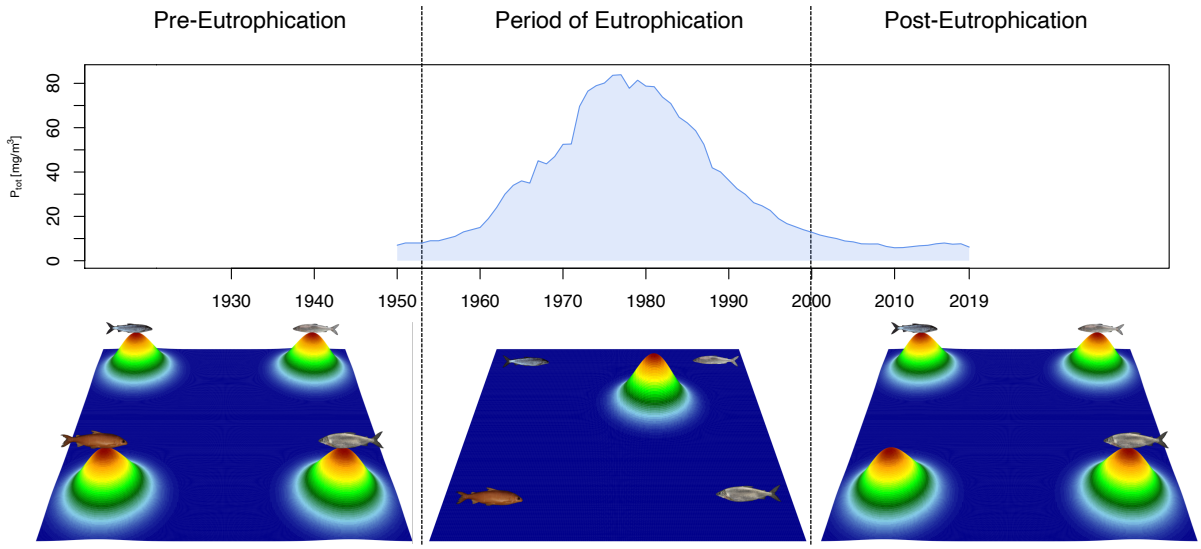


Figure 1: **Hypothetical change of the adaptive landscape over the eutrophication period**. On top, total phosphate concentration of Lake Constance over time is shown. Below, the hypothetical change of the adaptive landscape is illustrated. Before the onset of speciation reversal ("pre-eutrophication"), each of the four Lake Constance whitefish species was occupying a fitness peak in a multimodal fitness landscape. Eutrophic conditions altered the relationship between phenotype and fitness, resulting in a unimodal or flat adaptive landscape during the period of anthropogenic eutrophication. After the lake returned to oligotrophic conditions ("post-eutrophication"), the original or a similar adaptive landscape was probably restored. The niche originally occupied by *C. gutturosus* that went extinct during eutrophication and speciation reversal is vacant and might represent ecological opportunity.

Eutrophication-induced hybridization resulted in the maintenance of a considerable fraction of potentially adaptive genomic variation from the now extinct *C. gutturosus* within all three surviving species of the Lake Constance whitefish radiation. Furthermore, we found no indication for selection against introgression, while regions showing signals of positive selection in the extinct species were significantly enriched within introgressed genomic windows (see Chapter I). The absence of selection against introgressed variation is in line with the theoretical prediction of reduced divergent selection during speciation reversal. Due to a changed adaptive landscape (Fig. 1), hybrids might not experience a fitness disadvantage relative to the parental species during environmental change and speciation reversal.

Furthermore, considering the loss of genomic variation in all Lake Constance whitefish throughout the period of eutrophication, any increase in genomic diversity through hybridization could become beneficial. Thus, hybridization could at least partially counteract negative effects resulting from the reduction of genomic variation. Thereby, the process of hybridization itself could become adaptive during environmental change and speciation reversal. Recent work demonstrated that hybridization can reduce the vulnerability to environmental change (Brauer et al., 2023). Hence, the evolution of complete reproductive isolation between species might increase the extinction risk during (natural and anthropogenic) environmental change, because it precludes the possibility of adaptation to the changed environmental conditions through adaptive introgressive hybridization.

The two-sided nature of hybridization

Hybridization has the potential to result in the loss of biodiversity within a short timeframe (Grabenstein & Taylor, 2018), but it can also facilitate the evolution of entire adaptive radiations (Seehausen, 2004). The findings presented in this thesis (see Chapter I and Chapter II) share certain common aspects with the case of the Lake Victoria cichlids. Hybridization fueled the evolution of the Haplochromine adaptive radiation (Meier et al., 2017) but also resulted in a dramatic loss of cichlid species diversity (Seehausen et al., 1997). On the one hand, eutrophication-induced hybridization contributed to the rapid extinction of many endemic Alpine whitefish species (such as the deep water species *C. gutturosus* in Lake Constance), and to the loss of genomic differentiation (and hence possibly also adaptation) between surviving whitefish species. On the other, genomic variation derived from hybridization with *C. gutturosus* during eutrophication could potentially facilitate ongoing adaptation to deep water in *C. macrophthalmus*, the deepest spawning extant species. Taken

together, the results presented in this thesis highlight the two-sided nature of hybridization as evolutionary force that can both generate but also reduce biodiversity.

Ecological and genomic resilience

Today, most of the large Swiss lakes have returned to oligotrophic conditions, similar to the original conditions before eutrophication. Hence, the question arises of whether the lakes as ecosystems will fully recover from the disturbance through the eutrophic conditions during the last century. The concept of ecological resilience describes the ability of an ecosystem to withstand and recover from disturbance, without transitioning into an alternative stable state (Gunderson, 2000; Hirota, Holmgren, Van Nes, & Scheffer, 2011; Holling, 1973). Many resilience frameworks focus on relatively high levels of biological organization, such as communities or ecosystems (Capdevila, Stott, Beger, & Salguero-Gomez, 2020). However, Capdevila et al. (2020) described the concept of demographic resilience, a resilience framework with the focus on populations (or species) instead of higher levels of biological organization. Demographic resilience is defined as the inherent ability of a population to resist and recover from disturbance (Capdevila et al., 2020). Ecological disturbance and the associated demographic change can affect the genome, the genomic diversity and the evolutionary potential of a population (see e.g., Chapter I & Chapter III). This relationship opens up a genomic perspective on the concept of ecological resilience.

Genomic resilience can be defined as the ability to resist and recover from disturbance without losing the genomic integrity and variation (and thus evolutionary potential) that is needed to maintain the ecological function. In this context, the relevant level of biological organization is not necessarily a population or species (or its genomic variation), but rather all coexisting species that are able to hybridize and thus are connected by geneflow, respectively all their genomic variation and ecologically relevant alleles. Gene flow between coexisting

species (or populations) can reciprocally affect the demographic and genomic resilience of each single species through the exchange of ecologically relevant genomic variation. Thereby, hybridization might increase the potential for resistance or adaptation to changed environmental conditions of an entire adaptive radiation, a community or an ecosystem. However, hybridization can also result in the transition into an alternative stable state, which in the context of genomic resilience is the collapse into a hybrid swarm and the loss of linkage disequilibrium between alleles underlying ecologically relevant traits (and thus genomic integrity), resulting in the loss of ecological function.

Ecological resilience is not necessarily connected to the survival of a specific species in an ecosystem. If a given species with a certain ecological function goes extinct, it might be replaced by another species with the same or a similar ecological function. The ecosystem could still be considered ecologically resilient. Similarly, the ecological function or value of a given allele is not automatically confined to a single species, because it can be transported into any closely related coexisting species via hybridization and introgression. Thus, hybridization can rescue genomic variation of a species in the process of extinction and transport it into surviving species, where it can be reused and increase genomic resilience. The findings presented in this thesis (see Chapter II) suggest that alleles derived from hybridization with the extinct *C. gutturosus* are involved in the ongoing adaptation to deep water in the surviving *C. macrophthalmus* and represent a possible case of hybridization facilitating genomic resilience.

A high degree of genotypic redundancy is characterized by multiple genotypic solutions to the same selective pressure (Laruson, Yeaman, & Lotterhos, 2020). In the context of the Alpine whitefish radiation, genotypic redundancy could potentially play a role in (parallel) evolution of profundal species within the lake specific species flocks (see Chapter IV), as well as in the repeated evolution of other ecomorph contrasts across lakes (see De-

Kayne et al. (2022)). If environmental change results in the loss of alleles or genotypes underlying an ecologically relevant trait, genotypic redundancy might provide alternative genotypic solutions for the ecologically relevant trait, and thus compensate the loss of genomic variation through environmental change. Thereby, a high level of genotypic redundancy could increase the potential for genomic resilience.

Natural history collections and their value for evolutionary biology and biodiversity conservation

Natural history collections can represent a valuable resource to study the temporal dynamics of evolutionary processes, to expand the taxonomic sampling or to define the original and natural state of an ecosystem before anthropogenic disturbance (Jensen & Leigh, 2022). With the advance in sequencing technology, it is feasible to sequence historical samples with a lower DNA quality with less effort (Raxworthy & Smith, 2021). This makes it possible to use many natural history collections to generate sequence data, and thereby to study questions in evolutionary biology and ecology that could not be answered otherwise. One example is the inclusion of extinct species in the analysis of introgression dynamics, which otherwise would be limited to detecting introgression from "ghost lineages" (Ottenburghs, 2020; Tricou, Tannier, & De Vienne, 2022). In line with findings based on other historical samples such as herbaria (Bieker et al., 2020), dried insect specimen (Lalonde & Marcus, 2020), mammal skin tissue (Avila-Arcos et al., 2013) or egg shells (Grealy, Langmore, Joseph, & Holleley, 2021), the results presented in this thesis generated by sequencing historical fish scale samples highlight the value of natural history collections to study evolutionary dynamics of past and present populations. As herbaria, preserved specimen and mounted tissue collections are available for a wide range of study systems and taxa (Raxworthy & Smith, 2021), the potential applications of natural history collections in evolutionary biology are broad.

Apart from the field of evolutionary biology, natural history collections can be used to define suitable baselines for the state of an ecosystem before anthropogenic disturbance (Jensen & Leigh, 2022). This definition of suitable baselines is particularly relevant in the context of the shifting baselines syndrome (Pauly, 1995), because increased levels of disturbance and ecosystem degradation can lead to a lowered subjective threshold for the adequate state of an ecosystem. Data generated by using historical samples from natural history collections, such as whole-genome resequencing data, can be instrumental to document the change of species, communities and ecosystems against the shifting baseline syndrome (Jensen & Leigh, 2022).

Natural history collections became particularly relevant in the field of biodiversity conservation with the recent advances in sequencing technology. Next-generation sequencing technologies can provide critical insights in the temporal dynamics of e.g. genomic erosion, change in population structure or hybridization (Jensen & Leigh, 2022). The data generated for this thesis from historical fish scales demonstrated that results based on millions of SNPs can be different from results based on the exact same samples but using only few (and other types of) genetic markers (see Chapter III). Thus, conclusions can be divergent (and even contrasting) depending on the used approach. Hence, the application of next-generation sequencing technologies in combination with the use of samples from natural history collections can substantially contribute to the implementation of targeted and efficient conservation measures.

The balance between technological advance and reproducibility for combining genomic data sets

Technological progress during the last decades has rapidly advanced the field of evolutionary genomics and now permits the generation of large-scale data sets with decreasing financial efforts. Whilst the base-pair yield is increasing with every shift in

sequencing technology, such technological shifts can also introduce bias due to small differences in the sequencing process. When data generated by sequencing platform with technological differences is combined, these differences can translate into erroneous genotype calls, lead to the inaccurate interpretation of the data and finally result in misleading conclusions. Hence, the development of reproducible bioinformatic pipelines, ranging from the actual sequencing process, to pre-processing of raw data, to genotypes and finally to the population genomic measures of interest, is fundamental for a forward-looking evolutionary genomics framework (see Chapter V). Only genomic data generated by reproducible workflows allows to build up and combine datasets over longer timespans and a broader taxonomic range, enabling the curation of large sequencing data sets that allow to answer big questions in evolutionary biology.

Outlook and concluding remarks

Technological progress is rapidly advancing the field of evolutionary genomics. Per-base sequencing costs continue to decrease and the application of technologies such as long- or linked-read sequencing becomes feasible for many projects and study systems. Long- or linked-read sequencing approaches could represent a promising tool to investigate the dynamics of hybridization and introgression within the Alpine whitefish radiation in more detail. Because such data greatly improves the accuracy of statistical phasing and also provides actual phase information from the DNA fragments, it could enhance the inference of signals of selection along the genome. Further, this approach would allow to characterize introgressed blocks and infer their length and thus maximize the resolution in the detection of introgression. Moreover, information about introgressed block length could potentially be used to directly infer if and what type of selection is acting on introgressed variation (see e.g., Duranton et al., 2018).

The results presented in this thesis highlight the use of genomic data for the purpose of both evolutionary biology and conservation, which often are interconnected. However, even though genomic data can be a powerful tool for the purpose of biodiversity conservation, its use is sometimes still neglected by authorities and practitioners. In the context of the Alpine whitefish radiation, the establishment of a conservation genomics framework, intended to provide a basis for the acceptance of genomic data in management and practice, would facilitate the implementation of target-oriented conservation measures. Hence, the application of genomic data could substantially contribute to the conservation of the Alpine freshwater ecosystems and their diversity.

References

Avila-Arcos, M. C., Ho, S. Y. W., Ishida, Y., Nikolaidis, N., Tsangaras, K., Honig, K., . . . Greenwood, A. D. (2013). One hundred twenty years of koala retrovirus evolution determined from museum skins. *Molecular Biology and Evolution, 30*(2), 299-304. doi:10.1093/molbev/mss223

Bieker, V. C., Barreiro, F. S., Rasmussen, J. A., Brunier, M., Wales, N., & Martin, M. D. (2020). Metagenomic analysis of historical herbarium specimens reveals a postmortem microbial community. *Molecular Ecology Resources, 20*(5), 1206-1219. doi:10.1111/1755-0998.13174

Brauer, C. J., Sandoval-Castillo, J., Gates, K., Hammer, M. P., Unmack, P. J., Bernatchez, L., & Beheregaray, L. B. (2023). Natural hybridization reduces vulnerability to climate change. *Nature Climate Change*. doi:10.1038/s41558-022-01585-1

Capdevila, P., Stott, I., Beger, M., & Salguero-Gomez, R. (2020). Towards a comparative framework of demographic resilience. *Trends in Ecology & Evolution, 35*(9), 776-786. doi:10.1016/j.tree.2020.05.001

De-Kayne, R., Selz, O. M., Marques, D. A., Frei, D., Seehausen, O., & Feulner, P. G. D. (2022). Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nature Communications, 13*(4479), 1-13. doi:10.1038/s41467-022-32181-8

Deufel, J., Löffler, H., & Wagner, B. (1986). Auswirkungen der Eutrophierung und anderer anthropogener Einflüsse auf die Laichplätze einiger Bodensee-Fischarten. *Österreichs Fischerei, 39*, 325-336.

Duranton, M., Allal, F., Fraisse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.A. (2018). The origin and remolding of genomic islands of differentiaion in the European sea bass. *Nature Communications*, *9*(2518), 1-11. doi: 10.1038/s41467-018-04963-6

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology Evolution and* Systematics*, 34*, 487-515. doi:10.1146/annurev.ecolsys.34.011802.132419

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking barriers: Causes, consequences, and experimental utility of human-mediated hybridization. *Trends in Ecology & Evolution, 33*(3), 198-212. doi:10.1016/j.tree.2017.12.008

Grant, P. R., & Grant, B. R. (2019). Hybridization increases population variation during adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America, 116*(46), 23216-23224. doi:10.1073/pnas.1913534116

Grealy, A., Langmore, N. E., Joseph, L., & Holleley, C. E. (2021). Genetic barcoding of museum eggshell improves data integrity of avian biological collections. *Scientific Reports, 11*(1). doi:10.1038/s41598-020-79852-4

Gunderson, L. H. (2000). Ecological resilience - In theory and application. *Annual Review of Ecology and Systematics, 31*, 425-439. doi:10.1146/annurev.ecolsys.31.1.425

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., . . . Townshend, J. R. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances, 1*(2), 1-9. doi:10.1126/sciadv.1500052

Hirota, M., Holmgren, M., Van Nes, E. H., & Scheffer, M. (2011). Global resilience of tropical forest and savanna to critical transitions. *Science, 334*(6053), 232-235. doi:10.1126/science.1210657

Hoffmann, A. A., Sgro, C. M., & Kristensen, T. N. (2017). Revisiting adaptive potential, population size, and conservation. *Trends in Ecology & Evolution, 32*(7), 506-517. doi:10.1016/j.tree.2017.03.012

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics, 4*, 1-23. doi: 10.1146/annurev.es.04.110173.000245

Jensen, E. L., & Leigh, D. M. (2022). Using temporal genomics to understand contemporary climate change responses in wildlife. *Ecology and Evolution, 12*(9), 1-8. doi:https://doi.org/10.1002/ece3.9340

Lalonde, M. M. L., & Marcus, J. M. (2020). How old can we go? Evaluating the age limit for effective DNA recovery from historical insect specimens. *Systematic Entomology, 45*(3), 505-515. doi:10.1111/syen.12411

Laruson, A. J., Yeaman, S., & Lotterhos, K. E. (2020). The importance of genetic redundancy in evolution. *Trends in* Ecology *& Evolution, 35*(9), 809-822. doi:10.1016/j.tree.2020.04.009

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications, 8*(14363), 1-11. doi:10.1038/ncomms14363

Nümann, W. (1972). The Bodensee: Effects of exploitation and eutrophication on the Salmonid community. *Journal of the Fisheries Research Board of Canada, 29*(6), 833-884. doi:10.1139/f72-127

Ottenburghs, J. (2020). Ghost introgression: Spooky gene flow in the distant past. *Bioessays, 42*(6), 1-5. doi:10.1002/bies.202000012

Pauly, D. (1995). Anecdotes and the shifting base-line syndrome of fisheries. *Trends in Ecology & Evolution, 10*(10), 430-430. doi:10.1016/s0169-5347(00)89171-5

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution, 36*(11), 1049-1060. doi:10.1016/j.tree.2021.07.009

Reed, D. H. (2005). Relationship between population size and fitness. *Conservation Biology, 19*(2), 563-568. doi:10.1111/j.1523-1739.2005.00444.x

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation Biology, 17*(1), 230-237. doi:10.1046/j.1523-1739.2003.01236.x

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters, 8*(3), 336-352. doi:10.1111/j.1461-0248.2004.00715.x

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

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution, 19*(4), 198-207. doi:10.1016/j.tree.2004.01.003

Seehausen, O. (2006). Conservation: Losing biodiversity by reverse speciation. *Current Biology, 16*(9), 334-337. doi:10.1016/j.cub.2006.03.080

Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology, 17*(1), 30-44. doi:10.1111/j.1365-294X.2007.03529.x

Seehausen, O., van Alphen, J. J. M., & Witte, F. (1997). Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science, 277*(5333), 1808-1811. doi:10.1126/science.277.5333.1808

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology, 15*(2), 343-355. doi:10.1111/j.1365-294X.2005.02794.x

Tricou, T., Tannier, E., & De Vienne, D. M. (2022). Ghost lineages highly influence the interpretation of introgression tests. *Systematic Biology, 71*(5), 1147-1158. doi:10.1093/sysbio/syac011

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Muller, R., Lundsgaard-Hansen, B., . . . Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature, 482*(7385), 357-362. doi:10.1038/nature10824

Appendix

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 Fig. 3) 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 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 (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 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[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. 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 ($t=57.18$; $p<0.01$; $df=29.34$). 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. 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[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. 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 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. 9). 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 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

that had evolved in the extinct species prior to 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[48]. Nonetheless, profundal regions remain devoid of whitefish[49]. 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 of speciation[50]. However, re-emergence appears less likely the more species that 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 years or ~6 whitefish generations[51]) 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[5,52], 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 novel adaptive combinations of genotypes can accelerate adaptation and speciation[53]. Therefore, speciation reversal could potentially facilitate adaptation and diversification in response to changing or even entirely novel 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 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. 7). 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 Fig. 10 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 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[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 Fig. 3). 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 Fig. 10). 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-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 data set of genotype likelihoods obtained from angsd[61] (0.925) to only *C. gutturosus* individuals, and then inferred genotypes and phased these using beagle[64] 4.1. We then calculated the unstandardized nSL statistic with the software selscan[75] (version 1.3.0). Because the sample size consisted of 11 individuals, we included low frequency variants. We then used norm[75] (version 1.3.0) to normalize the unstandardized nSL calculations with default parameters in 50 kb windows along the genome. We considered windows with more than 51.1% of variable sites (top 1 percentile) with a normalized nSL score above 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[47] in angsd[61] (0.925) based on 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[61] (0.925) and then calculated the maximum likelihood estimate of the folded site allele frequency spectrum using realSFS of angsd[61] (0.925). We used the global site allele frequency spectrum to calculate theta per site in realSFS of angsd[61] (0.925), and then calculated Tajima's D in 50 kb windows using thetaStat of angsd[61] (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'[73] ($p<0.01$; W= 8352543). 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[56]; ENA accession: GCA_902810595.1) that overlap in their position with the identified windows under selection were identified using bedtools[76] 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[77] v.2.38.1 separately for each of the three ontology classes cellular component (CC), biological processes (BP), and molecular function (MF). 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[73] 'wilcox.test' and showed that there was no significant difference between the two sets of windows.

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) is deposited on the eawag research data institutional collections (https://doi.org/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.

Figures

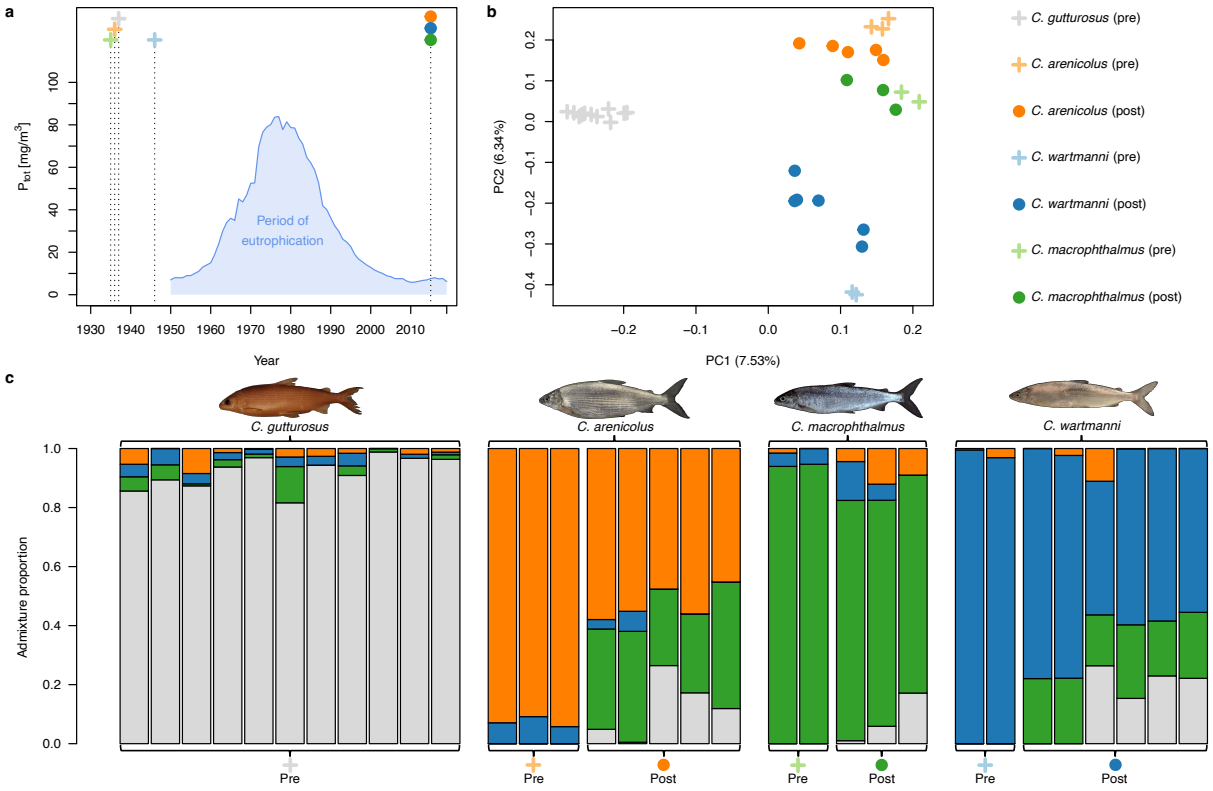


Fig. 1: Partial loss of genetic differentiation between Lake Constance whitefish species during eutrophication induced speciation reversal. a) Total phosphorus concentrations in Lake Constance over time as a proxy for severity of eutrophication. Time points when whitefish were sampled are indicated by 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.

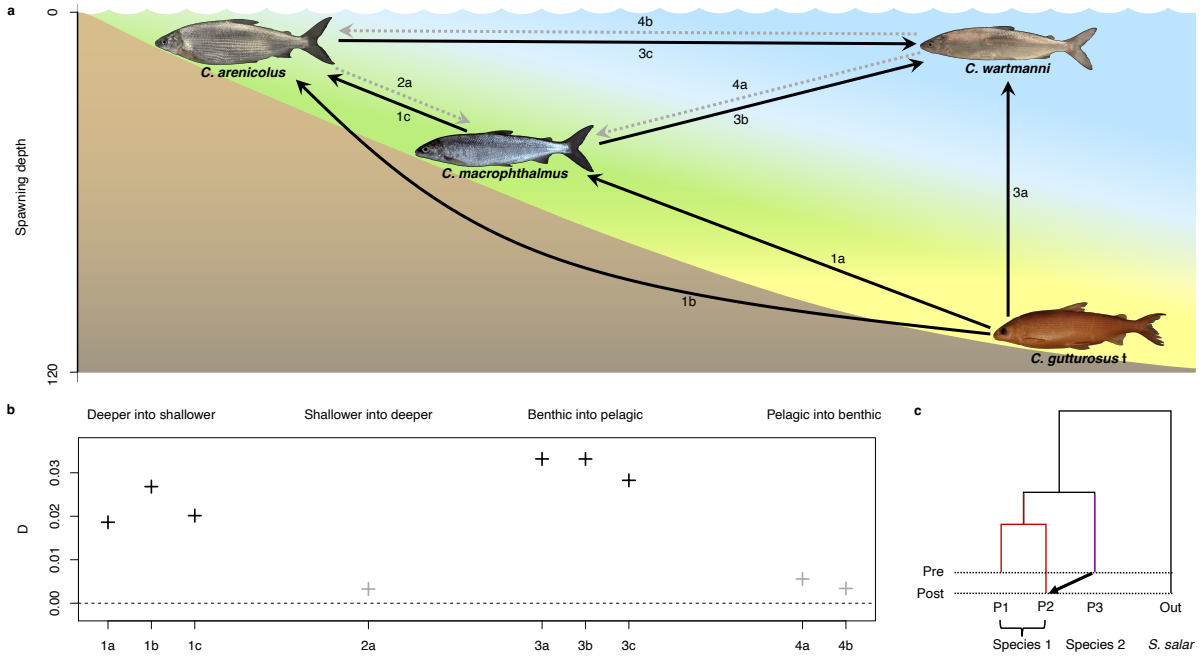


Fig. 2: Directionality of introgression during speciation reversal. a) Schematic representation of spawning habitat (water depth and benthic or pelagic habitat) of four Lake Constance whitefish species and the directionality of introgression. Significant tests for introgression during speciation reversal 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 Fig. 4). c**)** The topology shows the grouping of species for the underlying D-statistic test.

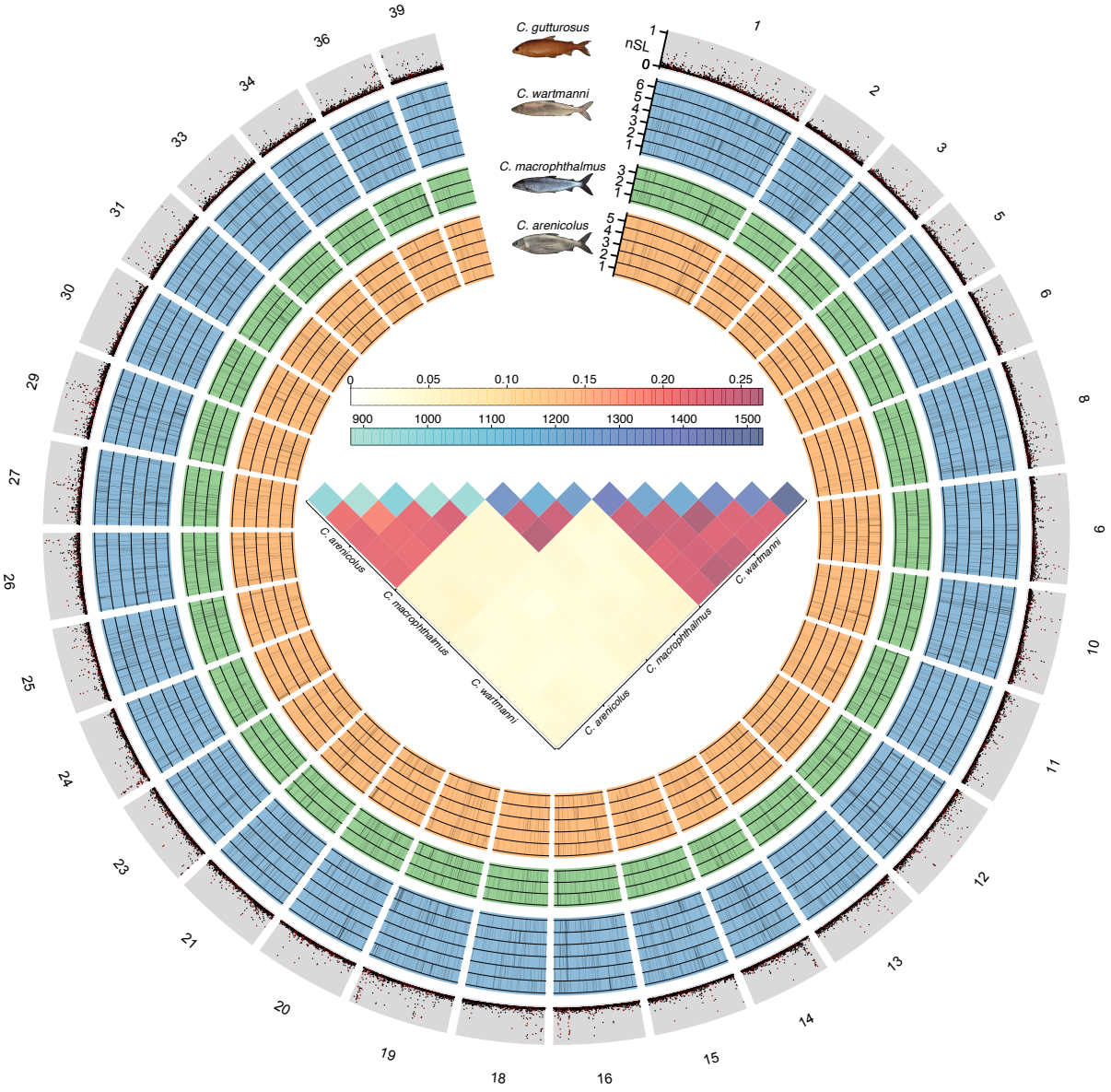


Fig. 3: Genomic distribution and characterization of introgression derived from extinct *C. gutturosus*. Each of the three inner tracks corresponds to a species (blue *C. wartmanni*, green *C. macrophthalmus* and orange *C. arenicolus*) and each track is subdivided into individual genomes. Each black 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).

36 Soraggi, S., Wiuf, C. & Albrechtsen, A. Powerful inference with the d-statistic on low-coverage whole-genome data. *G3-Genes Genomes Genet.* **8**, 551-566, doi:10.1534/g3.117.300192 (2018).

37 Wahl, B. & Loffler, H. Influences on the natural reproduction of whitefish (*Coregonus lavaretus*) in Lake Constance. *Can. J. Fish. Aquat. Sci.* **66**, 547-556, doi:10.1139/f09-019 (2009).

38 Nümann, W. The Bodensee: Effects of exploitation and eutrophication on the Salmonid community. *J. Fish. Res. Board of Can.* **29**, 833-884, doi:10.1139/f72-127 (1972).

39 Deufel, J., Löffler, H. & Wagner, B. Auswirkungen der Eutrophierung und anderer anthropogener Einflüsse auf die Laichplätze einiger Bodensee-Fischarten. *Österreichs Fischerei* **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, doi:http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-39858 (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, doi:10.3354/meps291249 (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, doi:10.3354/meps302233 (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, doi:10.1899/12-064.1 (2013).

44 Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429-438, doi:10.1534/genetics.116.194720 (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, doi:10.1111/2041-210x.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).

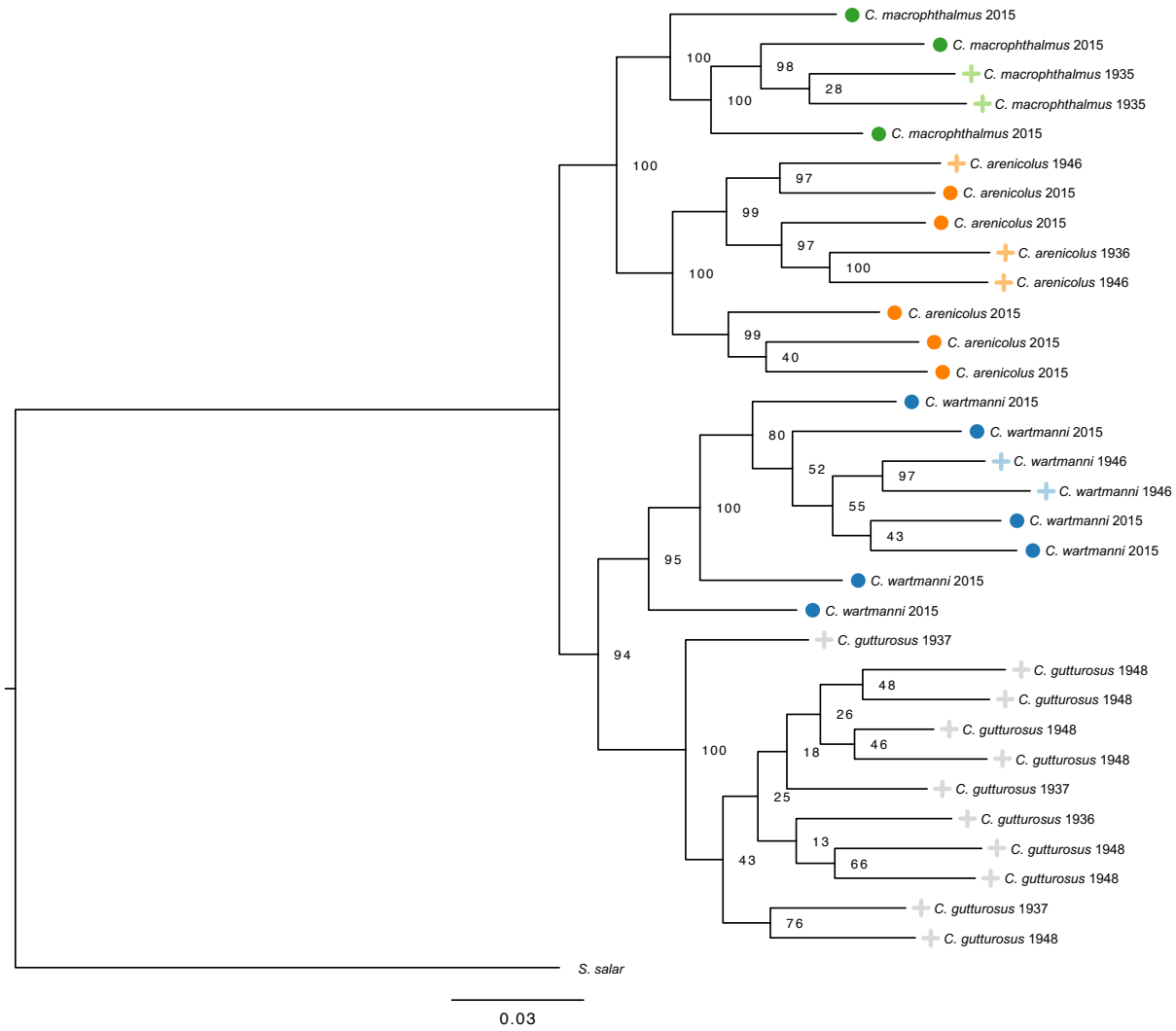
75 Szpiech, Z. A. & Hernandez, R. D. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824-2827, doi:10.1093/molbev/msu211 (2014).

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).

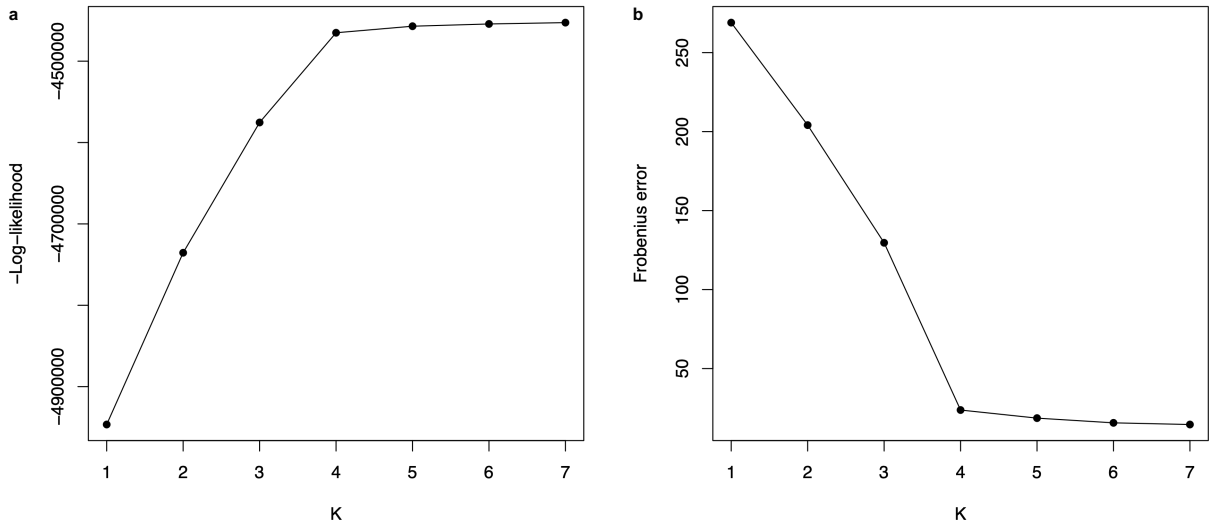
77 Alexa, A. & **Rahnenfuhrer**, J. *TopGO: Enrichment Analysis for Gene Ontology*. R package version 2.42.0. doi:10.18129/B9.bioc.topGO (2020).

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).

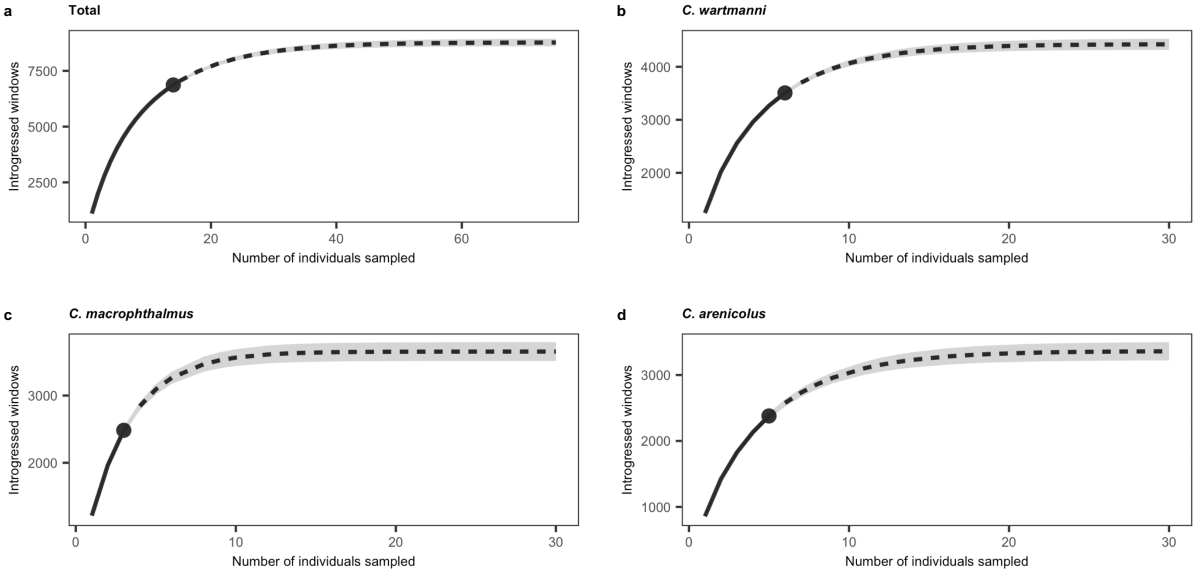
Extended Data Figures



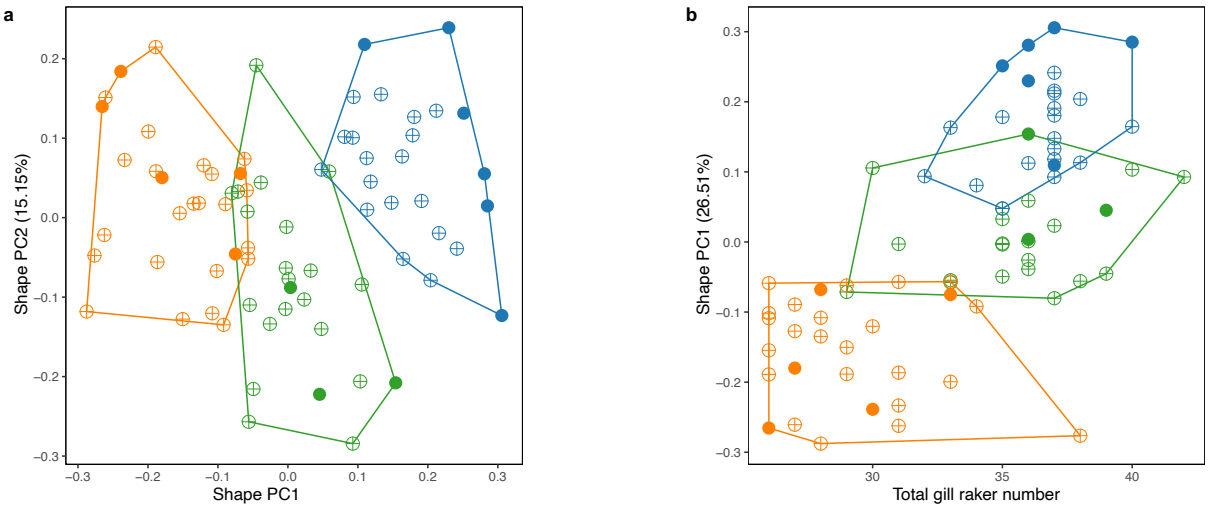
Extended Data Fig. 1: Maximum-likelihood phylogeny of all historical and contemporary samples. Maximum-likelihood phylogeny of all pre- (crosses) and post-eutrophication (points) individuals of the four Lake Constance whitefish species based on 58'831 SNPs. Colours correspond to species (see Fig. 1). 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).



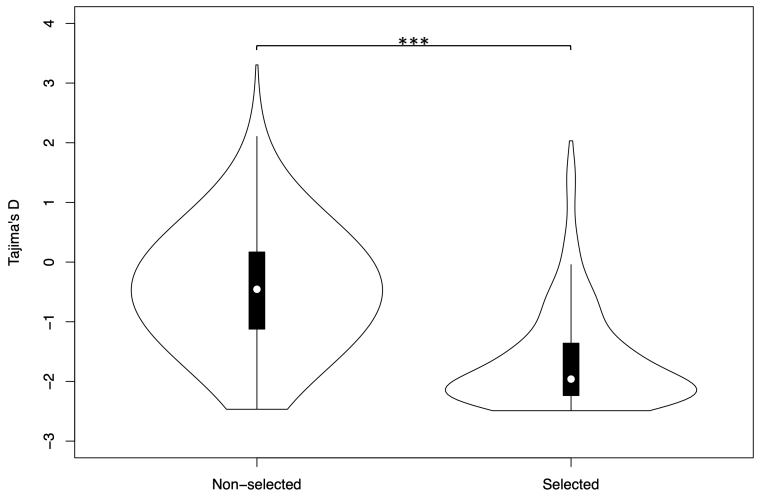
Extended Data Fig. 2: Log-likelihood and frobenius error for different K's of the admixture analysis. Log-likelihood values (**a**) and frobenius error (**b**) for different K's of the PCAngsd admixture analysis shown in Fig. 1c. K=4 turned out to represent the data best, also corresponding to the number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



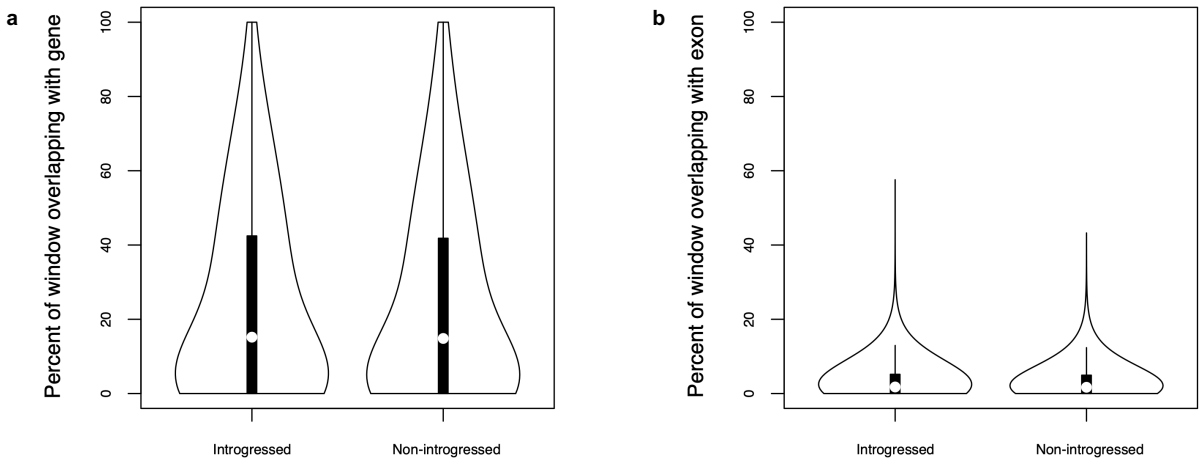
Extended Data Fig. 3: Rarefaction analysis of the *C. gutturosus* genome maintained in extant whitefish species. Rarefaction curves for all species combined showing the estimated number of introgressed windows in contemporary populations of Lake Constance whitefish (**a**), and for each extant species 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.



Extended Data Fig. 4: Morphological differentiation of contemporary Lake Constance whitefish. **a)** Shape PCA of the first two principal components based on body characters (PELVFB, PELVFS, PELVF, PECFB, PECF1, PECF2, 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.[30]. Morphological characters were measured and analysed following Selz et al.[30]. **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. arenicolus*, green *C. macrophthalmus* and blue *C. wartmanni*).



Extended Data Fig. 5: 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. 6: 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 no 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 no represented in the figure. 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 Tables

Extended Data Table 1: 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[3] 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*.

	Microsatellite data[3]		**SNP data**	
	F_{ST}	**n**	F_{ST}	**n**
Pre-eutrophication	0.108 (0.165)	68 (133)	0.046 (0.052)	18
Post-eutrophication	0.046	121	0.022	14

Extended Data Table 2: 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 Table 4).

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

Extended Data Table 3: *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. 3), and genome wide means of admixture proportions estimated with f_d (see Methods section).

	PCAngsd	**TWISST rarefaction**	f_d
C. arenicolus	0.12	0.11	0.11
C. macrophthalmus	0.08	0.12	0.10
C. wartmanni	0.14	0.14	0.14

Extended Data Table 4: 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.

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

Supplementary Information

Supplementary Table 1: Functional enrichment of windows under selection in *C. gutturosus.* GO enrichment analysis for all windows that were under positive selection in *C. gutturosus* (n=11) before its extinction, respectively before the eutrophic phase of the lake started. The first column shows the unique GO identifier for each enriched term, the second column gives the respective terminological description, the third column gives the number of genes annotated with the term within the genome, the fourth column gives the number how often the term was represented within windows with a signature of positive selection, the fifth column gives how often the term was expected in those windows by chance, the next two columns give the p-value using Fisher's exact method based on gene counts (accounting for the GO topology by weighting sixth column, or elimination seventh column) to test for a statistical overrepresentation of the term, and the last column gives one of the three ontologies of interest (CC cellular component, BP biological process, MF molecular function) that have been explored. For each category (CC, BP and MF), the top five entries are shown. Full output table is available as electronic supplementary material[1].

GO.ID	Term	Annotated	Significant	Expected	weight_fisher_P	elim_fisher_P	class
GO:0031094	platelet dense tubular network	11	3	0.07	4.9E-05	4.9E-05	CC
GO:0016528	sarcoplasm	122	5	0.83	0.0015	0.0015	CC
GO:0005747	mitochondrial respiratory chain complex I	43	3	0.29	0.0031	0.0031	CC
GO:0005952	cAMP-dependent protein kinase complex	17	2	0.12	0.0058	0.0058	CC
GO:0098588	bounding membrane of organelle	2294	28	15.55	0.0061	0.0213	CC
GO:0090330	regulation of platelet aggregation	27	3	0.19	0.0009	0.0412	BP
GO:0042311	vasodilation	30	3	0.21	0.0012	0.0012	BP
GO:0048210	Golgi vesicle fusion to target membrane	10	2	0.07	0.0022	0.0022	BP
GO:0060631	regulation of meiosis I	10	2	0.07	0.0022	0.0022	BP
GO:0035845	photoreceptor cell outer segment organization	37	3	0.26	0.0023	0.0023	BP
GO:0015278	calcium-release channel activity	26	4	0.18	3.2E-05	3.2E-05	MF
GO:0031681	G-protein beta-subunit binding	35	3	0.25	0.0019	0.0019	MF
GO:0004692	cGMP-dependent protein kinase activity	10	2	0.07	0.0022	0.0022	MF
GO:0099602	neurotransmitter receptor regulator activity	11	2	0.08	0.0026	0.0026	MF
GO:0031683	G-protein beta/gamma-subunit complex binding	86	4	0.61	0.0033	0.0033	MF

Supplementary References

1 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).

Acknowledgements

The time in KB shaped my thinking, both as a scientist and as a person. The supportive and inspiring atmosphere in KB is unique, and I am thankful for the opportunity to be part of the KB family. Especially, I want to thank Philine Feulner for giving me the chance to work in a fantastic research group. I am thankful for her incredible scientific, but also mental support throughout the past years. Although the pandemic was a challenging time to work and collaborate, the discussion and Fish Genomics group meetings always helped to overcome frustration and stress. I cannot thank Ole Seehausen enough for the supervision and support during my time as Bachelor, Master and PhD student. I am very thankful for his efforts to provide an inspiring and stimulating working atmosphere for the whole department, which had a huge positive impact on my wellbeing. It was an incredible experience to benefit from his feedback, input and knowledge throughout this past seven years and it influenced both my personal and scientific perspectives on many aspects of diversity.

A special thanks to Nadja Pepe, for always being there, with all kinds of different problems or issues. All her administrative and personal support was incredibly important, and I cannot thank her enough for her efforts and help for Fishec, for the Fish Genomics group and for myself.

Then, I want to thank all Eawag technicians for providing an outstanding infrastructure to do science. Thanks to Salome Mwaiko for keeping the molecular lab running, organized and tidy. I had so much fun working together with you. I am very thankful to Pascal Reichlin for sharing valuable knowledge about net-fishing, for helping me with my fieldwork, and also for many funny and interesting coffee breaks. I want to thank Beat Kienholz for his support with issues regarding all Eawag vehicles and boats, and for the numerous entertaining cigarette breaks (even though I have now quit smoking). Also, thanks

to Andreas Taverna for all the help with animal permits and LTK courses, for the help during fish duties and for the football discussions and betting games.

During the last years as a PhD student at Eawag, I very much enjoyed all the social interactions, collaborations and the scientific exchange with many good friends and colleagues in KB. I want to thank Oliver Selz for helping me with all whitefish questions and fieldwork, and for being a great mentor. I am thankful to Carmela Dönz for many stimulating scientific discussion and incredible mental support during all difficult situations. Thanks to Kurt Schmid and Philip Dermond for E.T. and all the funny coffee breaks at the lake (with the ducks and the swans). I am thankful to Kunio Takatsu for being the best office mate I could have wished for. I also thank Rishi De-Kayne, and Cas Retel for many inspiring Fish Genomics group meeting and Christmas parties. I am thankful to Kotaro Kagawa, whose expertise and input shaped the way I think about evolution. I thank Moritz Muschick and David Marques for sharing their experience and a lot of helpful feedback on my work. And, I thank Linda Haltiner for continuing to motivate and support me during the last months of my PhD, and especially for organizing the Seewandel writing retreat.

Finally, I want to thank my parents for fostering my fascination for nature, for giving me the opportunity to do this PhD, and for all your support. You have both played a key role in my education and influenced my scientific work more than you are probably aware of.

Erklärung

gemäss Art. 18 PromR Phil.-nat. 2019

Name/Vorname:	Frei David
Matrikelnummer:	09-116-740
Studiengang:	Ecology and Evolution
	Bachelor ☐ Master ☐ Dissertation ☑
Titel der Arbeit:	Genomic consequences of extinction by hybridization during eutrophication-induced speciation reversal
LeiterIn der Arbeit:	Philine Feulner und Ole Seehausen

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quel-len benutzt habe. Alle Stellen, die wörtlich oder sinn-gemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andern-falls der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes über die Universität vom 5. September 1996 und Artikel 69 des Universitätssta-tuts vom 7. Juni 2011 zum Entzug des Doktortitels be-rechtigt ist.
Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Univer-sität Bern das Recht, die dazu erforderlichen Perso-nendaten zu bearbeiten und Nutzungshandlungen vor-zunehmen, insbesondere die Doktorarbeit zu vervielfäl-tigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.

Goldach, 22.2.2023

Ort/Datum

Unterschrift