

MISS PHILINE G. D. FEULNER (Orcid ID : 0000-0002-8078-1788)

Article type : Original Article

## Genomic insights into the vulnerability of sympatric whitefish species flocks

Running title

Genomic insights on sympatric whitefish

Philine G.D. Feulner<sup>1,2</sup> & Ole Seehausen<sup>1,2</sup>

<sup>1</sup>Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Seestrasse 79, 6047 Kastanienbaum, Switzerland

<sup>2</sup>Division of Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland

Correspondence

Philine G.D. Feulner, Department of Fish Ecology and Evolution, Centre of Ecology, Evolution and Biogeochemistry, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Seestrasse 79, 6047 Kastanienbaum, Switzerland, Fax: +41 (0)58 765 21 68, mail: [philine.feulner@eawag.ch](mailto:philine.feulner@eawag.ch)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.14977

This article is protected by copyright. All rights reserved.

## Abstract

The erosion of habitat heterogeneity can reduce species diversity directly but can also lead to the loss of distinctiveness of sympatric species through speciation reversal. We know little about changes in genomic differentiation during the early stages of these processes, which can be mediated by anthropogenic perturbation. Here, we analyse three sympatric whitefish species (*Coregonus* spp) sampled across two neighbouring and connected Swiss pre-alpine lakes, which have been differentially affected by anthropogenic eutrophication. Our data set comprises 16,173 loci genotyped across 138 whitefish using restriction-site associated DNA sequencing (RADseq). Our analysis suggests that in each of the two lakes the population of a different, but ecologically similar, whitefish species declined following a recent period of eutrophication. Genomic signatures consistent with hybridisation are more pronounced in the more severely impacted lake. Comparisons between sympatric pairs of whitefish species with contrasting ecology, where one is shallow benthic and the other one more profundal pelagic, reveal genomic differentiation that is largely correlated along the genome, while differentiation is uncorrelated between pairs of allopatric provenance with similar ecology. We identify four genomic loci that provide evidence of parallel divergent adaptation between the shallow benthic species and the two different more profundal species. Functional annotations available for two of those loci are consistent with divergent ecological adaptation. Our genomic analysis indicates the action of divergent natural selection between sympatric whitefish species in pre-alpine lakes and reveals the vulnerability of these species to anthropogenic alterations of the environment and associated adaptive landscape.

## Keywords

ecological speciation, speciation reversal, *Coregonus* spp, RADseq

## Introduction

Ecologically based divergent natural selection, if persistent, can facilitate or initiate the evolution of reproductive isolation, a process known as ecological speciation (Mayr 1947; Rundle & Nosil 2005; Schluter 2001). Reproductive isolation between species that evolved by ecological speciation typically heavily relies on extrinsic environmental factors (Hendry 2009; Nosil 2012; Schluter 1996) and remains reversible for a very long time. Typically, it takes some period of geographic isolation (absence of gene flow) for complete irreversible reproductive isolation through intrinsic postzygotic hybrid incompatibilities to evolve, and until then reproductive isolation is dependent on prezygotic mechanisms (such as mate choice, breeding site choice, or time of mating) and extrinsic (ecology-dependent) postzygotic mechanisms (Seehausen 2006). Hence, the persistence of young sympatric species heavily depends on the maintenance of divergent natural and/or sexual selection and the efficiency of prezygotic isolation mechanisms to maintain distinctiveness until more permanent barriers eventually evolve. Such dependence on the environment makes many species arising from ecological speciation vulnerable to environmental disturbances and has widespread consequences for biodiversity (Seehausen *et al.* 2008).

Speciation reversal occurs when progression along an evolutionary trajectory toward complete speciation is reversed (Seehausen 2006), as might happen when environmental change weakens or eliminates a divergent selection regime. Speciation reversal is a quantitative reversal of the extent of reproductive isolation between young species and not a qualitative return to the ancestral state. Especially in times of rapid ecological changes, i.e. climate change and other anthropogenic perturbations, ecological speciation might often be reversed (Grabenstein & Taylor 2018). Indeed most cases of documented speciation reversal followed anthropogenic disturbances and changes in the environment (Darwin finches (De

León *et al.* 2011; Grant & Grant 2016; Kleindorfer *et al.* 2014) and several fish systems, including African cichlids (Konijnendijk *et al.* 2011; Seehausen *et al.* 1997), Canadian stickleback (Behm *et al.* 2010; Gow *et al.* 2007; Taylor *et al.* 2006), river herrings (Hasselman *et al.* 2014), North American ciscoes (Smith 1964; Todd & Stedman 1989), and European whitefish (Bhat *et al.* 2014; Vonlanthen *et al.* 2012). However, genomic insights based on dense genome-wide data are currently still rare for case studies of speciation reversal.

The feasibility of generating dense genome-wide marker sets for almost any organism has opened up the opportunity to address long standing evolutionary questions on the speciation process and its reversibility (Seehausen *et al.* 2014). Firstly, the new wealth of data permits describing patterns of differentiation between populations and species at an unprecedented fine scale. For example, dense marker sets have revealed subtle and previously cryptic population differentiation on a very fine geographic scale (Szulkin *et al.* 2016) and have disclosed signatures of admixture involving an extinct population (Feulner *et al.* 2013). In addition to the gain of ever-increasing resolution and power to detect subtle differences, genetic markers widely distributed along the genome also allow the identification of molecular signatures of selection within the genome (Nielsen 2005). Various population genomic approaches (Hohenlohe *et al.* 2010; Oleksyk *et al.* 2010) have assisted in addressing long standing questions regarding how many and which genes are involved in adaptation and speciation, which types of genetic variation are involved, and whether the involved genetic variants are pre-existing or novel (Barrett & Schluter 2008; Seehausen *et al.* 2014; Stapley *et al.* 2010). Utilising these varied genomic approaches, insightful observations have advanced our knowledge on the genomic changes occurring during speciation. A multitude of studies across various systems have found regions of exceptional differentiation to be widely spread

Accepted Article

across the genome, suggesting a polygenic basis of speciation (Feulner *et al.* 2015; Renaut *et al.* 2013; Soria-Carrasco *et al.* 2014). During the progressive establishment of reproductive isolation, distinct functional groups or specific pathways have been found to be overrepresented in regions shaped by divergent selection (Riesch *et al.* 2017). Regulatory changes have been shown to have a predominant appearance in repeated adaptive evolution (Jones *et al.* 2012). Evidence that regions of increased divergence contain ancient polymorphisms conferring the strongest resistance to gene flow has also been collected (Duranton *et al.* 2018; Meier *et al.* 2018). Hybridization has been identified as an influential mechanism fuelling adaptive radiations (Meier *et al.* 2017). While these and other studies have progressed our understanding of specific aspects of the speciation process, the underlying genomic landscape and the evolutionary processes shaping this landscape are still highly debated (Burri 2017; Ravinet *et al.* 2017; Wolf & Ellegren 2017).

Approaches that evaluate genetic differentiation based on relative allele frequency differences between populations have particularly been under scrutiny. These approaches are based on the assumption that demographics have in principle similar effects on all loci, such that loci showing exceptionally strong differentiation are indicative of being shaped by divergent selection and/or being shielded from homogenization by gene flow (Michel *et al.* 2010; Turner *et al.* 2005; Wu 2001). However, it has been posited that certain demographic histories, like bottlenecks and populations expansions are increasing the variance of genetic differentiation measured along the genome and might create signatures that can be mistaken as evidence of selection (Klopfstein *et al.* 2006). In this regard, repeated occurrences of the same regions of increased differentiation in several independently evolved population contrasts have been suggested as supporting evidence as such repeated occurrences are difficult to explain by neutral processes alone (Yeaman 2013). However, conclusively linking

genomic patterns of exceptional differentiation with reduced gene flow bears further challenges when aiming to identify loci important during speciation (Burri 2017; Cruickshank & Hahn 2014; Ravinet *et al.* 2017; Wolf & Ellegren 2017). Both background selection and directional selection, not connected to the speciation process, acting most strongly in regions of the genome where recombination is reduced, will result in very similar patterns of highly heterogeneous genomic differentiation even without any gene flow (Cruickshank & Hahn 2014; Nachman & Payseur 2012). These important insights build on previous work on the effect of background selection on genetic diversity (Charlesworth *et al.* 1997; Charlesworth *et al.* 1993) affecting measurements of relative differentiation. Hence, it is important to clearly think about the context of divergence and gene flow between study taxa and acknowledge that the context influences our ability to infer process from pattern (Wagner & Mandeville 2017).

Both young sympatric sister species and species affected by speciation reversal are useful study taxa for understanding genomic patterns of gene flow and reproductive isolation. In these systems, genomic signatures have not been obscured by post speciation events, such as background selection or other types of linked selection unrelated to the speciation process.

The *Coregonus lavaretus* species complex is a young radiation comprising of some 30 different whitefish species in the deep lakes of Switzerland alone (Hudson *et al.* 2010; Steinmann 1950). Up to 6 species coexist in some of the large, deep, and oligotrophic lakes in this region (Doenz *et al.* 2018; Hudson *et al.* 2017). Similar ecomorphs have evolved repeatedly and independently from each other in different lakes, following colonization of the pre-alpine region after the ice shields of the last glacial maximum retracted (Hudson *et al.* 2010). Sympatric species within the Alpine whitefish radiation differ in growth rate, body size, body shape, diet and feeding-related morphology, and habitat utilisation (Doenz *et al.*

2018; Hudson *et al.* 2017; Hudson *et al.* 2010; Ingram *et al.* 2012; Vonlanthen *et al.* 2012; Vonlanthen *et al.* 2009). Reproductive isolation between sympatric species in this radiation likely relies on extrinsic barriers like prezygotic isolation due to differences in spawning depth and time and postzygotic isolation, such as asynchronous hatch times in hybrids (Woods *et al.* 2009) and maybe also due to behavioural assortative mate choice.

In the second half of the 20th century the ecosystems of the pre-alpine lakes were altered dramatically including increases in phosphate input resulting in changes to algae biomass dynamics and composition (Gaedke & Schweizer 1993; Sommer *et al.* 1993), changes in zooplankton community structure (Burgi *et al.* 1999; Hairston Jr *et al.* 1999), and a dramatic reduction of oxygen levels in the deeper areas of lakes (Gächter & Müller 2003). The lake ecosystems became more homogenous and parts of the habitat (like the deep zone) became inaccessible to whitefish resulting in partial breakdown of reproductive isolation (Vonlanthen *et al.* 2012) and possibly also a relaxation of divergent selection between species (Hudson *et al.* 2013). During this period of intensive anthropogenic eutrophication of lakes, whitefish diversity decreased in most pre-alpine lakes (Vonlanthen *et al.* 2012). Sympatric species that survived in intermittently eutrophic lakes tend to have partially lost their ecological niche differentiation, as suggested by comparing present to past gill raker counts indicating a decrease of functional diversity (Vonlanthen *et al.* 2012). Speciation reversal resulted in decreases of morphological and genetic differentiation between sympatric species, varying within and between lakes, from slight introgression to complete extinction of species (Hudson *et al.* 2013; Vonlanthen *et al.* 2012). Across eight lakes the degree of decreased differentiation was indeed significantly correlated with the extent of phosphate intake, suggesting that eutrophication played a critical role (Vonlanthen *et al.* 2012). This history makes Alpine whitefish a prime study system for speciation genomics in young sympatric

Accepted Article

adaptive radiations. However, so far Alpine whitefish have not been investigated with next generation sequencing-based population genomic approaches. All previous genetic work on Alpine whitefish was based on microsatellites, Amplified Fragment Length Polymorphism data, or sequencing of a few genes (Bittner *et al.* 2010; Doenz *et al.* 2018; Douglas *et al.* 1999; Hudson *et al.* 2013; Hudson *et al.* 2017; Hudson *et al.* 2010; Ingram *et al.* 2012; Ostbye *et al.* 2005; Vonlanthen *et al.* 2012; Vonlanthen *et al.* 2009).

Here we compare the genomic patterns of differentiation between sympatric whitefish occurring in two adjacent and connected lakes with distinct eutrophication histories. In both lakes, Lake Zurich and Lake Walen, the same three whitefish species are taxonomically described (Kottelat & Freyhof 2007). However, more recent work has suggested that parts of this diversity might have been lost, with only two species still present (Vonlanthen *et al.* 2012). In our analysis based on dense genome wide single nucleotide polymorphism (SNP) data obtained using a restriction-site associated DNA sequencing (RADseq) approach, we intend to demonstrate the persistence of all three distinct sympatric species. We compare the abundances of the species in two lakes, which differ in their recent eutrophication history, and examine evidence for introgression between sympatric species. We analyse genomic differentiation between sympatric species from different habitats and between allopatric species and conspecific populations from similar habitats in different lakes. We observe the distribution of putatively divergently selected loci across the genome and determine whether these are confined to a few genomic regions. We investigate loci consistent with divergent selection between sympatric species for any overlap between the species pairs of the two lakes and for coinciding allele frequency differentials between shallow and deeper water habitats. We examine if those loci align close to genes with functions potentially relevant for habitat adaptation.



## Methods

### Study system and sampling

We investigated the population structure and species differentiation amongst whitefish inhabiting two neighbouring and connected lakes, Lake Zurich and Lake Walen, both part of the Limmat system. The lakes originate from a much larger postglacial Lake Limmat, which became separated into two adjacent and smaller basins by the rubble carried by the river Linth in the Holocene. In the early 19<sup>th</sup> century (1807 to 1822) the Linth was redirected to flow into Lake Walen with the latter overflowing to Lake Zurich via the Linth Canal. This reconstruction has severely impacted the ecosystem in Lake Walen (Steinmann 1950) and the direct massive inflow of glacial water from the Linth into the lake has made a lasting major change in water clarity. Later in the mid to late 20<sup>th</sup> century (with a peak around 1970) the lake ecosystems were again severely impacted by human activities, specifically by a large increase of phosphate inputs due to agricultural and household effluents. While phosphate levels in Lake Walen increased only modestly (max total phosphate concentration in the 1970s 26 µg/L, nowadays around 5 µg/L, (Vonlanthen *et al.* 2014)), phosphate levels in Lake Zurich increased more dramatically (max in the 1970s 119 µg/L, nowadays around 20 µg/L, (Alexander *et al.* 2017a)). Historical records refer to dramatic changes in whitefish populations following those environmental changes, and taxonomic work differentiated between two, three, or four distinct species (Fatio 1890; Kottelat 1997; Steinmann 1950; Wagler 1937). Most recently, Kottelat and Freyhof (2007) list three species occurring in both lakes. *Coregonus duplex* is a large benthivorous species, which spawns in shallow habitats during winter. *C. heglinus* is a small species, which spawns in the deep (20 - 80 m) potentially with summer and winter spawners. *C. zuerichensis* has been described as intermediate between the other species in body size and gill raker number, and is a planktivorous winter spawner (12 – 100 m). Morphologically *C. heglinus* and *C.*

*zuerichensis* can be difficult to distinguish and overlap in the distribution of many meristic traits. Previous studies suggested that *C. zuerichensis* had gone extinct during the eutrophication period, but doubts remained due to its morphological similarity to *C. heglungus* (Vonlanthen *et al.* 2012). Genetic studies based on microsatellites had shown weaker global differentiation between species in Lake Zurich compared to Lake Walen, and this was taken to suggest increased gene flow between species in the more disturbed lake (Vonlanthen *et al.* 2012). In total, our study used 180 tissue samples of whitefish caught in this system between 2004 and 2016. The samples originate from multiple sampling events in both lakes and in the Linth Canal (Alexander *et al.* 2017a; Karvonen *et al.* 2013; Vonlanthen *et al.* 2012; Vonlanthen *et al.* 2014), and include a laboratory family (parents plus two offspring) from Lake Thun being used for genotyping quality control (details see below). Table S1 summarizes details on locations and sampling time and depth (if available) as well as size (total length in mm) of every fish. Samples were stored in pure ethanol and deep-frozen (-80 °C). DNA was extracted from muscle or fin tissue following standard phenol chloroform procedure or using the Qiagen DNA easy tissue kit (as indicated in Table S1).

#### RAD sequencing and genotyping

Four Restriction-site Associated DNA (RAD) libraries were constructed using established procedures following Marques *et al.* (2016), a protocol slightly modified from Baird *et al.* (2008). In brief, genomic DNA was digested with *Sbf*I followed by shearing and size selection of 300 to 500 basepairs (bp). Equimolar proportions of DNA from 44 to 48 individuals were pooled into a single library and each library was sequenced (single end 100 bp) on one lane of Illumina HiSeq 2500 at Lausanne Genomic Technologies Facility. Sampled populations and to our best effort fishing actions were randomized across libraries (the last sequencing library was prepared subsequently, which only allowed for randomising

populations but not fishing actions; for details see NCBI SRP156755). Together with each library, we sequenced about 10% reads of bacteriophage PhiX genomic DNA (Illumina Inc.) to increase complexity at the first 10 sequenced base pairs. Read quality was assessed with fastQC v0.11.5 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). After removal of PhiX, reads were assigned to individuals based on their barcodes and reads without barcode or *SbfI* cut-site were filtered out and the remaining reads were trimmed to 90 bp making use of the process\_radtags module of stacks 1.26 (Catchen *et al.* 2013). Reads were filtered by removing reads with quality score below 10 for any base and if more than 5% of the bases were below a quality of 30 using FASTX Toolkit 0.0.13 ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)). All remaining reads (a total of 501,727,030 reads across 180 individuals) were combined to generate a read catalog by *de novo* assembling reads into unique loci (stacks) using ustacks with a minimum coverage per stack of 20 reads required and then building a consensus (reference loci) with cstacks (Catchen *et al.* 2013). Reads of each individual were mapped to these *de novo* pseudo reference loci (a catalog of 125,154 consensus loci each 90 bp long) using bowtie2 v2.2.6 (Langmead & Salzberg 2012) and genotypes were called with freebayes v1.0.0 (Garrison & Marth 2012). In order to utilize freebayes for this type of RAD data a heading and trailing N had to be added onto each reference locus (each stack). Changes to freebayes default setting included the exclusion of alignments with a mapping quality below 5, alleles with base quality below 5, and alternative alleles not supported by at least 5 reads. We allowed the detection of interrupted repeats and disabled prior expectations regarding read placement, strand balance probability, and read position probability (-V), and evaluated only best ranked SNP alleles (-n 10; an exhaustive search given that we later filter for biallelic SNPs). Genotypes were subsequently filtered in 8 steps: (1) Genotypes were kept if biallelic, having less than 50% missing data, a quality of more than 2, a minor allele frequency of more than

5%, and a minimal depth of 3. (2) Individuals were excluded if they had more than 50% missing genotypes. (3) Utilising dDocent ([https://github.com/jpuritz/dDocent/blob/master/scripts/dDocent\\_filters](https://github.com/jpuritz/dDocent/blob/master/scripts/dDocent_filters)) genotypes were filtered on criteria related to site depth, quality versus depth, strand presentation, and allelic balance in heterozygous. (4) Multiple allelic variants and indels were removed. (5) Using dDocent 359 sites were filtered if their alleles were out of Hardy-Weinberg-equilibrium in any of 4 well-characterized population samples of equal size. (6) Making use of four pedigree individuals (a pair of parents and two of their off-spring), 4,299 sites in Mendelian violations were detected, utilizing GATK (PhaseByTransmission; (McKenna *et al.* 2010)), and removed. (7) 71 loci/stacks with more than 4 SNPs were filtered out. (8) Genotypes and individuals with more than 30% missing data were removed. This filtering resulted in a file containing 138 individuals, 16,173 loci/stacks and 20,334 polymorphic sites. Whilst many sequence processing steps followed guidelines suggested by dDocent, we also used 25,266 genotypes available for a family (from the Lake Thun *C. sp.* “Balchen”), allowing us to remove 4,299 sites violating Mendelian segregation in this family from our overall data set. This additional step (6) was implemented to avoid erroneous genotypes from paralogous loci, which we expect to be frequently found in salmonids because of the relatively recent whole-genome duplication, which occurred 80-100 Mya in the shared ancestor of all salmonids (Macqueen & Johnston 2014).

#### Population genetic analysis

Nucleotide diversity ( $\pi$ ) for each population was estimated across all loci using all sites with vcfTools (Danecek *et al.* 2011). All other analyses are based on only 1 SNP per locus (of 90 bp). Population structure and species differentiation across the two lakes was assessed via PCA using SPNRelate v1.0.1 in R (Zheng *et al.* 2012) and via Bayesian clustering using

STRUCTURE v2.3.4 (Falush *et al.* 2003; Pritchard *et al.* 2000). We ran STRUCTURE ten times for each K from one to six, using a burn-in of 10,000 and running the chain for 20,000 generations. Utilising Structure Harvester (Earl & vonHoldt 2012), for each K the average likelihood values across the ten run and its standard deviation were summarised. For each K we plotted the clustering results (admixture proportions) for the run with the highest likelihood in R v3.1.3 (R Core Team 2015). Amova,  $F_{ST}$ , and  $F_{CT}$  were calculated and tested for significance by a permutation approach in Arlequin v3.5.2 (Excoffier & Lischer 2010). Outlier loci under selection were detected by the examination of the joint distribution of  $F_{ST}$  and heterozygosity under a hierarchical island model as implemented in Arlequin v3.5.2 (Excoffier & Lischer 2010). We calculated pairwise linkage disequilibrium (LD) with plink v1.07 (Purcell *et al.* 2007) between all pairs of outlier loci, to investigate if any of the outlier loci are closely physically linked.

#### Annotation

For positioning RAD loci onto the Atlantic Salmon genome (Lien *et al.* 2016) and the published whitefish scaffolds (Laporte *et al.* 2015), we used stampy v1.0.22 (Lunter & Goodson 2011). In addition, we blasted all outlier RAD loci (90 bp) against the non-redundant database using the default setting of blastn v2.2.28+ (Camacho *et al.* 2009). We report the best hit with any gene annotation (cds or mRNA) if at least 70 bp aligned or sequences match to 100%.

## Results

### Genetic diversity of populations consisting of three species

In total we genotyped 138 individuals across four libraries. The addition of four family samples (two parents and two laboratory bred offspring) allowed a thorough assessment of genotyping quality and the removal of false genotype calls due to paralogous loci. We evaluated genetic diversity for each sampled population and compared it between the five populations comprising three species ( $n = 14 - 44$  individuals, not considering *C. heglinus* from Lake Zurich  $n = 3$ ; Table 1), by using all the sequence information available across the 16,173 loci (a total of 1,455,570 bp, containing 20,334 SNPs). Average nucleotide diversity ( $\pi$ ) for the populations ranged between 0.0040 (*C. heglinus* from Lake Walen) and 0.0044 (*C. duplex* from Lake Zurich), showing only minor variation in diversity between populations and across the three species (see diagonal Table 1). Considering genetic diversity as a proxy for effective population size ( $N_e = \pi/4\mu$ ), the populations appeared to be of similar  $N_e$ . When applying a mutation rate ( $\mu$ ) of  $6.6 \times 10^{-8}$  (Recknagel *et al.* 2013) all population have a  $N_e > 15,000$  (max  $N_e = 16,662$  *C. duplex* from Lake Zurich). The observed minor differences in genetic diversity between the populations permit pairwise comparisons of relative differentiation (i.e.  $F_{ST}$ ), as estimates are unlikely to be affected by a reduction of diversity in one of the populations under consideration.

### Population structure and species differentiation

The samples investigated here were visually split into three morphologically distinct groups, corresponding to the three described species. The likelihood values from the Bayesian clustering (STRUCTURE) support three or four clusters almost equally well, with little differences in average likelihood values and variance across the ten runs (Figure 1a). Larger  $K$ s show a decrease in the average likelihood and increase in variance across runs, hence do

not explain the data well. However, with four genetic clusters (K=4; Figure 1b and Figure S1) the method differentiates the same three distinct groups and in addition a spatial gradient pattern across the *C. duplex* samples. A PCA based on all genotypes also resulted in three distinct non-overlapping clusters and a spatial genetic gradient across the *C. duplex* samples matching their geographic distribution (Figure 2b). The identified three distinct genetic clusters matched well with our morphological identifications and with the recorded sampling location and net depth. Out of 138 individuals four individuals were misidentified based on their phenotype. Whitefish species are difficult to identify as most of their distinguishing morphological features are relatively subtle and are fully developed only in adult, fully grown individuals. Additionally, there is a large amount of phenotypic variation within most whitefish species, some of which might be attributed to phenotypic plasticity. The three whitefish species of lakes Walen and Zurich (*C. duplex*, *C. zuerichensis*, and *C. heglingus*) show distinct but overlapping adult size distributions (Figure 2a). The average size of mature fish of the three species is significantly different (one-way anova,  $p < 2.2e-16$ ,  $F = 114.4$ ,  $DF = 2$  and  $128$ ), however the shapes of the size distributions are affected by net mesh size used and likely span a wide range of age classes (Figure 2a). The larger species, *C. duplex*, is during spawning season caught predominantly in shallow nets. While the two smaller species, *C. zuerichensis* and *C. heglingus*, are during spawning season caught in deeper nets set (see Table S1 and (Alexander *et al.* 2017a)). This distinction of shallow and deep (more profundal) spawning whitefish species is matched with the genetic differentiation observed. Sympatric species spawning in different depths (as evidential by the depth they are caught in during spawning in winter) are most strongly differentiated (see Table 1). Populations of the same species from the different lakes but also different species spawning at similar depth in different lakes are less strongly differentiated (see Table 1). The Linth Canal samples of *C. duplex* cluster between the populations of this species from the two lakes and connect them

(Figure 2b). Whereas most of the smaller whitefish from Lake Zurich turn out to be *C. zuerichensis*, three individuals genetically match *C. heglinus* from Lake Walen, suggesting that both deeper spawning species co-occur in this lake. By visual inspection of the PCA plot and the Bayesian clustering results (Figure 2b and 2c), we identify following potentially admixed individuals or hybrids. In Lake Walen, one intermediate hybrid between *C. duplex* and *C. heglinus* was identified (admixture proportion ~50%) and another *C. heglinus* individual showed a small admixture proportion of *C. duplex*. In Lake Zurich, all *C. zuerichensis* individuals appear to be partially admixed with *C. heglinus*, and four individuals were intermediates between *C. zuerichensis* and *C. duplex*. When extracting the admixture proportions from the Bayesian clustering approach (K=3), *C. duplex* appears to be on average more admixed in Lake Zurich (23.9%) than in Lake Walen (4.3%), with the Linth population showing intermediated admixture proportions (9.3%). The average admixture proportion in Lake Walen is largely driven by one intermediate individual and not much affected by the number of clusters (K=2: 4.0%; K=3: 4.3%; K=4: 4.0%). In Lake Zurich the admixture proportion varies with the number of clusters but is always larger than in Lake Walen (K=2: 16.8%; K=3: 23.9%; K=4: 8.2%). In summary, the combination of an extensive sample collection and a dense (16,173 SNPs spread across the genome) marker set allowed us to resolve the population structure and the extent of species differentiation of the three whitefish species inhabiting lakes Walen and Zurich. This unprecedented resolution revealed three genetically truly discrete whitefish species, which group into significantly differentiated, non-overlapping clusters reflecting their distinctiveness.



## Detection of outlier loci indicative of divergent selection

We tested if any of the RAD loci showed exceptionally strong genetic differentiation between species, a pattern consistent with the action of divergent selection. We compared pairwise contrasts of allopatric lake populations of the same species (*C. duplex* from Lake Zurich and Walen) and of different species with similar habitat niche (*C. zuerichensis* from Lake Zurich and *C. heglingus* from Lake Walen) with pairwise contrasts of sympatric species with different habitat niche (comparing *C. duplex* versus *C. heglingus* in Lake Walen and *C. duplex* versus *C. zuerichensis* in Lake Zurich). We could not include the Lake Zurich population of *C. heglingus* because of the small sample size, which in turn was caused by the rarity of this species in Lake Zurich. Outliers of exceptionally strong genetic differentiation were identified in comparisons to neutral simulations under a hierarchical island model (approach as implemented in Arlequin using a p-value cut-off of 0.001). The outlier analysis was performed without excluding any potential hybrid individuals, which were maintained as the species they were originally assigned by their morphology and sampling origin. In the allopatric within-niche comparison 22 outlier loci were detected for *C. duplex* and 33 outliers were detected in the allopatric similar-niche comparison of *C. zuerichensis* versus *C. heglingus*. Genetic differentiation across all loci was weakly correlated between these two allopatric pairs and outliers did not overlap (see Figure 3a,  $r^2 = 0.029$ ,  $p = 0.001$ ). Loci with strong genetic differentiation in one comparison, i.e. outlier loci, show only weak genetic differentiation in the other pair (see Figure 3a). In the two contrasts of sympatric shallow versus deep species we detected 33 outliers in Lake Walen and 36 in Lake Zurich, and four of those were shared between both pairs (see Table S2, shared outlier are highlighted with blue shading). At all the four shared outlier loci, allele frequency differentials have the same sign between the shallow and the deep spawning species in both lakes (see Table S3). Between sympatric contrasts genetic differentiation was positively correlated across loci ( $r^2 = 0.399$ ,  $p$

<< 0.001) and outliers in one comparison were often strongly differentiated in the other comparison too, beyond the four shared loci identified as outliers in each (see Figure 3b).

#### Distribution and annotation of outlier loci

In the absence of a whitefish reference genome we aligned the RAD loci against the Atlantic Salmon reference genome (Lien *et al.* 2016) to investigate relative positions of the detected outlier and their proximity to annotated genes. We focus on the two sympatric contrasts between species spawning in shallow versus deep water habitat. Based on the positioning on the salmon genome, outlier loci are well spread out along the genome and do not appear to be clustered in a few genomic regions (Figure 4). Three out of the four shared outlier loci could be aligned to three different salmon chromosomes (chromosomes 14, 18, 19). Only three outlier loci (detected in Lake Zurich only) matched to published scaffolds of the North American *C. clupeaformis* on which shape QTLs were detected. However, the match is not located near to the QTL associated SNPs (Laporte *et al.* 2015). On average, the RAD loci mapping to the *C. clupeaformis* scaffolds ( $n = 685$ ) show no elevation of genetic differentiation (Figure 4b, lake Zurich:  $t = 1.3829$ ,  $df = 738.745$ ,  $p = 0.1671$ , Lake Walen:  $t = 1.6884$ ,  $df = 741.008$ ,  $p = 0.09176$ ). For 12 out of the 65 (33 in Lake Walen, 36 in Lake Zurich, 4 shared between lakes) outlier loci from our sympatric pairs blast detected some sequence similarity with genes curated in the NCBI database (see Table S2). Two of the four outliers that were shared between our two sympatric species contrasts include the *Salmo salar* genes *retinitis pigmentosa 9* and *IgH locus B*. Outliers unique to Lake Walen align with *growth differentiation factor 11* and *defensin beta* annotated in *Oncorhynchus mykiss*, the Rainbow Trout, and a *charged multivesicular body protein 3* annotated in *S. salar*. Outliers unique to Lake Zurich include a *heat shock protein (HSP70)* found in *C. clupeaformis*, and an *olfactory receptor*, a *zink finger*, and a *GTPase* gene in *S. salar*. We also detected another

alignment match to the same *IgH locus B* of another RAD locus positioned onto a different salmon chromosome, but *IgH* is known to be duplicated in salmonids (Yasuike *et al.* 2010).

## Discussion

Species-specific demographic responses of whitefish between the two connected lakes

The genomic analysis presented here confirms the presence of three genetically distinct species (*C. duplex*, *C. heglingus*, and *C. zuerichensis*) in the two connected neighbouring

lakes, Lake Walen and Lake Zurich. All three species have been listed for both lakes

(Kottelat & Freyhof 2007) and also historical records suggested the presence of three distinct species in both lakes in the past (Wagler 1937). However, previous evidence suggested that

in both lakes, one of the species had been lost in the second half of the 20<sup>th</sup> century

(Vonlanthen *et al.* 2012) and the authors assumed that the same species (*C. zuerichensis*) got

lost in both lakes. More recent collections revealed phenotypic evidence for three different

taxa though in both lakes (Alexander *et al.* 2017a; Vonlanthen *et al.* 2014). Our genomic

analyses unambiguously reveal the sympatric occurrence of three genetically distinct species

in Lake Zurich, consistent with historical records (Kottelat & Freyhof 2007; Wagler 1937).

Hence, our data supports the contemporary presence of all three historically described

species. While we show that all three species co-occur in Lake Zurich, our samples from

Lake Walen only allowed confirming two species. It is likely that *C. zuerichensis* is not

absent but rather rare in Lake Walen. Unfortunately, we could not sequence the one fish that

was reported from this lake in the last major sampling event (Vonlanthen *et al.* 2014). In

Lake Zurich, we identified three individuals as *C. heglingus*. Both, the field observations

(Alexander *et al.* 2017a) and our sequencing results suggest that *C. heglingus* is rare in Lake

Zurich. However, our data does not resolve if the three *C. heglingus* individuals detected in

our sampling of Lake Zurich represent a recent introduction, a natural recolonization, or a

remaining rare population. Stocking and any other commercially motivated fish transfers are more likely focussed on the larger whitefish species (*C. duplex*), hence a smaller species spawning deeper in the lake, like *C. heglingsus*, is less likely to be affected by such human activities. In our one sampling during spawning season all the whitefish caught in the Linth Canal connecting the two lakes are *C. duplex*, hence we did find no evidence for any of the other species crossing the Linth Canal. However, all three *C. heglingsus* specimens were caught in different nets set during an untargeted sampling effort, which used standardised fishing protocol not targeted to specific sites or species (Alexander *et al.* 2017a; Vonlanthen *et al.* 2014). Interestingly, the three individuals are genetically not distinct from the *C. heglingsus* population in Lake Walen, which would suggest a recent recolonization. Based on our results here, and the phenotype-based results of quantitative fishing (Alexander *et al.* 2017a; Vonlanthen *et al.* 2014), we propose that the two smaller and more pelagic feeding whitefish species have responded differentially to the past changes in the ecosystems. That the deep spawning *C. heglingsus* became rare in Lake Zurich but remained abundant in Lake Walen can be explained because during eutrophication Lake Zurich experienced major and widespread anoxic conditions in deeper waters, whereas the less strongly nutrient-enriched Lake Walen remained oxygenated all the way to the greatest depths (Alexander *et al.* 2017a; Vonlanthen *et al.* 2014). The midwater spawning *C. zuerichensis* would have persisted in Lake Zurich and came to dominate this lake. The explanation for the decline of the latter species in Lake Walen is less apparent, but may have something to do with changes in the zooplankton community that could be associated with the increased turbidity of Lake Walen due to an increased influx of glacial melt water. This interpretation is consistent with our approximations of genetic effective population sizes, which show only minor differences and suggest similarly large effective population sizes for all three species.

In summary, the analysis of the genetic structure of populations and species presented here, suggests the contemporary presence of all three endemic whitefish species previously described for lakes Walen and Zurich. In addition, the results are consistent with distinct demographic responses to recent environmental change of the two deeper spawning species between the two lakes. Further, the ability of our dense genetic marker set to resolve sympatric species into clearly distinct and non-overlapping clusters highlights both, the distinctiveness of these species and the power of the approach in species designation among closely related sympatric species. It will be interesting to see if genomic studies at the same scale applied to more species rich pre-alpine whitefish lakes, will be able to provide a similar resolution. For example in the Lake Thun/Lake Brienz system, investigations using thousands of individuals genotyped at a dozen of microsatellite markers, while revealing six distinct species, had difficulty to demonstrate clear gaps in genotype space that would separate the species (Doenz *et al.* 2018).

#### Evidence for speciation reversal affecting sympatric whitefish species

Previous analysis of whitefish genetic differentiation based on microsatellite data in these lakes showed that global genetic differentiation in Lake Zurich is less pronounced than in Lake Walen (Vonlanthen *et al.* 2012). Phosphate data collected throughout the eutrophication period suggest that Lake Zurich was more heavily affected by eutrophication than Lake Walen (see methods section for details (Alexander *et al.* 2017a; Vonlanthen *et al.* 2014)). This is consistent with a more general pattern of a strongly negative correlation across lakes in the region between the extent of past eutrophication and the extent of current genetic differentiation between sympatric whitefish species (Vonlanthen *et al.* 2012). Hence, weaker genetic differentiation between species had been attributed to more introgression in Lake Zurich and was taken as evidence for partial speciation reversal. Our realization that the

abundant small species in the two lakes, assumed to be the same species in this previous study (Vonlanthen *et al.* 2012), are not the same species, complicates the interpretation.

However, our data is in agreement with the interpretation of weaker reproductive isolation, and more severe introgression in Lake Zurich. The shallow water spawning *C. duplex* that occurs in both lakes is clearly more strongly differentiated from the abundant sympatric deep spawning species in Lake Walen. In principle, this alone could also be due to more recent or less complete speciation between *C. duplex* and *C. zuerichensis* (the common species in Lake Zurich) than between *C. duplex* and *C. heglingus* (the common species in Lake Walen).

However, *C. duplex* of Lake Zurich shows increased proportions of *zuerichensis*-admixture relative to the same species sampled in Lake Walen (Figure 2c; on average 23.9% admixture in Lake Zurich and 4.3% admixture in Lake Walen *C. duplex*). Assuming that the three species and speciation events were historically shared between the two connected lakes, this can only be interpreted with post-speciation gene flow that affects the entire population of *C. duplex* in Lake Zurich but not its population in Lake Walen. In Lake Zurich, we additionally detected four more strongly admixed individuals (versus only one, or possibly two admixed individuals in Lake Walen). Nevertheless, the difference in the number of strongly admixed individuals might be coincidental, as our sampling size was smaller in Lake Walen (Table S1). Perhaps most importantly, *C. zuerichensis* of Lake Zurich itself is strongly admixed with *C. heglingus*, whereas the reverse is not true. This might suggest that introgression was strongly asymmetric from *C. heglingus* into *C. zuerichensis* and not vice versa. Alternatively, introgression between the two species occurred mainly in Lake Zurich leading to the loss of *C. heglingus*. The rare *C. heglingus* that we discovered in Lake Zurich, would then be a population that has recently recolonized Lake Zurich from Lake Walen potentially after eutrophication decreased. Consistent with this, the *C. heglingus* of Lake Zurich is genetically indistinguishable from the population in Lake Walen, whereas conspecific populations of *C.*

*duplex* are very significantly differentiated between these lakes. Conspecific populations of any species are also significantly differentiated between lakes Brienz and Thun that are connected by a short stretch of river much like lakes Walen and Zurich (Doenz *et al.* 2018). Our data confirms and extends on previous evidence (Vonlanthen *et al.* 2012) and suggests that the major ecological perturbations that the lakes experienced in recent history was associated with a loss of whitefish species diversity mediated by differential demographic responses between species and increased genetic admixture. This reinforces the importance of ecosystem stability for the maintenance of the diversity of endemic salmonid species in pre-alpine lakes, and probably for their evolutionary origins (Alexander *et al.* 2017b; Seehausen 2006; Seehausen *et al.* 2008).

Widespread genetic differentiation between sympatric whitefish species associated with contrasting habitats

Genome-wide data as presented here allows the identification of genomic regions important for divergent adaptation to shallow versus deep spawning habitats, as well as more benthic versus more pelagic feeding. These are key adaptations associated with ecological speciation in the whitefish radiation of the pre-alpine lakes and in salmonid lacustrine radiations more broadly. We can identify parts of the genome that have been influenced by divergent selection between the habitats, learn about how those regions are distributed across the genome, and which gene functions are encoded in these genomic regions. However, genome scan approaches aiming to detect signatures of selection in genomic data and pinpointing relevant regions of the genome have been criticised for detecting false positive signals.

Evolutionarily independent replicates of differentiation might help to address concerns regarding drift, especially in the context of certain demographic histories, such as bottlenecks and/or population expansions scenarios, which can impede the confident identification of outliers due to selection (Bank *et al.* 2014; Crisci *et al.* 2012; Klopstein *et al.* 2006; Marques *et al.* 2016). In addition, when aiming to reveal divergent selection signatures left by the speciation process, additional concerns relate to the possibility of being misled by the action of background selection that would generate similar genomic landscapes of differentiation in independent speciation events (Ravinet *et al.* 2017; Wolf & Ellegren 2017). Especially genome scans based on relative measurements of population differentiation (such as  $F_{ST}$ ) might also pick up signatures of directional or background selection unrelated to the speciation process (Cruickshank & Hahn 2014; Nachman & Payseur 2012). The jury is still out on how well different signatures of evolutionary processes can be distinguished from one another, however especially comparisons between young sympatric species, for which independent evidence suggests the presence of some gene flow, minimise the risk of detecting signatures that are not related to the speciation process (Burri 2017; McGee *et al.* 2015; Meier *et al.* 2018). However, if patterns of increased genetic differentiation between sister species, repeated across replicate speciation events, are shaped by a conserved recombination landscape, ruling out background selection is challenging. Across multiple comparisons background selection associated with low recombination regions should leave the same signature in pairwise comparisons of  $F_{ST}$  between sympatric as well as between allopatric populations, and between populations occupying the same type of environment and those occupying contrasting environments, or having similar or very different phenotypes (McGee *et al.* 2015; Meier *et al.* 2018). On the contrary, outlier loci repeatedly found in sympatric contrasts with different ecologies and phenotypes but absent in allopatric comparisons of similar ecologies and phenotypes, might be best explained by the action of



divergent selection between species pairs (Meier *et al.* 2018). The latter is exactly the pattern we detected in our whitefish species comparison, where we find shared outlier loci when comparing sympatric species with contrasting ecologies to each other but not when we compare allopatric populations or species with similar ecologies to each other (Figure 3). In conclusion, we specifically highlight four outlier loci shared between two sympatric contrasts as the best candidates of being shaped by divergent selection between whitefish species adapting to divergent spawning habitats and feeding ecologies. Importantly, all four loci have allele frequency differentials with the same sign between shallow and deep spawning species in both species pairs.

Aside from identifying genomic regions shaped by divergent selection, we also gain a better understanding of the speciation process by investigating the distribution of differentiated regions across the genome. Exceptionally differentiated loci detected in our study are widely spread across the genome and not confined to few genomic regions (see Figure 4 and Figure S3). This is similar to patterns detected in sympatric species in the North American sister lineage to European whitefish (Gagnaire *et al.* 2013) and suggests that divergence between whitefish species is polygenic on both continents. It is also consistent with studies showing that most ecologically important phenotypic traits in North American Lake Whitefish (e.g., growth rate, depth selection, activity) are quantitative traits with a polygenic basis involving multiple genes of moderate to small effect (Gagnaire *et al.* 2013; Rogers & Bernatchez 2007). A comparison of genetic differentiation patterns across sympatric species pairs in five North American lakes, in which a normal benthic and a dwarf limnetic species co-occur, showed only partial parallelism, often between only two lakes and only one occasion of parallelism across all five lakes (Gagnaire *et al.* 2013). This may be surprising given that all species pairs result from secondary contact between the same two glacial refugial lineages

(Acadian and Atlantic Mississippian). The low degree of parallelism in the divergence that is maintained in secondary contact suggests that independent outcomes of the coupling process following secondary contact of the same two genetic backgrounds (Acadian and Atlantic Mississippian) in different lakes produced partially different genetic architectures of postzygotic isolation (Gagnaire *et al.* 2013). The whitefish species investigated in our study have most likely diverged in primary sympatry and the genomic signatures of parallel divergence we observed are consistent with previous suggestions of speciation along a depth gradient of spawning sites for Alpine whitefish (Hudson *et al.* 2017; Ingram *et al.* 2012; Vonlanthen *et al.* 2012; Vonlanthen *et al.* 2009).

In our study the functional annotations available for a subset of our outlier loci further support a scenario of speciation along an ecological gradient. While functional annotations are not conclusive, they can support evidence and serve as starting points for new hypotheses. One of the four shared adaptive loci falls close to an immune relevant gene (*IgH locus B*). This is interesting as parasite studies on Alpine whitefish, including work on the Lake Zurich species pair, revealed that fish caught during the spawning season in shallow nets (mostly *C. duplex*) carried a high parasite burden, while fish from deep nets (>35 meters; mostly *C. zuerichensis*) hardly had any detectable macroparasites (Karvonen *et al.* 2013). Most fish macroparasites are limited to shallow waters as they often rely on invertebrates as intermediate hosts. This reveals one plausible biotic difference between water depth habitats within a lake, which could drive adaptive divergence (Karvonen & Seehausen 2012). However, other factors might also be relevant along a depth gradient; some abiotic ones might be light, pressure, and temperature. Interestingly, another tentative annotation for one of the four shared outliers is a gene with a function in vision (*retinitis pigmentosa 9*). Taken together, these functional annotations support previous notions that divergence between

sympatric whitefish is often driven by ecological factors along depth gradients (Hudson *et al.* 2017; Ingram *et al.* 2012; Vonlanthen *et al.* 2012; Vonlanthen *et al.* 2009). Perturbations of the habitats or of the conditions for feeding and/or reproducing in these habitats, is therefore expected to affect coexistence of, and genetic differentiation between the species (Seehausen *et al.* 2008). Once reproductive isolation breaks down, speciation might reverse and species eventually become homogenized due to introgressive hybridisation. As we do not currently have genetic data covering the pre-eutrophication period in Lake Zurich, the extent and rate of genetic homogenization is difficult to evaluate. Historical collections and perhaps ancient DNA from fossils burrowed in the lake sediments might allow in the future to directly access past conditions. Genetic data for neutral microsatellite loci from historical scale collections from another pre-alpine lake (Lake Constance) could indeed demonstrate loss of genetic differentiation between sympatric whitefish species, and the complete loss of one species, coinciding with the period of intense eutrophication (Vonlanthen *et al.* 2012).

## Conclusion

As a result of the combined action of ecological (species-specific demographic responses) and evolutionary (speciation reversal) processes, two adjacent and connected lakes (historically inhabited by the same three whitefish species) with distinct eutrophication histories nowadays differ in the relative abundance and genetic distinctiveness of sympatric whitefish species. Specifically, we found that the smallest and most deeply spawning of the three species is abundant in Lake Walen but rare in Lake Zurich, whereas the mid-size species that spawns at intermediate water depths is most abundant in Lake Zurich, but very rare in Lake Walen (and not genetically confirmed yet). The largest species that spawns inshore is moderately abundant in both lakes. In Lake Zurich, historically more heavily affected by eutrophication, we find evidence of considerable genetic admixture between the

abundant mid-size species and both, the large shallow-spawning species and the small deep-water spawning species. Admixture proportions imply gene flow from the smaller and deeper spawning into the larger and shallower spawning species in both cases, consistent with previous evidence for partial eutrophication-associated speciation reversal and the prediction that the hypoxic conditions associated with eutrophication force deeper spawning species to spawn shallower, i.e. on the spawning grounds of the shallower spawning species. It is possible that the small population of the very deep spawning species that we recovered from Lake Zurich is due to recent recolonization from Lake Walen after extinction of this species in Lake Zurich. Sequence data from historical collections or subfossils, or demographic modelling of whole genome sequence data from contemporary fish may help to resolve this unambiguously in the future. Genetic differentiation between sympatric species in both lakes is wide spread across the genome. Functional annotations of loci showing parallel differentiation in two species pairs suggest ecological habitat differences as one driving force of genomic divergence. Our results demonstrate genomic signatures of ecological speciation but also its sensitivity to anthropogenic perturbation of the ecological conditions. In general, our study supports previous evidence from the Alpine whitefish and other study systems suggesting that anthropogenic perturbations can have both direct ecological and indirect evolutionary consequences. This is important to note, as anthropogenic perturbations might not only cause demographic decline, but also affect the genetic diversity and composition of extant species and hence have potential long-term consequences for the remaining species.

## Acknowledgements:

We thank Salome Mwaiko for her support during the laboratory work. We thank Baenz Lundsgaard-Hansen, Pascal Vonlanthen, Alan Hudson, Timothy Alexander and the Projet Lac team and the local fishermen and fisheries wardens for kindly providing or sharing samples with us. We thank Oliver Selz, Timothy Alexander, and Carmela Doenz for sharing their expertise on the whitefish system. We thank Blake Matthews and Rishi De-Kayne for their critical input during the preparation of the manuscript. We thank three anonymous reviewers for their insightful comments, which helped to improve the clarity of the manuscript. Data analysis for this paper was supported by the collaboration with the Genetic Diversity Centre (GDC), ETH Zurich. We acknowledge Verena Kälin for the whitefish illustrations. Sampling was supported by the Bafu through “Projet Lac” and by Eawag Action field grant AquaDiverse to OS. This work was supported by the Swiss Science Foundation grant SNSF 31003A\_163446 awarded to PGDF.

## References:

- Alexander TJ, Vonlanthen P, Periat G, *et al.* (2015) Evaluating gillnetting protocols to characterize lacustrine fish communities. *Fisheries Research* **161**, 320-329.
- Alexander TJ, Vonlanthen P, Périat G, *et al.* (2017a) Artenvielfalt und Zusammensetzung der Fischgemeinschaft im Zürichsee. Projet Lac, Eawag, Kastanienbaum.
- Alexander TJ, Vonlanthen P, Seehausen O (2017b) Does eutrophication-driven evolution change aquatic ecosystems? *Philos Trans R Soc Biol Sci* **372**.
- Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**, e3376.
- Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD (2014) Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends Genet* **30**, 540-546.
- Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends Ecol Evol* **23**, 38-44.
- Behm JE, Ives AR, Boughman JW (2010) Breakdown in postmating isolation and the collapse of a species pair through hybridization. *Am Nat* **175**, 11-26.
- Bhat S, Amundsen P-A, Knudsen R, *et al.* (2014) Speciation reversal in European whitefish (*Coregonus lavaretus* (L.)) caused by competitor invasion. *PLoS ONE* **9**, e91208.

- Bittner D, Excoffier L, Largiadèr CR (2010) Patterns of morphological changes and hybridization between sympatric whitefish morphs (*Coregonus* spp.) in a Swiss lake: a role for eutrophication? *Mol Ecol* **19**, 2152-2167.
- Burgi HR, Heller C, Gaebel S, Mookerji N, Ward JV (1999) Strength of coupling between phyto- and zooplankton in Lake Lucerne (Switzerland) during phosphorus abatement subsequent to a weak eutrophication. *J Plankton Res* **21**, 485-507.
- Burri R (2017) Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters* **1**, 118-131.
- Camacho C, Coulouris G, Avagyan V, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421-421.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* **22**, 3124-3140.
- Charlesworth B, Morgan M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Res* **70**, 155-174.
- Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-1303.
- Crisci JL, Poh YP, Bean A, Simkin A, Jensen JD (2012) Recent progress in polymorphism-based population genetic inference. *J Hered* **103**, 287-296.
- Cruikshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* **23**, 3133-3157.
- Danecek P, Auton A, Abecasis G, et al. (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.
- De León L, Fernando, Raeymaekers J, A. M., Bermingham E, et al. (2011) Exploring possible human influences on the evolution of Darwin's finches. *Evolution* **65**, 2258-2272.
- Doenz C, Bittner D, Vonlanthen P, Wagner C, Seehausen O (2018) Rapid buildup of sympatric species diversity in Alpine whitefish. *Ecol Evol*.
- Douglas MR, Brunner PC, Bernatchez L (1999) Do assemblages of *Coregonus* (Teleostei: Salmoniformes) in the Central Alpine region of Europe represent species flocks? *Mol Ecol* **8**, 589-603.
- Duranton M, Allal F, Fraïsse C, et al. (2018) The origin and remolding of genomic islands of differentiation in the European sea bass. *Nature Communications* **9**, 2518.
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* **10**, 564-567.
- Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567 - 1587.
- Fatio V (1890) Faune des Vertebres de la Suisse. Vol. 5. *Histoire naturelle des Poissons Part 2*.
- Feulner PGD, Chain FJJ, Panchal M, et al. (2015) Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genet* **11**, e1004966.
- Feulner PGD, Gratten J, Kijas JW, et al. (2013) Introgression and the fate of domesticated genes in a wild mammal population. *Mol Ecol* **22**, 4210-4221.
- Gächter R, Müller B (2003) Why the phosphorus retention of lakes does not necessarily depend on the oxygen supply to their sediment surface. *Limnol Oceanogr* **48**, 929-933.

- Gaedke U, Schweizer A (1993) The first decade of oligotrophication in Lake Constance. *Oecologia* **93**, 268-275.
- Gagnaire P-A, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution* **67**, 2483-2497.
- Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. **1207**, arXiv:1207.3907.
- Gow JL, Peichel CL, Taylor EB (2007) Ecological selection against hybrids in natural populations of sympatric threespine sticklebacks. *J Evol Biol* **20**, 2173-2180.
- Grabenstein KC, Taylor SA (2018) Breaking barriers: Causes, consequences, and experimental utility of human-mediated hybridization. *Trends Ecol Evol* **33**, 198-212.
- Grant PR, Grant BR (2016) Introgressive hybridization and natural selection in Darwin's finches. *Biol J Linn Soc* **117**, 812-822.
- Hairston Jr NG, Lampert W, Cáceres CE, *et al.* (1999) Rapid evolution revealed by dormant eggs. *Nature* **401**, 446.
- Hasselmann DJ, Argo EE, McBride MC, *et al.* (2014) Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. *Mol Ecol* **23**, 1137-1152.
- Hendry AP (2009) Ecological speciation! Or the lack thereof? *Can J Fish Aquat Sci* **66**, 1383-1398.
- Hohenlohe PA, Phillips PC, Cresko WA (2010) Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *Int J Plant Sci* **171**, 1059-1071.
- Hudson A, Vonlanthen P, Bezault E, Seehausen O (2013) Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish. *BMC Evol Biol* **13**, 108.
- Hudson AG, Lundsgaard-Hansen B, Lucek K, Vonlanthen P, Seehausen O (2017) Managing cryptic biodiversity: Fine-scale intralacustrine speciation along a benthic gradient in Alpine whitefish (*Coregonus* spp.). *Evolutionary Applications* **10**, 251-266.
- Hudson AG, Vonlanthen P, Seehausen O (2010) Rapid parallel adaptive radiations from a single hybridogenetic ancestral population. *Proc R Soc Biol Sci Ser B* **278**, 58-66.
- Ingram T, Hudson AG, Vonlanthen P, Seehausen O (2012) Does water depth or diet divergence predict progress towards ecological speciation in whitefish radiations? *Evol Ecol Res* **14**, 487-502.
- Jones FC, Grabherr MG, Chan YF, *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61.
- Karvonen A, Lundsgaard-Hansen B, Jokela J, Seehausen O (2013) Differentiation in parasitism among ecotypes of whitefish segregating along depth gradients. *Oikos* **122**, 122-128.
- Karvonen A, Seehausen O (2012) The role of parasitism in adaptive radiations - when might parasites promote and when might they constrain ecological speciation? *International Journal of Ecology* **2012**, 20.
- Kleindorfer S, O'Connor JA, Dudaniec RY, *et al.* (2014) Species collapse via hybridization in Darwin's tree finches. *The American Naturalist* **183**, 325-341.
- Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* **23**, 482-490.
- Konijnendijk N, Joyce DA, Mrosso HDJ, Egas M, Seehausen O (2011) Community genetics reveal elevated levels of sympatric gene flow among morphologically similar but not among morphologically dissimilar species of Lake Victoria cichlid fish. *International Journal of Evolutionary Biology* **2011**, 12.

- Kottelat M (1997) European freshwater fishes. *Biologia, Bratislava*, 1-271.
- Kottelat M, Freyhof J (2007) *Handbook of European freshwater fishes* Publications Kottelat, Cornol and Freyhof, Berlin.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357.
- Laporte M, Rogers SM, Dion-Cote AM, *et al.* (2015) RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in Lake Whitefish (*Coregonus clupeaformis*) species pairs. *G3 (Bethesda)* **5**, 1481-1491.
- Lien S, Koop BF, Sandve SR, *et al.* (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200.
- Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* **21**, 936-939.
- Macqueen DJ, Johnston IA (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proc R Soc Biol Sci Ser B* **281**, 20132881.
- Marques DA, Lucek K, Meier JJ, *et al.* (2016) Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet* **12**, e1005887.
- Mayr E (1947) Ecological factors in speciation. *Evolution* **1**, 263-288.
- McGee M, D., Neches R, Y., Seehausen O (2015) Evaluating genomic divergence and parallelism in replicate ecomorphs from young and old cichlid adaptive radiations. *Mol Ecol* **25**, 260-268.
- McKenna A, Hanna M, Banks E, *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303.
- Meier JJ, Marques DA, Mwaiko S, *et al.* (2017) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications* **8**, 14363.
- Meier JJ, Marques DA, Wagner CE, Excoffier L, Seehausen O (2018) Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol Biol Evol* **35**, 1489-1506.
- Michel AP, Sim S, Powell THQ, *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proc Natl Acad Sci* **107**, 9724-9729.
- Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc Biol Sci* **367**, 409-421.
- Nielsen R (2005) Molecular signatures of natural selection. *Annu Rev Genet* **39**, 197-218.
- Nosil P (2012) *Ecological Speciation* Oxford University Press.
- Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Biol Sci* **365**, 185-205.
- Ostbye K, Bernatchez L, Naesje TF, Himberg KJM, Hindar K (2005) Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Mol Ecol* **14**, 4371-4387.
- Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945 - 959.
- Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Gen* **81**, 559-575.
- R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ravinet M, Faria R, Butlin RK, *et al.* (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol* **30**, 1450-1477.



- Recknagel H, Elmer KR, Meyer A (2013) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3 (Bethesda)* **3**, 65-74.
- Renaut S, Grassa CJ, Yeaman S, *et al.* (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat Commun* **4**, 1827.
- Riesch R, Muschick M, Lindtke D, *et al.* (2017) Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution* **1**, 0082.
- Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs. *Mol Biol Evol* **24**, 1423-1438.
- Rundle HD, Nosil P (2005) Ecological speciation. *Ecol Lett* **8**, 336-352.
- Schluter D (1996) Ecological speciation in postglacial fishes. *Philos Trans R Soc Lond B Biol Sci* **351**, 807-814.
- Schluter D (2001) Ecology and the origin of species. *Trends Ecol Evol* **16**, 372-380.
- Seehausen O (2006) Conservation: Losing biodiversity by reverse speciation. *Curr Biol* **16**, R334-R337.
- Seehausen O, Butlin RK, Keller I, *et al.* (2014) Genomics and the origin of species. *Nat Rev Genet* **15**, 176-192.
- Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol Ecol* **17**, 30-44.
- Seehausen O, vanAlphen JJM, Witte F (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808-1811.
- Smith S, H. (1964) Status of the deepwater cisco population of Lake Michigan. *Trans Am Fish Soc* **93**, 155-163.
- Sommer U, Gaedke U, Schweizer A (1993) The first decade of oligotrophication of Lake Constance. *Oecologia* **93**, 276-284.
- Soria-Carrasco V, Gompert Z, Comeault AA, *et al.* (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**, 738-742.
- Stapley J, Reger J, Feulner PGD, *et al.* (2010) Adaptation genomics: the next generation. *Trends Ecol Evol* **25**, 705-712.
- Steinmann P (1950) Monographie der schweizerischen Koregonen. *Schweizerische Zeitschrift für Hydrologie* **12**, 109-112.
- Szulkin M, Gagnaire PA, Bierne N, Charmantier A (2016) Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. *Mol Ecol* **25**, 542-558.
- Taylor EB, Boughman JW, Groenenboom M, *et al.* (2006) Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Mol Ecol* **15**, 343-355.
- Todd TN, Stedman RM (1989) Hybridization of ciscoes (*Coregonus* spp.) in Lake Huron. *Can J Zool* **67**, 1679-1685.
- Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* **3**, e285.
- Vonlanthen P, Bittner D, Hudson AG, *et al.* (2012) Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482**, 357-362.
- Vonlanthen P, Périat G, Seehausen O, *et al.* (2014) Artenvielfalt und Zusammensetzung der Fischgemeinschaft im Walensee. *Projet Lac, Eawag, Kastanienbaum*.
- Vonlanthen P, Roy D, Hudson AG, *et al.* (2009) Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *J Evol Biol* **22**, 498-514.

- Accepted Article
- Wagler E (1937) Die Coregonen in den Seen des Voralpengebietes. IX. Die Systematik der Voralpencoregonen. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* **35**, 345-446.
- Wagner CE, Mandeville EG (2017) Speciation, species persistence and the goals of studying genomic barriers to gene flow. *J Evol Biol* **30**, 1512-1515.
- Wolf JBW, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet* **18**, 87-100.
- Woods PJ, Müller R, Seehausen O (2009) Intergenomic epistasis causes asynchronous hatch times in whitefish hybrids, but only when parental ecotypes differ. *J Evol Biol* **22**, 2305-2319.
- Wu CI (2001) The genic view of the process of speciation. *J Evol Biol* **14**, 851-865.
- Yasuike M, de Boer J, von Schalburg KR, *et al.* (2010) Evolution of duplicated IgH loci in Atlantic salmon, *Salmo salar*. *BMC Genomics* **11**, 486.
- Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc Natl Acad Sci U S A* **110**, E1743-1751.
- Zheng X, Levine D, Shen J, *et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-3328.

Data accessibility:

Raw read sequencing files (fastq files for all 180 individuals) are deposited on short read archive SRA (PRJNA485027 and SRP156755). The sequences of the reference loci (fasta format) and genotypes (vcf format) are available at the dryad repository doi:10.5061/dryad.gp25h48.

Author's contributions:

PGDF produced the data, analysed the data, and wrote the manuscript. PGDF and OS designed the experiment, discussed data and results, and edited and revised the manuscript.

Tables and Figures:

**Table 1: Genetic diversity of and differentiation between whitefish populations in lakes Walen and Zurich and the connecting Linth Canal.** Genetic diversity  $\pi$  for each of six populations are given in the diagonal. Above the diagonal relative differentiation  $F_{ST}$  is given for all pairwise comparison, respective p-values are given below the diagonal. Samples sizes for each population are given in brackets and values of the *C. heglingus* population of Lake Zurich are faded to reflect the low sample size.

**Figure 1: Genetic clustering (STRUCTURE) of whitefish species in lakes Walen and Zurich.** (a) Genetic clustering was based on genotypes of 138 whitefish individuals at 16,173 loci. Likelihood values of ten replicated clustering runs for K clusters from one to six are summarised. Circles indicate mean likelihood values for a given K, while bars represent the variance across ten replicated runs. (b) Allocations of individuals (horizontal bars) to genetic clusters and respective admixture proportions are given for the K clusters with the best likelihood and least variance across runs. For each K the run with the highest likelihood value was plotted. Individuals of the same species and sampling location were plotted next to each other.

**Figure 2: Distinction of whitefish species in lakes Walen and Zurich.** (a) Distributions of standard length for all three species (pink - *C. heglingus*, green - *C. zurichensis*, blue - *C. duplex*) are overlapping but differ in their means (one-way anova,  $p < 2.2e-16$ ,  $F = 114.4$ ,  $DF = 2$  and 128). Scientific illustrations (©Verena Kälin) outline additional differences in appearance between the species. (b) PCA plot illustrating genetic differences between 138 whitefish based on 16,173 loci evaluated. Individuals of the three species are colour coded (see a), sampling locations are indicated by different symbols (diamonds Lake Walen, triangles Linth Canal, squares Lake Zurich). The three *C. heglingus* of Lake Zurich are further highlighted by filled symbols for ease of identification. (c) Map of lakes Zurich and Walen connected by the Linth. The genetic composition (admixture proportions as estimated by STRUCUTRE) of whitefish sampled at each of the three locations is plotted alongside. Each bar represents an individual and the three colours indicate the proportion of each of three different genetic contributions.

**Figure 3: Comparison between two pairwise estimates of genetic differentiation  $F_{ST}$  across 16,173 loci.** (a) Pairwise  $F_{ST}$  values are not correlated ( $r^2 = 0.029$ ,  $p = 0.0001$ ) when the two pairwise comparisons are comparisons between allopatric populations of the same or ecologically similar species (b) but are positively correlated ( $r^2 = 0.399$ ,  $p << 0.0001$ ) when the two comparisons are pairwise comparisons of sympatric species pairs with contrasting ecology. Loci with an exceptional differentiation as determined by an outlier test are highlighted in colour (blue – allopatric contrast between *C. duplex* from Lake Walen to Lake Zurich, pink – allopatric contrast between *C. heglingus* from Lake Walen and *C. zurichensis* from Lake Zurich, green – sympatric contrast within Lake Walen comparing *C. heglingus* and *C. duplex*, light blue – sympatric contrast within Lake Zurich comparing *C. zurichensis* and *C. duplex*). Turquoise stars highlight four outlier loci shared in both sympatric comparisons.

**Figure 4: Distribution of genetic differentiation  $F_{ST}$  along the genome based on pairwise comparisons between sympatric whitefish species in lakes Zurich and Walen.** (a) Both sympatric contrasts (same values as in Figure 2b) are plotted along the salmon genome. Loci were position onto the salmon reference genome (along 29 chromosomes) by sequence

similarity. Alternating shadings (white versus light grey) indicated chromosomal boundaries. Loci with an exceptional differentiation as determined by an outlier test are highlighted in colour (green – sympatric contrast within Lake Walen comparing *C. heglingus* and *C. duplex*, light blue – sympatric contrast within Lake Zurich comparing *C. zurichensis* and *C. duplex*). The positions of three of the four shared outlier (turquoise stars in Figure 2b) are highlighted in with turquoise bars. (b) No evidence for any increase in genetic differentiation  $F_{ST}$  at loci showing sequence similarity in the proximity of known shape QTLs. Genetic differentiation  $F_{ST}$  is not more different at 681 loci (qtl), which mapped to published whitefish scaffolds that included QTLs for shape in the North American sister genus (Laporte et al G3 2015), then in the 15'485 other loci (ao) that do not match with those scaffolds. This is true for both, Lake Zurich and Lake Walen (Lake Zurich:  $t = 1.3829$ ,  $df = 738.745$ ,  $p = 0.1671$ , Lake Walen:  $t = 1.6884$ ,  $df = 741.008$ ,  $p = 0.09176$ ) pairwise comparisons between sympatric whitefish species, which differ significantly in their size (see Figure 1a).

**Table S1: Detailed information on the sampling locations for all whitefish individual.** Database ID and lab ID for each individual are given. Shading indicates individuals for which genotyping failed. Aside for species assignment, total length, sampling location, date and depth, as well as DNA extraction methods are given.

**Table S2: Salmon reference position and available annotations for any outlier loci.** Outlier loci of the four pairwise comparisons are given on four sheets. For each consensus ID of each RAD locus, a salmon reference genome (Lien *et al.* 2016) position is given, if the locus could be mapped. The same if the locus could be mapped to any of the publicly available scaffold for *C. clupeaformis* (Laporte et al. 2015). Heterozygosity, genetic differentiation, and the p-value determining the outlier status for each locus are given as well. Any best blast hits are given indicating their accession number, description, and alignment statistics.

**Table S3: Allele frequencies of reference and alternative allele given for the four shared outlier loci.** For each shared outlier loci, number of observed alleles (N\_CHR), and frequency of reference (FREQ\_REF) and alternative (FREQ\_ALT) are given as observed in *C. duplex* from Lake Walen and Lake Zurich and *C. heglingus* from Lake Walen and *C. zurichensis* from Lake Zurich.

**Figure S1: Genetic clustering (STRUCTURE) of whitefish species in lakes Walen and Zurich.** Genetic clustering was based on genotypes of 138 whitefish individuals at 16,173 loci. Allocations of individuals (horizontal bars) to genetic clusters and respective admixture proportions are given for the K clusters from two to six. For each K the run with the highest likelihood value was plotted. Individuals of the same species and sampling location were plotted next to each other.

**Figure S2: Distribution of genetic differentiation  $F_{ST}$  along the genome based on pairwise comparisons between allopatric whitefish populations between lakes Zurich and Walen.** Both allopatric contrasts (same values as in Figure 2a) are plotted along the salmon genome. Loci were position onto the salmon reference genome (along 29 chromosomes) by sequence similarity. Alternating shadings (white versus light grey) indicated chromosomal boundaries. Loci with an exceptional differentiation as determined by an outlier test are highlighted in colour (blue – allopatric contrast between *C. duplex* from Lake Walen to Lake Zurich, pink – allopatric contrast between *C. heglingus* from Lake Walen and *C. zurichensis* from Lake Zurich).

**Figure S3: Pairwise association between any of the loci showing exceptional differentiation between sympatric whitefish species in lakes Zurich and Walen.** The pairwise association of markers within each of four populations is given in two heatmaps indicating linkage  $r^2$ . Loci plotted are outlier in the comparison between sympatric whitefish species in Lake Zurich (a) or in Lake Walen (b). Measurements of linkage disequilibrium ( $r^2$ ) in the two *C. duplex* populations are plotted above the diagonal, below estimated for the populations of *C. heglingus* and *C. zurichensis* are plotted. No systematic strong associations could be detected, suggesting that none of the outlier loci are physically linked with each other, matching the distribution pattern observed in Figure 3.

		<i>C. duplex</i> Walen (n=15)	<i>C. duplex</i> Linth (n=14)	<i>C. duplex</i> Zurich (n=39)	<i>C. heglingus</i> Walen (n=23)	<i>C. zuerichensis</i> Zurich (n=44)	<i>C. heglingus</i> Zurich (n=3)
<i>C. duplex</i>	Walen	0.0042	0.004	0.023	0.119	0.087	0.120
<i>C. duplex</i>	Linth	$p = 0.07$	0.0043	0.004	0.101	0.066	0.112
<i>C. duplex</i>	Zurich	$p < 0.01$	$p < 0.01$	0.0044	0.087	0.052	0.089
<i>C. heglingus</i>	Walen	$p < 0.01$	$p < 0.01$	$p < 0.01$	0.0040	0.037	0.006
<i>C. zuerichensis</i>	Zurich	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	0.0042	0.040
<i>C. heglingus</i>	Zurich	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p = 0.62$	$p < 0.01$	0.0038





