DR KAY LUCEK (Orcid ID : 0000-0002-2253-2556)

Article type : Research Papers

Distinct colonization waves underlie the diversification of the freshwater sculpin (*Cottus gobio*) in the Central European Alpine region

Kay Lucek[1,2,3,*], Irene Keller[4], Arne W. Nolte[5,6], Ole Seehausen[2,3]

[1] Department of Environmental Sciences, University of Basel, Schönbeinstrasse 6, 4056 Basel, Switzerland

[2] Department of Aquatic Ecology and Macroevolution, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland

[3] Department of Fish Ecology and Evolution, EAWAG Swiss Federal Institute of Aquatic Science and Technology, Center of Ecology, Evolution and Biochemistry, Seestrasse 79, CH-6047 Kastanienbaum, Switzerland

[4] Department of BioMedical Research and Swiss Institute of Bioinformatics, University of Bern, Murtenstrasse 40, 3012 Bern, Switzerland

[5] Carl von Ossietzky University Oldenburg, Institute for Biology, Carl von Ossietzky Str. 9-11, 26111 Oldenburg, Germany

[6] Department for Evolutionary Genetics, Max-Planck Institute for Evolutionary Biology, August Thienemann Strasse 2, 24306 Plön, Germany

* kay.lucek@unibas.ch

Abstract

Ecological speciation and adaptive radiation are key processes shaping northern temperate freshwater fish diversity. Both often involve parapatric differentiation between stream and lake populations and less often, sympatric intralacustrine diversification into habitat- and resource-associated ecotypes. However, few taxa have been studied, calling for studies of others to investigate the generality of these processes. Here, we test for diversification within catchments in freshwater sculpins in a network of peri-Alpine lakes and streams. Using 8,047 and 13,182 restriction site associated (RADseq) SNPs respectively we identify three deeply divergent phylogeographic lineages associated with different major European drainages. Within the Aare/Rhine catchment, we observe populations from geographically distant lakes to be genetically more similar to each other than to populations from nearby streams. This pattern is consistent with two distinct colonization waves, rather than by parapatric ecological speciation after a single colonization wave. We further find two distinct depth distribution modes in three lakes of the Aare catchment, one in very shallow and one in very deep water, and significant genome-wide differentiation between these in one lake. Sculpins in the Aare catchment appear to represent an early stage adaptive radiation involving the evolution of a lacustrine lineage distinct from parapatric stream sculpins, and the repeated onset of depth-related intralacustrine differentiation.

Key words: ecological speciation, intralacustrine radiation, postglacial diversification, *Cottus gobio*, character displacement

Introduction

Adaptive radiation, i.e. the diversification of a single taxon into phenotypically, ecologically and genetically differentiated populations and ultimately species, is a key process in the evolution of biodiversity (Schluter, 2000b; Nosil, 2012). Factors that promote or impede adaptive radiation are best studied at early stages of the process where phenotypic and genotypic segregation are likely to be incomplete (Yoder *et al.*, 2010). The relevance of adaptive radiations for the generation of biodiversity has been highlighted for fishes in insular freshwater habitats, like deep lakes in the tropics or temperate zones (Seehausen & Wagner, 2014). Postglacial diversification amongst the latter may occur between distinct habitats, such as lakes and streams as well as within the same macrohabitat, where for example species within lakes segregate according to water depth and/or trophic niches (Hendry, 2009; Seehausen & Wagner, 2014). Despite the prominent discussion of these processes in speciation research, postglacial ecological speciation and adaptive radiation have only been identified in very few fish taxa, specifically in several salmonid fishes: e.g. whitefish (Bernatchez, 2004; Hudson *et al.*, 2007) or arctic charr (Jonsson & Jonsson, 2001) and in the threespine stickleback (McKinnon & Rundle, 2002; see Seehausen & Wagner (2014) for a review). Few examples are known in other temperate fish taxa (Seehausen & Wagner, 2014). To understand the extent to which lineage specific traits facilitate or constrain postglacial diversification, additional taxa in similar and unimpacted environments need to be investigated (Seehausen & Wagner, 2014).

Because of its broad geographic distribution, the European bullhead or freshwater sculpin (*Cottidae*: *Cottus gobio* Linnaeus, 1758) is a good candidate to test for postglacial diversification. It is the most widely distributed of the 16 known and described, predominantly allopatric European freshwater *Cottus* species. It is thought to predominantly occur in streams from Scandinavia to Northern Italy, characterized by fast flowing, well-oxygenated water and structured benthic habitats where it feeds predominantly on benthic invertebrates (Kottelat & Freyhof, 2007; Goto *et al.*, 2015). In other parts of the northern hemisphere other sculpin

species are also reported from littoral zones of large cold lakes with some rare accounts of profundal lacustrine occurrences (Goto *et al.*, 2015). Albeit, *Cottus gobio* has occasionally been reported from the littoral zones of a lake (Wanzenböck *et al.*, 2000; Kontula & Vainola, 2001), profundal populations have not been studied in detail so far (Goto *et al.*, 2015).

European sculpins have been the focus of several biogeographic studies because of their wide distribution where population structure is thought to be relatively little impacted by contemporary human translocations. This is due to their current economical insignificance in contrast to other co-occurring freshwater fish such as trout (Hänfling & Brandl, 1998a; b; Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008). With one exception (Kontula & Vainola, 2001), biogeographic studies have involved almost only riverine populations and suggest that much of the currently occupied geographic range has been colonized during postglacial range expansions after the last glacial maximum ~12,000 years ago (Englbrecht *et al.*, 2000). Contemporary populations derive from distinct glacial refugia that are linked to the major European drainages (Englbrecht *et al.*, 2000; Nolte *et al.*, 2005), but natural watershed crossings have occurred in the course of the last glacial retreat (Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008). The biogeographic structure of European sculpins is thus largely shaped by differentiation between drainages, where recent taxonomic work suggests that separate drainages may often contain distinct species (Freyhof *et al.*, 2005). Genetic admixture between such species has at least in one case led to the emergence of a hybrid species that occupies an environment distinct from either parental species, and distinct from all other species in the genus (Nolte *et al.*, 2005).

Population genetic structure in sculpins within a given drainage is also common and thought to reflect low dispersal range, geographic isolation or anthropogenic river fragmentation (Hänfling & Weetman, 2006; Junker *et al.*, 2012). The possibility of habitat dependent divergence has, however, received only little attention among European sculpins (Davey *et al.*, 2005; Goto *et al.*, 2015), despite the fact that sculpins evolved an impressive

intralacustrine species radiation in Lake Baikal, where they form a species flock comprising at least 33 taxa with diverse adaptations (Kontula *et al.*, 2003; Goto *et al.*, 2015).

The Alps are one area for which biogeography and molecular data imply relatively recent colonization (Slechtová *et al.*, 2004). The evolutionary dynamics of sculpins in this area are therefore likely recent (Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008). Despite their broad distribution, occurring in both streams (Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008) and lakes (Alexander *et al.*, 2015), the potential for habitat dependent differentiation has not been assessed. The deep, cold and oligotrophic lakes in this region moreover make for suitable sculpin habitat, distinct from streams and have been in continuous existence since the retreat of the glaciers. Here, we take advantage of a recent fish diversity assessment of peri-alpine lakes around the European Alps that found sculpins to occur in many of the lakes with individuals occupying both shallow and very deep water (Alexander *et al.*, 2015; *see also* Figure 5c). Using many thousand genomic markers we first assess the population structure of sculpins across Switzerland and subsequently test for a genomic footprint of habitat-dependent differentiation between lake dwelling populations and the common stream populations within our most densely sampled catchment, the Aare (Figure 1). To do so, we use individual based phylogenomic approaches (Huson & Bryant, 2006), assignment statistics (Raj *et al.*, 2014), outlier detection analyses (Stucki *et al.*, 2017) as well as population based estimates of genomic differentiation between habitats (F_{ST} and F'_{ST}). In the biodiversity assessment, we observed a bimodal depth distribution of sculpins within three lakes of the Aare catchment, with sculpins occupying the profundal and the littoral habitat but not the intermediate depths (Alexander *et al.*, 2015). We thus used our genomic data to further test if this depth related ecological divergence is associated with genetic differentiation, consistent with the onset of intralacustrine differentiation. For each lake we employed outlier detection analyses (Stucki *et al.*, 2017) and estimated patterns of genetic differentiation across the genome between profundal and littoral individuals.

Material and Methods

Sample collection

We collected 437 sculpins from 31 sites with different sampling methods: We used electrofishing in streams and in lakes as well as gill nets that were randomly distributed across all habitats and depths, including the littoral and profundal zones (Table S1; Figure 1). Additional fishing effort for the very shallow (<2m) parts of the lake was done using electrofishing also at randomly selected places (see Alexander *et al.*, 2015 for details). Lake dwelling specimens were collected during a large fish diversity assessment in perialpine lakes - "*Projet Lac*" - where for each specimen the catching depth and exact location were recorded (Alexander *et al.*, 2015). Our sampling design further covered distinct drainages draining into the North Sea (Aare/Rhine catchment), the Mediterranean Sea (Doubs/Rhone catchment) or the Adriatic Mediterranean Sea (Maggia/Po and Poschiavino/Po catchments) see Table S1 & Figure 1 for details. Fish were euthanized with an overdose of MS222 and a fin clip was taken and stored in absolute ethanol for further genetic analyses.

Molecular methods

Restriction site associated DNA (RAD) markers were generated following Marques *et al.* (2016), using 400 ng genomic DNA per sample, which was digested for 12 hours with four units of *Sbf*I-HF (New England Biolabs, Switzerland). Individuals were multiplexed after ligating sample-specific P1 adapters (synthesized by Microsynth, Switzerland) with custom 5-8 base pair barcodes that had a minimal distance of two bases between any two barcodes. The pooled DNA was subsequently sheared on an S220 series Adaptive Focused Acoustic ultra-sonicator (Covaris Inc., USA) with the manufacturer's settings for a mean fragment size of 400bp. Sheared fragments between 300–500bp were size-selected on a 1.25% agarose gel. Libraries were PCR

amplified in four aliquots with 50 µl reaction volumes each. All replicates were combined prior to the final size selection step. Each library was single-end sequenced on one lane of an Illumina HiSeq 2000 platform together with 5–20% bacteriophage PhiX genomic DNA (Illumina Inc., USA) to increase complexity at the first 10 sequenced base pairs. We generated seven libraries, each containing 60-64 individually barcoded specimens.

Data analysis

We filtered the raw sequencing reads from each individual for an intact *Sbf*I restriction site, de-multiplexed and barcode-trimmed them using PROCESS_RADTAGS 1.26 (Catchen *et al.*, 2013). We aligned reads for each individual against a repeat-masked reference assembly of *Cottus rhenanus*, a species that was only recently split from *C. gobio* (Freyhof *et al.*, 2005). The assembly consisted of 88,957 contigs (N50=6,900bp; Smolka *et al.*, 2015). We used end-to-end alignment in BOWTIE2 2.2.6 with default parameters (Langmead & Salzberg, 2012). Contigs were moreover anchored against the threespine stickleback (*Gasterosteus aculeatus*) genome allowing us to obtain SNP position relative to the latter (Cheng *et al.*, 2013; Dennenmoser *et al.*, 2017). Raw sequencing reads were also aligned against the PhiX 174 reference genome (accession: NC_001422; Sanger *et al.*, 1977), known variation was masked and PhiX-alignments were used to create a base quality score recalibration table for each library using BASERECALIBRATOR from GATK 3.2-0-g289df4b (McKenna *et al.*, 2010). We subsequently recalibrated the base quality scores of sculpin alignments to remove potential library effects with the GATK tool PRINTREADS.

We used the GATK tool UNIFIEDGENOTYPER to call variants and genotypes in a combined fashion for all individuals, using the following parameters: minimal phred-scaled base quality score threshold of 20, a genotype likelihood model calling both SNPs and insertions/deletions (indels). Using VCFTOOLS v0.1.12b (Danecek *et al.*, 2011), genotypes with quality < 28 or depth < 6 were set to missing. Individuals with more than 50% missing data were subsequently

removed from the data set. Variants with quality < 28 or > 50% missing genotypes per sampling site, monomorphic sites, SNPs with > 2 alleles, indels and SNPs 10 bp around indels were further removed from the dataset. This filtering step was performed using either all specimens or separately for each drainage (North Sea, Mediterranean Sea, Adriatic Mediterranean Sea).

To assess the potential effects of gene flow among populations that can result in discordance in the phylogenetic signal (Seehausen, 2004), we reconstructed a network using SPLITSTREE4 with a GTR model (Huson & Bryant, 2006).

Further analyses were done on datasets for which we applied a minor allele frequency filter of 5% to increase statistical power and to accommodate the underlying assumptions of the respective software packages. To infer population structure and individual assignment to genetic groups, we first used a Bayesian framework implemented in the program FASTSTRUCTURE (Raj *et al.*, 2014), which was run using either all individuals combined or separately for each drainage system to infer potential genetic substructure that may be hidden in the overall analysis. For the Aare/Rhine catchment, we removed individuals from Lake Geneva (Rhone, Mediterranean drainage), as contemporary natural gene flow between the two watersheds is absent (Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008). In each case, we ran an admixture model with a 'simple prior', i.e. a flat prior over population-specific allele frequencies at each locus (Raj *et al.*, 2014). We conducted analyses for different numbers of genetic clusters (K), ranging from K = 1 to an initial arbitrary upper limit of K = 20. The *chooseK.py* script of the FASTSTRUCTURE software package was then used to infer the best-supported K based on model complexity that maximizes marginal likelihood (Raj *et al.*, 2014).

We also calculated the average levels of pairwise genetic differentiation (F_{ST}) among our sampled populations using pairwise locus-by-locus AMOVAs in ARLEQUIN 3.5.1.4 (Excoffier & Lischer, 2010), where we used a bootstrap approach with 1,000 replicates to infer significances. Within the Aare catchment F_{ST} estimates between lakes, between streams and between lake and streams were subsequently compared using an ANOVA with a Tukey HSD *post hoc* correction.

Because F_{ST} estimates can be biased towards lower values in cases where the level of heterozygosity within populations is high (Meirmans & Hedrick, 2011), the analysis for the Aare catchment was repeated using F'_{ST} (Hedrick, 2005) calculated in R with the package DIVERSITY (Keenan *et al.*, 2013). Using ADEGENET 1.4-2 (Jombart *et al.*, 2010), we calculated the expected heterozygosity (*Hs*) within each site and compared *Hs* between lake and stream populations within the Aare/Rhine catchment using a *t*-test. The single individual from Lake Neuchatel was omitted from these analyses (Table S1).

We used SAMβADA 0.5.1 (Stucki *et al.*, 2017) to test for alleles that show a significant association with lake or stream habitat within the Aare/Rhine catchment. SAMβADA uses logistic regression models to identify statistical associations between the presence of an allele and a given response variable. The significance of the resulting univariate models was assessed using the implemented log-likelihood ratios (G-scores) providing *p* values based on χ2-tests (Stucki *et al.*, 2017). We then tested whether genes associated with outliers were enriched for gene ontology (GO) terms using the STRING v10.5 database (Szklarczyk *et al.*, 2015) with a Bonferroni-corrected alpha level of 0.05. We further highlight loci with alleles that show significant differentiation between lake and stream habitats as suggested by SAMβADA along the genome for the three geographically close lake-stream contrasts, i.e. Aare-KM vs. Lake Thun, Aare-RU vs. Lake Lucern, Aare-RU vs. Lake Zurich.

Lastly, we performed a Principal Component (PC) analysis to test for putative habitat dependent population structure, i.e. between all lakes and all streams within the Aare/Rhine catchment using the package SNPRELATE (Zheng *et al.*, 2012). We performed PCAs using either all polymorphic SNPs within the Aare/Rhine catchment that were included in the FASTSTRUCTURE analysis or only those identified as outliers by SAMβADA. In either case, the scores of the two leading PCs were further tested for a difference between lake and stream dwelling individuals using one-way ANOVAs in R.

Given the bimodal depth distribution of sculpins observed in the field within lakes Lucern, Thun and Walen (Figure 6c), we tested for intralacustrine genetic differentiation between littoral and profundal dwelling specimens: We first estimated differentiation based on SNP sites that were polymorphic within each lake using a discriminant analysis of principal components (DAPC; Jombart *et al.* 2010). For each lake, DAPC was based on the 10 leading PC axes, where the discriminant function separated individuals that were caught within the shallower (0-50m depth) distribution mode (Figure 6c) and specimens that were caught at a deeper depth. The scores of each PC were further compared between shallow and profundal individuals using one-way ANOVAs in R. We then estimated pairwise F_{ST} estimates between shallow and profundal individuals with pairwise locus-by-locus AMOVAs in ARLEQUIN. Significance was assessed with 1,000 bootstrap replicates. Lastly, we tested for depth-associated alleles within each lake with SAMβADA using the recorded individual catch depth as response variable for each allele.

Results

Sequencing resulted in 583 million raw reads of which 356 million mapped uniquely to the sculpin reference assembly (average mapped reads per individual: 816,100; 95% confidence interval: 23,690-2,495,300 reads). Due to an excess of missing data, we subsequently removed 82 individuals from our data set. This resulted in a filtered data set of 8,047 or 13,182 polymorphic SNPs respectively for 355 individuals (Table S1), when applying either a minor allele frequency filter of 5% or not.

Biogeography

The combined phylogenetic and population genetic analyses suggest the existence of three major lineages related to the three main drainages that we studied. SPLITSTREE resolved three deeply divergent clades: (1) an Aare/Rhine clade including Lake Geneva, (2) a Doubs clade and (3) an Adriatic clade, which was further subdivided into individuals from the Ticino and the Poschiavino river catchments (Figure 2). FASTSTRUCTURE identified further genetic substructure, with the best supported number of genetic clusters K=7 for the overall data set (Figure 3a). Individuals from the Doubs and the Adriatic drainages each fell into distinct clusters. River dwelling populations from the Aare/Rhine catchment fell into three genetic clusters, each related to one stream with restricted admixture between them. The exception is the Aare-RU population, which showed mostly admixture or shared ancestry with lake populations (see also Figure 5). Conversely, lake dwelling individuals from the Aare/Rhine catchment were predominantly assigned to only two genetic clusters, both distinct from the three clusters of Aare/Rhine stream fish: one comprising sculpins from Lake Lucerne and the other comprising individuals from most other lakes sampled within the Aare/Rhine catchment. Admixture between distinct genetic clusters was apparent in lakes Constance and Geneva.

When we ran FASTSTRUCTURE separately for each drainage, the Adriatic Mediterranean Sea drainage became further subdivided into two genetic clusters, reflecting different river catchments, i.e. the Ticino and the Poschiavino. The Doubs population also became subdivided into three genetic clusters, associated with the upper, middle and lower reaches of the River Doubs (Figure 3a). The Aare/Rhine catchment became subdivided into a total of ten genetic clusters. K=10 remained the best number of clusters even if the four individuals from Lake Geneva were included (results not shown). These additional clusters reflect geographic structure, the propensity for ongoing gene flow, as well as lake versus stream population structure. For example, the genetic cluster comprising most lake populations in the overall analysis (see above) became subdivided into two clusters comprising either specimens from the

geographically close-by and interconnected lakes Zurich and Walen or Brienz and Thun respectively (Figure 1; 3a). Further genetic substructure likewise became apparent amongst sites from the Thur and Sense stream catchments.

Genetic differentiation along the lake-stream axis

Within the Aare/Rhine catchment, we found different lines of evidence for genetic differentiation between lake and stream populations: First, both SPLITSTREE and FASTSTRUCTURE find that most lake populations cluster closely together despite being geographically distant (e.g. lakes Thun and Zurich, Figure 1), whereas geographically proximate stream populations were genetically distinct from the lake populations, e.g. the population of the Upper Aare River versus that from Lake Thun (Figures 2 & 3a). Secondly, both pairwise F_{ST} and F'_{ST} estimates were significantly lower within the Aare catchment for comparisons between geographically distant lake populations than between lake and stream populations (TukeyHSD; F_{ST}: p<0.001; F'_{ST}: p<0.001) or between stream populations (F_{ST}: p<0.001 & Figure 3b; F'_{ST}: p<0.001 & Figure S1).

SAMβADA identified 781 out of 14,927 (i.e. 5.2%) SNP alleles that differ significantly between lake and stream dwelling individuals within the Aare/Rhine catchment. Of these, 328 alleles overlapped with 237 genes of which 194 (82%) were annotated in the STRING database. However, we did not find enrichment for gene ontology categories among these. Significant differences in allele frequencies between lake and stream dwelling individuals occurred across the genome. Although we used individuals from across lakes and streams in SAMβADA, the level of genetic differentiation was consistently higher for SAMβADA outlier loci identified than the rest of the genome for all three pairwise comparisons (Aare-KM vs. Lake Thun: $F_{1,1970}$= 29.4, p < 0.001; Aare-Ru vs. Lake Lucern: $F_{1,2349}$= 15.7, p < 0.001; Aare-Ru vs. Lake Zurich: $F_{1,2251}$= 9.7, p = 0.002; Figure 4). Lastly, heterozygosity was overall significantly higher in lakes than in streams

within the Aare/Rhine catchment ($t_{1,15}$=3.10, p=0.014; average *Hs* lakes: 0.17±0.04 SD, average *Hs* streams: 0.11±0.03 SD).

The PC analyses of polymorphic SNPs within the Aare/Rhine catchment further highlight the genetic distinctiveness of all lake populations on the one hand and stream populations on the other within this catchment (Figure 5). The latter being further enhanced when using SAMβADA outliers (Figure 5). The residuals differed significantly between habitats for both datasets (all SNPs: PC1 – $F_{1,217}$ = 167.3, p < 0.001; PC2 – $F_{1,217}$ = 42.62, p < 0.001; outlier SNPs: PC1 – $F_{1,217}$ = 459.4, p < 0.001; PC2 – $F_{1,217}$ = 24.2, p < 0.001). Consistent with our FASTSTRUCTURE analysis (Figure 3) we find all lake-dwelling populations to be more closely related to each other than to the stream dwelling clusters, with the exception of the stream dwelling Aare-RU population that clusters with the lake cluster (Figure 5).

Contrary to the Aare/Rhine catchment, we did not find habitat-associated patterns of differentiation in our only lake-stream pair belonging to the Adriatic Mediterranean Sea watershed, i.e. the Poschiavino River/Lake Poschiavo system (Figures 2a & 3a).

Intralacustrine differentiation

The DAPC analysis highlights genomic differentiation between littoral and profundal dwelling individuals (Figure 6a). Differentiation between individuals from both habitats was however most strongly captured by different PC axes (Thun: PC4 – $F_{1,37}$=7.1, p=0.011; Lucern: PC2 – $F_{1,23}$=7.1, p=0.014; Walen: PC1 – $F_{1,19}$=4.9, p=0.040; Figure 6b & S2) accounting for 5.3, 3.6 and 8.6% of the total variation respectively (Figure S3). Depth associated differentiation based on global F_{ST} occurred within Lake Walen, showing a subtle, yet significant level of genetic differentiation (F_{ST} = 0.004, p = 0.025). Average genomic differentiation between littoral and profundal dwelling sculpins was not observed in lakes Lucern and Thun (Lucern: F_{ST} = 0.001, p = 0.129; Thun: F_{ST} = 0.002, p = 0.130). However our sample size for Lake Thun profundal

sculpins was very limited with just 5 fish. Nonetheless, all intralacustrine comparisons between littoral and profundal showed loci with much increased genetic differentiation, i.e. 11.5%, 15.7% and 17.7% of all loci with an $F_{ST} > 0.05$ for Lakes Lucern, Thun and Walen respectively, where loci were distributed across the genome (Figure 7). Although SAMβADA did not detect any water depth associated alleles within any lake (results not shown), we identified in all three lakes four common loci that had an $F_{ST} > 0.05$ between littoral and profundal dwelling individuals in all three lakes (Figure 7). Only one of these, a locus on chromosome 3 overlapped with a gene, i.e. catenin (cadherin-associated protein), delta 2b - *ctnnd2b* whose function in fish has not been determined (Hsu *et al.*, 2012).

Discussion

Secondary contact and habitat-associated differentiation

Strong genetic differentiation amongst European sculpin taxa commonly predates the last glaciation period, with distinct lineages restricted to particular drainage systems (Englbrecht *et al.*, 2000; Hänfling *et al.*, 2002). Consistent with postglacial colonization from distinct glacial refugia (Hänfling & Brandl, 1998a; b; Englbrecht *et al.*, 2000; Hänfling & Weetman, 2006) we found three deeply divergent allopatric sculpin lineages, likely belonging to the so-called danubian lineage of *C. gobio* (Englbrecht *et al.*, 2000) that are yet limited to distinct drainages (Figure 2&3). Unexpectedly we found two less divergent lineages to coexist in one of these drainages, i.e. the Aare/Rhine catchment, where individuals are associated with different environments, i.e. lake and streams, depending on their respective lineage (Figures 2,3 & 5).

The postglacial colonization of the peri-alpine region implies that any subsequent diversification within a drainage system is younger than allopatric differentiation for sculpins across Europe. We find considerable genetic substructure within and among stream dwelling

sculpins in each of our studied drainages. Previous studies have attributed population genetic structure within streams to low dispersal rates of sculpins and hence geographic isolation combined with more recent anthropogenic river fragmentation (Hänfling & Weetman, 2006; Junker *et al.*, 2012). Headwater stream populations in particular have been suggested to often harbour gene pools that are distinct from downstream populations, potentially also due to local adaptation (Junker *et al.*, 2012). In line with this, we find several of our headwater populations to be genetically most distinct within a particular river (i.e. Sense, Thur and Doubs; Figure 1, 3a; Table S2).

Much more surprisingly, we find genetic differentiation between lake and stream dwelling populations within the Aare catchment. Although sculpins are known to occur in either environment (Kontula & Vainola, 2001; Kottelat & Freyhof, 2007; Goto *et al.*, 2015), such a lake-stream differentiation within a drainage was unknown for *Cottus gobio* (Hänfling & Brandl, 1998a; b; Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008; Seehausen & Wagner, 2014; Goto *et al.*, 2015). Habitat dependent ecotype formation along a lake-stream axis is well known in some other freshwater fishes where it occurs with some regularity (Seehausen & Wagner, 2014). This is likely because lake and stream environments often require adaptation to different physical (flow, light, temperature) and biotic conditions (predation regimes, parasite communities, food resources; Seehausen & Wagner, 2014). However, our results differ importantly from classical stickleback and salmonid examples of lake-stream divergence, where distinct ecotypes have evolved repeatedly in parapatry from the same founding lineage. In these cases, geographically close populations of different ecotypes are usually more closely related than geographically distant populations of the same ecotype (Lucek *et al.*, 2013; Seehausen & Wagner, 2014; Theis *et al.*, 2014). In contrast, we find that within the Aare catchment geographically distant populations of lake dwelling sculpins cluster together while there are more distantly related to populations living in nearby streams (Figure 2&3). Parapatric populations from stream and lake show substantial genetic differentiation across the genome (Figure 4) with very limited evidence for gene flow (Figures 2,3 & 5).

This pattern is consistent with the colonization of the Aare catchment by two distinct lineages – one that now occurs mainly in streams and one that occurs predominantly in lakes. It is currently difficult to place these lineages in a detailed phylogeographic context because previous European phylogeographic studies did not provide the resolution to determine potential refugial areas around the Alps (Slechtová *et al.*, 2004; Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008). However, tributaries of Lake Constance are known to harbour the so-called danubian lineage of *C. gobio* that is distinct from those in the lower Rhine (Englbrecht *et al.*, 2000; Freyhof *et al.*, 2005). Our Lake Constance fish were distinct from those of the Aare lakes and cluster with geographically nearby stream populations as well as with fish from Lake Geneva (Figure 2&3), suggesting that they all belong to the same lineage. This would be the lineage described earlier from Lake Constance, potentially deriving from a Danubian refugium (Englbrecht *et al.*, 2000). The watershed crossing from the Aare/Rhine catchment into Lake Geneva was dated to 10-20 kyrs ago (Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008) further suggests that the colonization of northern Switzerland by the now geographically widely distributed and predominantly stream dwelling clade likely predates the colonization of the Aare catchment by the lineage that is now confined to lakes.

The possibility that the lake-dwelling populations in the Aare catchment represent a distinct genetic lineage is highlighted by i) by the low level of genetic differentiation of populations between even geographically distant lakes, which is significantly lower than that between geographically adjacent lake and stream populations or that among stream populations (Figure 3b, S1). ii) the clustering of lake populations in our PC analyses (Figure 5). iii) loci associated with lake-stream divergence are widespread across the genome rather than being confined to few outliers regions (Figure 4) which is consistent with allopatric genomic divergence (Feder *et al.*, 2012; Seehausen *et al.*, 2014). iv) the discordance in our phylogenetic network is consistent with gene flow among lake dwelling populations of the Aare catchment that has historically exceeded gene flow between lake and stream populations even if they were geographically adjacent (Figure 2). The potential for contemporary gene flow among lake

populations is naturally limited given the low dispersal rates of sculpins and the numerous anthropogenic migration barriers in Switzerland that prevent upstream migration in particular (Junker *et al.*, 2012). Downstream migration may be more likely, especially if lacustrine sculpins have pelagic larval stages that passively drift downstream, as has been observed in Lake Hallstätt (Austria; Wanzenböck *et al.*, 2000). Given the wide geographical distribution of the lacustrine lineage within the Aare catchment, and its complete geographical range overlap with the predominantly stream lineage, we observe surprisingly little genetic admixture suggesting reproductive isolation (Figure 3). However, fine-scaled parapatric gradients would need to be sampled to assess the degree of admixture.

The large geographical overlap between our genetically differentiated lake and stream sculpins is consistent with a scenario of postglacial secondary contact and niche (i.e. lake and stream) differentiation, potentially due to lineage interactions. Such character displacement, i.e. the shift of a species or lineage away from the ancestral state in regions of sympatry with a competing species of similar ancestral state is an important cause of ecological diversification (Schluter, 2000a; Stuart & Losos, 2013). It is possible that the first colonizing lineage did not populate lakes, but alternatively, it may have been displaced from that habitat by the second wave of colonization (Schluter, 2000a; Stuart & Losos, 2013). Our data indicates such a displacement given that we found the otherwise stream-dwelling lineage to occur in Lakes Geneva and Constance (Fig. 2 & 3), albeit in only low abundance (Alexander *et al.*, 2015). Both are isolated from the range of the Aare lake lineage by formidable dispersal barriers that formed early after the retreat of the glacial ice sheet and would have prevented a subsequent colonization: Lake Constance is isolated by the Rhine Falls at Schaffhausen (Figure 1) and Lake Geneva lost its connectivity to the Aare catchment and drains southwards into the Rhone (Vonlanthen *et al.*, 2007; Neuenschwander *et al.*, 2008). Further testing is, however, required to

distinguish between our hypothesis of character displacement and alternative scenarios such as competitive exclusion (Schluter, 2000a).

The onset of intralacustrine radiation

Intralacustrine radiations into different habitat and/or trophic niches are common among some salmonid and coregonid fish, and also occur in stickleback in one small part of the world, yet they have rarely been reported for other fish taxa in the northern temperate zone (Hendry *et al.*, 2009; Seehausen & Wagner, 2014). We found genetic evidence for intralacustrine differentiation between littoral and profundal sculpins in Lake Walen, coinciding with a bimodal depth distribution that cannot be explained by a sampling bias (Alexander *et al.*, 2015). Despite similar bimodal depth and genetic distributions in Lakes Thun and Lucern (Figure 6), genome-wide differentiation was non-significant in these lakes. Nonetheless, several loci showed substantial genetic differentiation especially so in Lake Thun (Figure 7), suggesting that our small sample sizes of the deep-water sculpins may have constrained our ability to detect genome-wide differentiation. Larger sample sizes are required to confirm or reject the genomic distinctiveness of deep-water sculpins in these lakes in the future, but these fish are difficult to catch at depths of 150 to 250 meters. We cannot rule out that the segregation into littoral and profundal sculpins could be accounted for by plasticity (Edelaar & Bolnick, 2012). However, given the abundance of loci within each lake that show differentiation, the observed depth-related differentiation is likely to have an at least partly heritable basis.

Conclusions

Taken together, our results provide a case for postglacial habitat dependent divergence in sculpins along the lake-stream axis at a catchment scale. The observed pattern could reflect character displacement in the early-arrived lineage upon arrival of the second lineage. In the absence of such habitat contrasts, secondary contact among distinct sculpin lineages usually

leads to strong tension zones within streams (Nolte *et al.*, 2005; 2009). We further find some evidence for depth related early stage lacustrine radiations between sculpins inhabiting the littoral and profundal habitats within the lake lineage. Both findings were surprising given that population structure in *Cottus* has often been deemed to be driven by low dispersal range, geographic isolation or anthropogenic river fragmentation (Hänfling & Weetman, 2006; Junker *et al.*, 2012). This highlights the necessity of sampling a broad geographic and habitat range as well as using a large number of genomic markers to detect intraspecific variation in the appropriate context (Bickford *et al.*, 2007). This is particularly important for biodiversity assessment and conservation, as single taxonomic units may comprise ecologically and genetically distinct subunits or incipient species that require diversity-aware management.

Data accessibility

BAM files with aligned de-multiplexed and base-quality score recalibrated reads are available through the short read archive (www.ncbi.nlm.nih.gov/sra). BioProject ID: PRJNA416944.

Acknowledgement

This research was supported through a grant from the ETH Competence Centre for Environment and Sustainability in the project GeneMig, and Eawag/Bafu strategic project "Projet Lac" to. was further supported by an SNSF Early Postdoc.Mobility grant P2BEP3_152103 and by SNSF grant 31003A_166322. We thank Nicola Nadeau and one anonymous reviewer for their constructive input on an earlier draft. We thank everyone who helped with fieldwork, in particular Pascal Vonlanthen, Guy Periat and Tim Alexander for extensive effort to collect lacustrine profundal sculpins during Projet Lac. We further thank B. Almasi, C. Baumgartner, L. Costa, J.-M. Fierz, B. Germann, M. Grünenfelder, S. Haertel, J. Hellmann, J. Junker, J. Koegler, A. Lièvre, A. Peter, B. Polli, C. Rau, E. Schager, J. Schuler, D. Senn, B. Tschirren, H. Walther, P.

Warnier, and A. Westram for help during fieldwork. We are grateful to the fisheries authorities of Cantons Bern, Graubünden, Jura, Luzern, St. Gallen, Solothurn, Thurgau, Ticino, Uri and the province of Sondrio, Italy, for their support of this project.

Figure captions

Figure 1

Map of Switzerland with the sampled locations indicated (see Table S1 for details) and the respective drainages highlighted by colour, i.e. North Sea (Aare/Rhine catchment), Mediterranean Sea (Doubs/Rhone catchment), Adriatic Mediterranean Sea (Maggia/Po and Poschiavino/Po catchments) . Green triangles indicate lake sites and blue circles stream sites. Grey triangles indicate sites that were surveyed but not included in the final genetic data set following quality filtering. The asterisk indicates the Rhine fall, which separates the Lake Constance catchment from the lower Rhine. Black arrows indicate the water flow of the respective rivers.

Figure 2:

Individual based phylogenetic relationship established from a SPLITSTREE analysis.

Figure 3:

Summary of the population structure across populations. a) Individual based assignment using FASTSTRUCTURE comprising either all individuals combined (top; *K*=7) or ran separately for each drainage (*K*=10, 3, 2 respectively). Shown are in each case the individual based assignments for the best-supported number of genetic clusters (i.e. *K*; see main text for details).

The population from Lake Geneva was excluded from the Aare/Rhine specific analysis. b) Boxplot summarizing pairwise AMOVA based F_{ST} among lake and stream dwelling populations from the Aare catchment (see Table S2 for details). Significances are based on an ANOVA with a *post hoc* TukeyHSD correction.

Figure 4:

Genomic differentiation across the genome between lake and stream sculpins. a) Lake Zurich versus Aare at Ruppoldingen (Ru), b) Lake Lucern versus Aare at Ruppoldingen (Ru), c) Lake Thun versus Aare at Kiesen (Km). SNP positions are in relation to the stickleback genome (see main text for details). Common outlier loci between lake and stream dwelling populations identified by SAMβADA within the Aare/Rhine catchment are highlighted in red. Boxplots summarizing F_{ST} for outlier (red) and non-outlier (grey) loci respectively. *P*-values are based on one-way ANOVAs.

Figure 5:

SNP based principal components analyses for individuals from the Aare/Rhine catchment using either a) all loci or b) outlier loci as identified by SAMβADA. Colours represent individuals from lakes (green) or streams (blue). Individuals from the Aare-Ru stream site are highlighted in orange. The 95% confidence ellipsoids for different geographic groups are highlighted.

Figure 6:

Genetic structure and distribution of individuals within lakes Thun, Lucern and Walen: a) Density distributions of individuals along the main discriminant axis (DF1) based on a DAPC analysis using the 10 leading PC axes (see main text for details). Density distributions are given for littoral (blue) and profundal (red) individuals. b) Principal component (PC) plots for each

lake. For Thun, PC1 was plotted against PC4, which showed the strongest differentiation between habitats (see Figure S2 & S3). c) Frequency distribution of sculpins caught at a given depth for each lake. d) Representative images for individuals from littoral (top) and profundal (bottom) habitats for each lake.

Figure 7:

Genomic differentiation between profundal and littoral individuals within lakes Thun, Lucern and Walen. SNP positions are in relation to the stickleback genome (see main text for details). Red dots indicate SNPs with $F_{ST} > 0.05$ in all three comparisons.

Figure S1:

Boxplot summarizing pairwise F'_{ST}s among lake and stream dwelling populations from the Aare catchment (see Table S3 for details). Significances are based on an ANOVA with a *post hoc* TukeyHSD correction.

Figure S2:

Boxplots summarizing PC scores for littoral (blue) and profundal (red) dwelling individuals for lakes Thun, Lucern and Walen. *P*-values are based on one-way ANOVAs (see main text for details).

Figure S3:

Scree plots for the principal component (PC) analyses used for the DAPC analysis, showing the explained variance for each PC axis.

References

Alexander, T.J., Vonlanthen, P., Periat, G. & Degiorgi, F., Raymond, J.C., Seehausen, O. 2015. Estimating whole-lake fish catch per unit effort. *Fish. Res.* **117**: 287-302.

Bernatchez, L. 2004. Ecological theory of adaptive radiation - an empirical assessment from coregonine fishes (Salmoniformes). In: *Evolution Illuminated: Salmon and their relatives* (A. P. Hendry & S. C. Stearns, eds), pp. 175–207. Oxford University Press, Oxford, UK.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., *et al.* 2007. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol*. **22**: 148–155.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**: 3124–3140.

Cheng, J., Czypionka, T. & Nolte, A.W. 2013. The genomics of incompatibility factors and sex determination in hybridizing species of *Cottus* (Pisces). *Heredity* **111**: 520–529.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Davey, A., Hawkins, S.J. & Turner, G.F. 2005. Size-dependent microhabitat use and intraspecific competition in *Cottus gobio*. *J. Fish Biol.* **67**: 428–443.

Dennenmoser, S., Vamosi, S.M., Nolte, A.W. & Rogers, S.M. 2017. Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (*Cottus asper*) revealed by Pool-Seq. *Mol. Ecol.* **26**: 25–42.

Edelaar, P. & Bolnick, D.I. 2012. Non-random gene flow: an underappreciated force in evolution and ecology. *Trends Ecol. Evol.* **27**: 659–665.

Englbrecht, C.C., Freyhof, J., Nolte, A., Rassmann, K., Schliewen, U. & Tautz, D. 2000. Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-Pleistocene origin of the major central European populations. *Mol. Ecol.* **9**: 709–722.

Excoffier, L. & Lischer, H.E.L.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**: 564–567.

Feder, J.L., Egan, S.P. & Nosil, P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet.* **28**: 342–350.

Freyhof, J., Kottelat, M. & Nolte, A. 2005. Taxonomic diversity of European *Cottus* with description of eight new species (*Teleostei: Cottidae*). *Ichth. Exp. Freshw.* **16**, 107-172.

Goto, A., Yokoyama, R. & Sideleva, V.G. 2015. Evolutionary diversification in freshwater sculpins (Cottidea): a review of two major adaptive radiations. *Env. Biol. Fish*. **98**, 307-335.

Hänfling, B. & Brandl, R. 1998a. Genetic differentiation of the bullhead *Cottus gobio* L. across watersheds in Central Europe: evidence for two taxa. *Heredity*. **80**, 110-117.

Hänfling, B. & Brandl, R. 1998b. Genetic variability, population size and isolation of distinct

populations in the freshwater fish *Cottus gobio* L. *Mol. Ecol*. **7**, 1625-1632.

Hänfling, B. & Weetman, D. 2006. Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the river sculpin, *Cottus gobio*. *Genetics.* **173**: 1487–1501.

Hänfling, B., Hellemans, B., Volckaert, F.A.M. & Carvalho, G.R. 2002. Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Mol. Ecol.* **11**: 1717–1729.

Hedrick, P.W. 2005. A standardized genetic differentiation measure. *Evolution* **59**: 1633–1638.

Hendry, A.P. 2009. Ecological speciation! Or the lack thereof? *Can. J. Fish. Aquat. Sci.* **66**: 1383–1398.

Hendry, A.P., Bolnick, D.I., Berner, D. & Peichel, C.L. 2009. Along the speciation continuum in sticklebacks. *J. Fish Biol.* **75**: 2000–2036.

Hsu, C.L., Muerdter, C.P., Knickerbocker, A.D., Walsh, R.M., Zepeda-Rivera, M.A., Depner, K.H., *et al.* 2012. Cdc42 GTPase and Rac1 GTPase act downstream of p120 catenin and require GTP exchange during gastrulation of zebrafish mesoderm. *Dev. Dyn.* **241**: 1545–1561.

Hudson, A.G., Vonlanthen, P., Müller, R. & Seehausen, O. 2007. Review: The geography of speciation and adaptive radiation in coregonines. *Advanc. Limnol.* **60**: 111-146.

Huson, D.H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**: 254–267.

Jombart, T., Devillard, S. & Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**: 94.

Jonsson, B. & Jonsson, N. 2001. Polymorphism and speciation in Arctic charr. *J. Fish Biol.* **58**: 605–638.

Junker, J., Peter, A., Wagner, C.E., Mwaiko, S., Germann, B., Seehausen, O., *et al.* 2012. River fragmentation increases localized population genetic structure and enhances asymmetry of dispersal in bullhead (*Cottus gobio*). *Cons. Genet.* **13**, 545–556.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W. & Prodöhl, P.A. 2013. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* **4**: 782–788.

Kontula, T. & Vainola, R. 2001. Postglacial colonization of Northern Europe by distinct phylogeographic lineages of the bullhead, *Cottus gobio*. *Mol. Ecol.* **10**: 1983–2002.

Kontula, T., Kirilchik, S.V. & Väinölä, R. 2003. Endemic diversification of the monophyletic cottoid fish species flock in Lake Baikal explored with mtDNA sequencing. *Mol. Phylo. Evol.* **27**: 143–155.

Kottelat, M. & Freyhof, J. 2007. *Handbook of European freshwater fishes*. Kottelat, Cornol,

Switzerland.

Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie2. *Nat. Methods* **9**: 357–359.

Lucek, K., Sivasundar, A., Roy, D. & Seehausen, O. 2013. Repeated and predictable patterns of ecotypic differentiation during a biological invasion: lake-stream divergence in parapatric Swiss stickleback. *J. Evol. Biol.* **26**: 2691–2709.

Marques, D.A., Lucek, K., Meier, J.I., Mwaiko, S., Wagner, C.E., Excoffier, L., *et al.* 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet.* **12**: e1005887.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**: 1297–1303.

McKinnon, J.S. & Rundle, H. 2002. Speciation in nature: the threespine stickleback model systems. *Trends Ecol. Evol.* **17**: 480–488.

Meirmans, P.G. & Hedrick, P.W. 2011. Assessing population structure: FST and related measures. *Mol. Ecol. Resour.* **11**, 5-18.

Neuenschwander, S., Largiadèr, C.R., Ray, N., Currat, M., Vonlanthen, P. & Excoffier, L. 2008. Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol. Ecol.* **17**: 757–772.

Nolte, A.W., Freyhof, J., Stemshorn, K. & Tautz, D. 2005. An invasive lineage of sculpins, *Cottus* sp (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from hybridization between old phylogeographic groups. *P. R. Soc. B.* **272**: 2379–2387.

Nolte, A.W., Gompert, Z. & Buerkle, C.A. 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Mol. Ecol.* **18**: 2615–2627.

Nosil, P. 2012 *Ecological speciation*. Oxford University Press, Oxford, UK.

Raj, A., Stephens, M. & Pritchard, J.K. 2014. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **197**: 573–589.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., *et al.* 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**: 687–695.

Schluter, D. 2000a. Ecological character displacement in adaptive radiation. *Am. Nat.* **156**: S4–S16.

Schluter D. 2000 *The ecology of adaptive radiation*. Oxford University Press, Oxford, UK.

Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**: 198–207.

Seehausen, O. & Wagner, C.E. 2014. Speciation in freshwater fishes. *Annu. Rev. Ecol. Evol. S.* **45**:

621–651.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., *et al.* 2014. Genomics and the origin of species. *Nat. Rev. Genet.* **15**: 176–192.

Slechtová, V., Bohlen, J., Freyhof, J., Persat, H. & Delmastro, G.B. 2004. The Alps as barrier to dispersal in cold-adapted freshwater fishes? Phylogeographic history and taxonomic status of the bullhead in the Adriatic freshwater drainage. *Mol. Phylo. Evol.* **33**: 225–239.

Smolka, M., Rescheneder, P., Schatz, M.C., Haeseler, von, A. & Sedlazeck, F.J. 2015. Teaser: Individualized benchmarking and optimization of read mapping results for NGS data. *Genome Biol.* **16**: 235.

Stuart, Y.E. & Losos, J.B. 2013. Ecological character displacement: glass half full or half empty? *Trends Ecol. Evol.* **28**: 402–408.

Stucki, S., Orozco-terWengel, P., Forester, B.R., Duruz, S., Colli, L., Masembe, C., *et al.* 2017. High performance computation of landscape genomic models including local indicators of spatial association. *Mol. Ecol. Resour.* **17**: 1072–1089.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., *et al.* 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**: D447–452.

Theis, A., Ronco, F., Indermaur, A., Salzburger, W. & Egger, B. 2014. Adaptive divergence between lake and stream populations of an East African cichlid fish. *Mol. Ecol.* **23**: 5304–5322.

Vonlanthen, P., Excoffier, L., Bittner, D., Persat, H., Neuenschwander, S. & Largiadèr, C.R. 2007. Genetic analysis of potential postglacial watershed crossings in Central Europe by the bullhead (*Cottus gobio* L.). *Mol. Ecol.* **16**: 4572–4584.

Wanzenböck, J., Lahnsteiner, B. & Maier, K. 2000. Pelagic early life phase of the bullhead in a freshwater lake. *J. Fish Biol.* **56**, 1553-1557.

Yoder, J.B., Clancey, E., Roches, Des, S., Eastman, J.M., Gentry, L., Godsoe, W., *et al.* 2010. Ecological opportunity and the origin of adaptive radiations. *J. Evol. Biol.* **23**: 1581–1596.

Zheng, X.W., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. & S, W.B. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**: 3326–3328.